UNIVERSITY OF CALIFORNIA SAN DIEGO

Association of TET loss of function with heterochromatin DNA hypomethylation and genome instability

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

in

Bioinformatics and Systems Biology

by

Isaac F. López Moyado

Committee in charge:

Professor Anjana Rao, Chair
Professor Vineet Bafna, Co-Chair
Professor Terry Gaasterland
Professor Olivier Harismendy
Professor Bing Ren
Professor Sheng Zhong

2019

The Dissertation of Isaac F. López Moyado is approved, and is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Chair

University of California San Diego

2019

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 2

Chapter 3

ACKNOWLEDGEMENTS

Chapter 2, in full, is a reprint with modifications as it appears in Paradoxical association of TET loss of function with genome-wide hypomethylation, *PNAS* (2019), published online August 1st, 2019. doi: 10.1073/pnas.1903059116. The dissertation author was the primary author of this paper. Other authors include Ageliki Tsagaratou, Hiroshi Yuita, Hyungseok Seo, Benjamin Delatte, Sven Heinz, Chris Benner & Anjana Rao.

Chapter 3, sections 3.2.1-3.2.3, is a reprint with modifications as it appears in Paradoxical association of TET loss of function with genome-wide hypomethylation, *PNAS* (2019), published online August 1st, 2019. doi: 10.1073/pnas.1903059116. The dissertation author was the primary author of this paper. Other authors include Ageliki Tsagaratou, Hiroshi Yuita, Hyungseok Seo, Benjamin Delatte, Sven Heinz, Chris Benner & Anjana Rao.

Chapter 3, sections 3.2.4-3.2.5, is part of a manuscript under preparation. The dissertation author is the primary author of this paper along with Romain O. Georges. Other authors include Romain O. Georges, Ageliki Tsagaratou, Benjamin Delatte & Anjana Rao.

VITA

2012 Bachelor of Science, National Autonomous University of Mexico

2019 Doctor of Philosophy, University of California San Diego

PUBLICATIONS

1. **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. Proc Natl Acad Sci U S A.
2. Seo, H., Chen, J., González-Avalos, E., Samaniego-Castruita, D., Das, A., Wang, Y.H., **López-Moyado, I.F.**, Georges, R.O., Zhang, W., Onodera, A.*, et al.* (2019). TOX and TOX2 transcription factors cooperate with NR4A transcription factors to impose CD8. Proc Natl Acad Sci U S A *116*, 12410-12415.
3. Chen, J., **López-Moyado, I.F.**, Seo, H., Lio, C.J., Hempleman, L.J., Sekiya, T., Yoshimura, A., Scott-Browne, J.P., and Rao, A. (2019). NR4A transcription factors limit CAR T cell function in solid tumours. Nature *567*, 530-534.
4. Scott-Browne, J.P., **López-Moyado, I.F.**, Trifari, S., Wong, V., Chavez, L., Rao, A., and Pereira, R.M. (2016). Dynamic Changes in Chromatin Accessibility Occur in CD8. Immunity *45*, 1327-1340.
5. Stroustrup, N., Anthony, W.E., Nash, Z.M., Gowda, V., Gomez, A., **López-Moyado, I.F.**, Apfeld, J., and Fontana, W. (2016). The temporal scaling of Caenorhabditis elegans ageing. Nature *530*, 103-107.
6. 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. Nat Commun *6*, 10071.
7. Bakowski, M.A., Desjardins, C.A., Smelkinson, M.G., Dunbar, T.L., Dunbar, T.A., **Lopez-Moyado, I.F.**, Rifkin, S.A., Cuomo, C.A., and Troemel, E.R. (2014). Ubiquitin-mediated response to microsporidia and virus infection in C. elegans. PLoS Pathog *10*, e1004200.
8. Fernandes de Abreu, D.A., Caballero, A., Fardel, P., Stroustrup, N., Chen, Z., Lee, K., Keyes, W.D., Nash, Z.M., **López-Moyado, I.F.**, Vaggi, F.*, et al.* (2014). An insulin-to-insulin regulatory network orchestrates phenotypic specificity in development and physiology. PLoS Genet *10*, e1004225.
9. Stroustrup, N., Ulmschneider, B.E., Nash, Z.M., **López-Moyado, I.F.**, Apfeld, J., and Fontana, W. (2013). The Caenorhabditis elegans Lifespan Machine. Nat Methods *10*, 665-670.

ABSTRACT OF THE DISSERTATION

Association of TET loss of function with heterochromatin DNA hypomethylation and genome instability

by

Isaac F. López Moyado

Doctor of Philosophy in Bioinformatics and Systems Biology

University of California San Diego, 2019

Professor Anjana Rao, Chair
Professor Vineet Bafna, Co-Chair

Cancer genomes are characterized by focal increases in DNA methylation, co-occurring with widespread hypomethylation. Here we show that TET loss-of-function results in a similar genomic footprint. Both 5hmC in wildtype genomes, and DNA hypermethylation in TET-deficient genomes, are largely confined to the active euchromatic compartment, consistent with the known functions of TET proteins in DNA demethylation and the known distribution of 5hmC at transcribed

genes and active enhancers. In contrast, an unexpected DNA hypomethylation noted in multiple TET-deficient genomes is primarily observed in the heterochromatin compartment. In a mouse model of T cell lymphoma driven by TET deficiency (*Tet2/3 DKO* T cells), genomic analysis of malignant T cells revealed DNA hypomethylation in the heterochromatic genomic compartment, as well as reactivation of repeat elements and enrichment for single nucleotide alterations, primarily in heterochromatic regions of the genome. Furthermore, like DNMT-deficient cells, expanded *Tet2/3 DKO* T cells displayed recurrent aneuploidies and increased accumulation of DNA double-strand breaks. Noteworthy, hematopoietic stem/precursor cells (HSPC) doubly deficient for *Tet2* and *Dnmt3a* displayed greater losses of DNA methylation than HSPC singly deficient for *Tet2* or *Dnmt3a* alone, potentially explaining the unexpected synergy between *DNMT3A* and *TET2* mutations in myeloid and lymphoid malignancies. *Tet1*-deficient cells showed decreased localization of Dnmt3a in the heterochromatin compartment compared to WT cells, pointing to a functional interaction between TET and DNMT proteins and providing a potential explanation for the hypomethylation observed in *TET*-deficient genomes. Our data suggest that TET loss-of-function may at least partially underlie the characteristic pattern of global hypomethylation coupled to regional hypermethylation observed in diverse cancer genomes, propose that heterochromatin DNA hypomethylation may at least partially explain some of the shared features between TET and DNMT mutants, and highlight the potential contribution of heterochromatin hypomethylation to oncogenesis.

Chapter 1

Introduction

In addition to the four major bases present in the DNA - adenine, guanine, cytosine and thymine, it has been known for almost 45 years that the cytosine base can be modified to produce a new nucleotide known as 5-methylcytosine (5mC), which could act as an epigenetic mark (Holliday and Pugh 1975, Riggs 1975), and whose distribution has been generally associated with transcriptional silencing (Schübeler 2015, Bestor, Edwards and Boulard 2015). Furthermore, in 2009 a collaboration between the Aravind and Rao labs (Iyer et al. 2009, Tahiliani et al. 2009) led to the discovery of a new family of enzymes, the TET (Ten-Eleven Translocation) proteins, which would provide an answer to the long-standing question of the mechanism of cytosine demethylation.

1.1 TET enzymes

TET enzymes are Fe(II) and α-ketoglutarate-dependent dioxygenases that mediate DNA demethylation through sequential oxidation of the methyl group of 5-methylcytosine (5mC) to 5-hydroxymethyl, 5-formyl and 5-carboxylcytosine (5hmC, 5fC and 5caC)(Tahiliani et al. 2009, Ito et al. 2011, He et al. 2011). The oxidized methylcytosines (oxi-mC) generated by TET proteins are intermediates in at least two pathways of DNA demethylation: (i) replication-dependent loss of methylation, reflecting inability of the DNMT1/UHRF1 complex to methylate unmodified CpGs on newly replicated DNA strands if an oxi-mC (rather than 5mC) is present on the template strand, and (ii) a replication-independent process in which thymine DNA glycosylase (TDG) excises 5fC and 5caC, which are then replaced with unmodified cytosine through base excision repair (Pastor, Aravind and Rao 2013, Wu and Zhang 2017).

Even in the absence of TET coding region mutations, TET loss-of-function and low 5hmC levels are strongly associated with cancer (Ko et al. 2010, Marçais et al. 2017, Lemonnier et al. 2018, Huang and Rao 2014). *TET2* mutations are frequent in diverse hematopoietic malignancies, including myelodysplastic syndromes (MDS), acute myeloid leukemias (AML) and peripheral T

cell lymphomas (PTCL) (Delhommeau et al. 2009, Langemeijer et al. 2009, Lemonnier et al. 2012, Odejide et al. 2014). However, both solid tumors and hematopoietic malignancies display TET loss-of-function without TET coding region mutations, as a result of TET promoter methylation, increased degradation of TET proteins, or aberrant microRNA expression (Cimmino et al. 2015, Ko, An and Rao 2015, Raffel et al. 2017, Wu et al. 2018). In addition, hypoxia and a variety of metabolic alterations impair the enzymatic activity of TET and other dioxygenases, by decreasing the levels of the substrates α-ketoglutarate and molecular oxygen or increasing the levels of the inhibitor 2-hydroxyglutarate (Losman and Kaelin 2013, Kaelin and McKnight 2013, Huang and Rao 2014, Ko et al. 2015).

Based on the known biochemical activities of TET-family proteins in oxidizing 5mC (Pastor et al. 2013), *TET* loss-of-function mutations are expected to result in gains of DNA methylation. In fact, increased methylation as a result of TET loss-of-function has been documented at many genomic regions including promoters, enhancers and CTCF sites (Cimmino et al. 2015, Rasmussen et al. 2015, Flavahan et al. 2016, Yue et al. 2016, Tsagaratou et al. 2017). Unexpectedly, however, a small number of studies have noted widespread decreases of DNA methylation in the genomes of TET-deficient cells, by mapping differentially methylated regions (DMRs) or by comparing methylation values at random genomic regions after whole-genome bisulfite sequencing (WGBS) (Hon et al. 2014, Lu et al. 2014, An et al. 2015, Tsagaratou et al. 2017). Because the hypomethylated regions observed in TET loss-of-function did not overlap with active or regulatory regions of the genome, these findings were largely ignored.

1.2 Nuclear compartments: Euchromatin and Heterochromatin

Principal component analysis of the interaction matrix obtained from Hi-C data has been used to compartmentalize the genome into an A compartment (positive PC1 values) and a B compartment (negative PC1 values) that exhibit the hallmark characteristics of euchromatin and

heterochromatin, respectively (Lieberman-Aiden et al. 2009, van Steensel and Belmont 2017). The euchromatic A compartment is rich in expressed genes in the cell type under consideration, whereas the heterochromatic B compartment is gene-poor and bears epigenetically “repressive” chromatin marks, including H3K9me2/3 (van Steensel and Belmont 2017, Lieberman-Aiden et al. 2009). Moreover, the Hi-C B compartment overlaps with lamina-associated domains and corresponds to late-replicating regions of the genome, whereas the Hi-C A compartment corresponds to early replicating genomic regions and is not lamina-associated (Hiratani et al. 2008, Peric-Hupkes et al. 2010). Notably, the extended, partially methylated domains (PMDs) observed in cancer genomes overlap with Hi-C B compartment, late-replicating, nuclear lamina-associated domains (Berman et al. 2011, Hon et al. 2012, Zhou et al. 2018). In the remainder of this dissertation we will refer to the Hi-C A and B compartments as euchromatic and heterochromatic compartments, respectively.

1.3 DNA methylation in cancer genomes

Cancer genomes are characterized by two opposing patterns of aberrant DNA methylation: focal or regional hypermethylation (Jones and Baylin 2002, Baylin and Jones 2016) and widespread DNA hypomethylation (Feinberg and Vogelstein 1983, Ehrlich 2009). DNA hypermethylation at promoters and enhancers contributes to oncogenesis through transcriptional silencing of genes involved in DNA damage repair and tumor suppressors (Jones and Baylin 2002, Baylin and Jones 2016). Local hypermethylation has been shown to reflect the impaired expression or activity of TET proteins. Despite our understanding of the biochemical and biological consequences of local hypermethylation, however, the causes and consequences of DNA hypomethylation in cancer genomes are less well understood.

DNA hypomethylation has been associated with many biological consequences, including reactivation of repeat elements, genome instability, chromosome segregation defects and

aneuploidies, and sharply increased mutation rates (Walsh, Chaillet and Bestor 1998, Gaudet et al. 2003, Eden et al. 2003, Chen et al. 1998). For instance, mice with a hypomorphic mutation in *Dnmt1* displayed a T cell lymphoma that occurred in 80% of the cases and was characterized by recurrent aneuploidies of chromosome 15 and hypomethylation of centromeric repeats (Gaudet et al. 2003); and mice with disruption of the *Dnmt3b* gene showed hypomethylation of centromeric minor satellite repeats and early embryonic lethality (Okano et al. 1999). Reactivation of transposable elements is prevalent in cancer genomes, and is associated with the formation of RNA-DNA hybrids and R-loops (Zeller et al. 2016, Zhu et al. 2018), which in turn have been linked to DNA damage and the appearance of DNA double-strand breaks (Crossley, Bocek and Cimprich 2019). Moreover, mutations in proteins associated with the maintenance of heterochromatin integrity are frequent in cancer, leading to the postulate that many cancers may arise as a consequence of heterochromatin dysfunction (Janssen, Colmenares and Karpen 2018).

1.4 Scope of this work

Here we use a combination of Hi-C and WGBS data to document the DNA methylation changes associated with TET loss-of-function in diverse TET-deficient cell types. We show that 5hmC in wildtype genomes, and DNA hypermethylation in TET-deficient genomes, are largely confined to the euchromatic Hi-C A compartment. This finding is consistent with the known functions of TET proteins in DNA demethylation and the known distribution of 5hmC at active enhancers and in the gene bodies of highly transcribed genes. In contrast, we show that the unexpected DNA hypomethylation noted in TET-deficient genomes is primarily present in the heterochromatic Hi-C B compartment. Moreover, TET-deficient cells showed a strong increase in chromosome segregation defects and aneuploidies, as also observed in DNMT-deficient cells (Xu et al. 1999, Gaudet et al. 2003, Bourc'his and Bestor 2004). They also showed reactivation of

repeat elements and a pronounced enrichment for single-nucleotide variations (SNVs) in the Hi-C B compartment; this feature is characteristic of cancer genomes, in which mutation rates are elevated in genome compartments marked by H3K9me3 (Schuster-Böckler and Lehner 2012). We also show that Dnmt3a relocalizes from the heterochromatic compartment to the euchromatic compartment in *Tet1*-deficient mESC, providing a potential mechanism for the heterochromatin hypomethylation observed in TET-deficient genomes. Our results are consistent with the co-occurrence of *DNMT3A* and *TET2* mutations in human cancers and the more pronounced leukemic phenotype observed in double *Tet2/Dnmt3a*-deficient mice compared to mice with individual disruption of *Tet2* or *Dnmt3a* alone. Taken together, these data point to a functional interaction between TET proteins and DNMTs, and highlight the potential contribution of heterochromatic dysfunction to oncogenesis.

1.5 References

An, J., E. González-Avalos, A. Chawla, M. Jeong, I. F. López-Moyado, W. Li, M. A. Goodell, L. Chavez, M. Ko & A. Rao (2015) Acute loss of TET function results in aggressive myeloid cancer in mice. *Nat Commun,* 6**,** 10071.

Baylin, S. B. & P. A. Jones (2016) Epigenetic Determinants of Cancer. *Cold Spring Harb Perspect Biol,* 8.

Berman, B. P., D. J. Weisenberger, J. F. Aman, T. Hinoue, Z. Ramjan, Y. Liu, H. Noushmehr, C. P. Lange, C. M. van Dijk, R. A. Tollenaar, D. Van Den Berg & P. W. Laird (2011) Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat Genet,* 44**,** 40-6.

Bestor, T. H., J. R. Edwards & M. Boulard (2015) Notes on the role of dynamic DNA methylation in mammalian development. *Proc Natl Acad Sci U S A,* 112**,** 6796-9.

Bourc'his, D. & T. H. Bestor (2004) Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature,* 431**,** 96-9.

Chen, R. Z., U. Pettersson, C. Beard, L. Jackson-Grusby & R. Jaenisch (1998) DNA hypomethylation leads to elevated mutation rates. *Nature,* 395**,** 89-93.

Cimmino, L., M. M. Dawlaty, D. Ndiaye-Lobry, Y. S. Yap, S. Bakogianni, Y. Yu, S. Bhattacharyya, R. Shaknovich, H. Geng, C. Lobry, J. Mullenders, B. King, T. Trimarchi, B. Aranda-Orgilles, C. Liu, S. Shen, A. K. Verma, R. Jaenisch & I. Aifantis (2015) TET1 is a tumor suppressor of hematopoietic malignancy. *Nat Immunol,* 16**,** 653-62.

Crossley, M. P., M. Bocek & K. A. Cimprich (2019) R-Loops as Cellular Regulators and Genomic Threats. *Mol Cell,* 73**,** 398-411.

Delhommeau, F., S. Dupont, V. Della Valle, C. James, S. Trannoy, A. Massé, O. Kosmider, J. P. Le Couedic, F. Robert, A. Alberdi, Y. Lécluse, I. Plo, F. J. Dreyfus, C. Marzac, N. Casadevall, C. Lacombe, S. P. Romana, P. Dessen, J. Soulier, F. Viguié, M. Fontenay, W. Vainchenker & O. A. Bernard (2009) Mutation in TET2 in myeloid cancers. *N Engl J Med,* 360**,** 2289-301.

Eden, A., F. Gaudet, A. Waghmare & R. Jaenisch (2003) Chromosomal instability and tumors promoted by DNA hypomethylation. *Science,* 300**,** 455.

Ehrlich, M. (2009) DNA hypomethylation in cancer cells. *Epigenomics,* 1**,** 239-59.

Feinberg, A. P. & B. Vogelstein (1983) Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature,* 301**,** 89-92.

Flavahan, W. A., Y. Drier, B. B. Liau, S. M. Gillespie, A. S. Venteicher, A. O. Stemmer-Rachamimov, M. L. Suvà & B. E. Bernstein (2016) Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature,* 529**,** 110-4.

Gaudet, F., J. G. Hodgson, A. Eden, L. Jackson-Grusby, J. Dausman, J. W. Gray, H. Leonhardt & R. Jaenisch (2003) Induction of tumors in mice by genomic hypomethylation. *Science,* 300**,** 489-92.

He, Y. F., B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He & G. L. Xu (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science,* 333**,** 1303-7.

Hiratani, I., T. Ryba, M. Itoh, T. Yokochi, M. Schwaiger, C. W. Chang, Y. Lyou, T. M. Townes, D. Schübeler & D. M. Gilbert (2008) Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol,* 6**,** e245.

Holliday, R. & J. E. Pugh (1975) DNA modification mechanisms and gene activity during development. *Science,* 187**,** 226-32.

Hon, G. C., R. D. Hawkins, O. L. Caballero, C. Lo, R. Lister, M. Pelizzola, A. Valsesia, Z. Ye, S. Kuan, L. E. Edsall, A. A. Camargo, B. J. Stevenson, J. R. Ecker, V. Bafna, R. L. Strausberg, A. J. Simpson & B. Ren (2012) Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res,* 22**,** 246-58.

Hon, G. C., C. X. Song, T. Du, F. Jin, S. Selvaraj, A. Y. Lee, C. A. Yen, Z. Ye, S. Q. Mao, B. A. Wang, S. Kuan, L. E. Edsall, B. S. Zhao, G. L. Xu, C. He & B. Ren (2014) 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Mol Cell,* 56**,** 286-297.

Huang, Y. & A. Rao (2014) Connections between TET proteins and aberrant DNA modification in cancer. *Trends Genet,* 30**,** 464-74.

Ito, S., L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He & Y. Zhang (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science,* 333**,** 1300-3.

Iyer, L. M., M. Tahiliani, A. Rao & L. Aravind (2009) Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle,* 8**,** 1698-710.

Janssen, A., S. U. Colmenares & G. H. Karpen (2018) Heterochromatin: Guardian of the Genome. *Annu Rev Cell Dev Biol,* 34**,** 265-288.

Jones, P. A. & S. B. Baylin (2002) The fundamental role of epigenetic events in cancer. *Nat Rev Genet,* 3**,** 415-28.

Kaelin, W. G. & S. L. McKnight (2013) Influence of metabolism on epigenetics and disease. *Cell,* 153**,** 56-69.

Ko, M., J. An & A. Rao (2015) DNA methylation and hydroxymethylation in hematologic differentiation and transformation. *Curr Opin Cell Biol,* 37**,** 91-101.

Ko, M., Y. Huang, A. M. Jankowska, U. J. Pape, M. Tahiliani, H. S. Bandukwala, J. An, E. D. Lamperti, K. P. Koh, R. Ganetzky, X. S. Liu, L. Aravind, S. Agarwal, J. P. Maciejewski & A. Rao (2010) Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature,* 468**,** 839-43.

Langemeijer, S. M., R. P. Kuiper, M. Berends, R. Knops, M. G. Aslanyan, M. Massop, E. Stevens-Linders, P. van Hoogen, A. G. van Kessel, R. A. Raymakers, E. J. Kamping, G. E. Verhoef,

E. Verburgh, A. Hagemeijer, P. Vandenberghe, T. de Witte, B. A. van der Reijden & J. H. Jansen (2009) Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat Genet,* 41**,** 838-42.

Lemonnier, F., L. Couronné, M. Parrens, J. P. Jaïs, M. Travert, L. Lamant, O. Tournillac, T. Rousset, B. Fabiani, R. A. Cairns, T. Mak, C. Bastard, O. A. Bernard, L. de Leval & P. Gaulard (2012) Recurrent TET2 mutations in peripheral T-cell lymphomas correlate with TFH-like features and adverse clinical parameters. *Blood,* 120**,** 1466-9.

Lemonnier, F., E. Poullot, A. Dupuy, L. Couronné, N. Martin, L. Scourzic, V. Fataccioli, J. Bruneau, R. A. Cairns, T. W. Mak, O. A. Bernard, L. de Leval & P. Gaulard (2018) Loss of 5-hydroxymethylcytosine is a frequent event in peripheral T-cell lymphomas. *Haematologica,* 103**,** e115-e118.

Lieberman-Aiden, E., N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander & J. Dekker (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science,* 326**,** 289-93.

Losman, J. A. & W. G. Kaelin (2013) What a difference a hydroxyl makes: mutant IDH, (R)-2-hydroxyglutarate, and cancer. *Genes Dev,* 27**,** 836-52.

Lu, F., Y. Liu, L. Jiang, S. Yamaguchi & Y. Zhang (2014) Role of Tet proteins in enhancer activity and telomere elongation. *Genes Dev,* 28**,** 2103-19.

Marçais, A., L. Waast, J. Bruneau, K. Hanssens, V. Asnafi, P. Gaulard, F. Suarez, P. Dubreuil, A. Gessain, O. Hermine & C. Pique (2017) Adult T cell leukemia aggressivenness correlates with loss of both 5-hydroxymethylcytosine and TET2 expression. *Oncotarget,* 8**,** 52256-52268.

Odejide, O., O. Weigert, A. A. Lane, D. Toscano, M. A. Lunning, N. Kopp, S. Kim, D. van Bodegom, S. Bolla, J. H. Schatz, J. Teruya-Feldstein, E. Hochberg, A. Louissaint, D. Dorfman, K. Stevenson, S. J. Rodig, P. P. Piccaluga, E. Jacobsen, S. A. Pileri, N. L. Harris, S. Ferrero, G. Inghirami, S. M. Horwitz & D. M. Weinstock (2014) A targeted mutational landscape of angioimmunoblastic T-cell lymphoma. *Blood,* 123**,** 1293-6.

Okano, M., D. W. Bell, D. A. Haber & E. Li (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell,* 99**,** 247-57.

Pastor, W. A., L. Aravind & A. Rao (2013) TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat Rev Mol Cell Biol,* 14**,** 341-56.

Peric-Hupkes, D., W. Meuleman, L. Pagie, S. W. Bruggeman, I. Solovei, W. Brugman, S. Gräf, P. Flicek, R. M. Kerkhoven, M. van Lohuizen, M. Reinders, L. Wessels & B. van Steensel (2010) Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol Cell,* 38**,** 603-13.

Raffel, S., M. Falcone, N. Kneisel, J. Hansson, W. Wang, C. Lutz, L. Bullinger, G. Poschet, Y. Nonnenmacher, A. Barnert, C. Bahr, P. Zeisberger, A. Przybylla, M. Sohn, M. Tönjes, A.

Erez, L. Adler, P. Jensen, C. Scholl, S. Fröhling, S. Cocciardi, P. Wuchter, C. Thiede, A. Flörcken, J. Westermann, G. Ehninger, P. Lichter, K. Hiller, R. Hell, C. Herrmann, A. D. Ho, J. Krijgsveld, B. Radlwimmer & A. Trumpp (2017) BCAT1 restricts αKG levels in AML stem cells leading to IDHmut-like DNA hypermethylation. *Nature,* 551**,** 384-388.

Rasmussen, K. D., G. Jia, J. V. Johansen, M. T. Pedersen, N. Rapin, F. O. Bagger, B. T. Porse, O. A. Bernard, J. Christensen & K. Helin (2015) Loss of TET2 in hematopoietic cells leads to DNA hypermethylation of active enhancers and induction of leukemogenesis. *Genes Dev,* 29**,** 910-22.

Riggs, A. D. (1975) X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet,* 14**,** 9-25.

Schuster-Böckler, B. & B. Lehner (2012) Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature,* 488**,** 504-7.

Schübeler, D. (2015) Function and information content of DNA methylation. *Nature,* 517**,** 321-6.

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

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

van Steensel, B. & A. S. Belmont (2017) Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell,* 169**,** 780-791.

Walsh, C. P., J. R. Chaillet & T. H. Bestor (1998) Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet,* 20**,** 116-7.

Wu, D., D. Hu, H. Chen, G. Shi, I. S. Fetahu, F. Wu, K. Rabidou, R. Fang, L. Tan, S. Xu, H. Liu, C. Argueta, L. Zhang, F. Mao, G. Yan, J. Chen, Z. Dong, R. Lv, Y. Xu, M. Wang, Y. Ye, S. Zhang, D. Duquette, S. Geng, C. Yin, C. G. Lian, G. F. Murphy, G. K. Adler, R. Garg, L. Lynch, P. Yang, Y. Li, F. Lan, J. Fan, Y. Shi & Y. G. Shi (2018) Glucose-regulated phosphorylation of TET2 by AMPK reveals a pathway linking diabetes to cancer. *Nature,* 559**,** 637-641.

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

Xu, G. L., T. H. Bestor, D. Bourc'his, C. L. Hsieh, N. Tommerup, M. Bugge, M. Hulten, X. Qu, J. J. Russo & E. Viegas-Péquignot (1999) Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature,* 402**,** 187-91.

Yue, X., S. Trifari, T. Äijö, A. Tsagaratou, W. A. Pastor, J. A. Zepeda-Martínez, C. W. Lio, X. Li, Y. Huang, P. Vijayanand, H. Lähdesmäki & A. Rao (2016) Control of Foxp3 stability through modulation of TET activity. *J Exp Med,* 213**,** 377-97.

Zeller, P., J. Padeken, R. van Schendel, V. Kalck, M. Tijsterman & S. M. Gasser (2016) Histone H3K9 methylation is dispensable for Caenorhabditis elegans development but suppresses RNA:DNA hybrid-associated repeat instability. *Nat Genet,* 48**,** 1385-1395.

Zhou, W., H. Q. Dinh, Z. Ramjan, D. J. Weisenberger, C. M. Nicolet, H. Shen, P. W. Laird & B. P. Berman (2018) DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat Genet,* 50**,** 591-602.

Zhu, Q., N. Hoong, A. Aslanian, T. Hara, C. Benner, S. Heinz, K. H. Miga, E. Ke, S. Verma, J. Soroczynski, J. R. Yates, T. Hunter & I. M. Verma (2018) Heterochromatin-Encoded Satellite RNAs Induce Breast Cancer. *Mol Cell,* 70**,** 842-853.e7.

Chapter 2

Paradoxical association of TET loss of function with genome-wide DNA hypomethylation

Abstract

Cancer genomes are characterized by focal increases in DNA methylation, co-occurring with widespread hypomethylation. Here we show that TET loss-of-function results in a similar genomic footprint. Both 5hmC in wildtype genomes, and DNA hypermethylation in TET-deficient genomes, are largely confined to the active euchromatic compartment, consistent with the known functions of TET proteins in DNA demethylation and the known distribution of 5hmC at transcribed genes and active enhancers. In contrast, an unexpected DNA hypomethylation noted in multiple TET-deficient genomes is primarily observed in the heterochromatin compartment. In a mouse model of T cell lymphoma driven by TET deficiency (*Tet2/3 DKO* T cells), genomic analysis of malignant T cells revealed DNA hypomethylation in the heterochromatic genomic compartment. Moreover, hematopoietic stem/precursor cells (HSPC) doubly deficient for *Tet2* and *Dnmt3a* displayed greater losses of DNA methylation than HSPC singly deficient for *Tet2* or *Dnmt3a* alone, potentially explaining the unexpected synergy between *DNMT3A* and *TET2* mutations in myeloid and lymphoid malignancies. *Tet1*-deficient cells showed decreased localization of Dnmt3a in the heterochromatin compartment compared to WT cells, pointing to a functional interaction between TET and DNMT proteins and providing a potential explanation for the hypomethylation observed in *TET*-deficient genomes. Our data suggest that TET loss-of-function may at least partially underlie the characteristic pattern of global hypomethylation coupled to regional hypermethylation observed in diverse cancer genomes.

2.1 Introduction

TET enzymes are Fe(II) and α-ketoglutarate-dependent dioxygenases that mediate DNA demethylation through sequential oxidation of the methyl group of 5-methylcytosine (5mC) to 5-hydroxymethyl, 5-formyl and 5-carboxylcytosine (5hmC, 5fC and 5caC)(Tahiliani et al. 2009, Ito et al. 2011, He et al. 2011). The oxidized methylcytosines (oxi-mC) generated by TET proteins are intermediates in at least two pathways of DNA demethylation: (i) replication-dependent loss of methylation, reflecting inability of the DNMT1/UHRF1 complex to methylate unmodified CpGs on newly replicated DNA strands if an oxi-mC (rather than 5mC) is present on the template strand, and (ii) a replication-independent process in which thymine DNA glycosylase (TDG) excises 5fC and 5caC, which are then replaced with unmodified cytosine through base excision repair (Pastor, Aravind and Rao 2013, Wu and Zhang 2017).

Based on the known biochemical activities of TET-family proteins in oxidizing 5mC (Pastor et al. 2013), *TET* loss-of-function mutations are expected to result in gains of DNA methylation. In fact, increased methylation as a result of TET loss-of-function has been documented at many genomic regions including promoters, enhancers and CTCF sites (Cimmino et al. 2015, Rasmussen et al. 2015, Flavahan et al. 2016, Yue et al. 2016, Tsagaratou et al. 2017). Unexpectedly, however, a small number of studies have noted widespread decreases of DNA methylation in the genomes of TET-deficient cells, by mapping differentially methylated regions (DMRs) or by comparing methylation values at random genomic regions after whole-genome bisulfite sequencing (WGBS) (Hon et al. 2014, Lu et al. 2014, An et al. 2015, Tsagaratou et al. 2017). Because the hypomethylated regions observed in TET loss-of-function did not overlap with active or regulatory regions of the genome, these findings were largely ignored.

Cancer genomes are characterized by two opposing patterns of aberrant DNA methylation: focal or regional hypermethylation (Jones and Baylin 2002, Baylin and Jones 2016) and widespread DNA hypomethylation (Feinberg and Vogelstein 1983, Ehrlich 2009). DNA

hypermethylation at promoters and enhancers contributes to oncogenesis through transcriptional silencing of genes involved in DNA damage repair and tumor suppressors (Jones and Baylin 2002, Baylin and Jones 2016). Local hypermethylation has been shown to reflect the impaired expression or activity of TET proteins. Despite our understanding of the biochemical and biological consequences of local hypermethylation, however, the causes and consequences of DNA hypomethylation in cancer genomes are less well understood.

Principal component analysis of the interaction matrix obtained from Hi-C data has been used to compartmentalize the genome into an A compartment (positive PC1 values) and a B compartment (negative PC1 values) that exhibit the hallmark characteristics of euchromatin and heterochromatin, respectively (Lieberman-Aiden et al. 2009, van Steensel and Belmont 2017). The euchromatic A compartment is rich in expressed genes in the cell type under consideration, whereas the heterochromatic B compartment is gene-poor and bears epigenetically “repressive” chromatin marks, including H3K9me2/3 (van Steensel and Belmont 2017, Lieberman-Aiden et al. 2009). Moreover, the Hi-C B compartment overlaps with lamina-associated domains and corresponds to late-replicating regions of the genome, whereas the Hi-C A compartment corresponds to early replicating genomic regions and is not lamina-associated (Hiratani et al. 2008, Peric-Hupkes et al. 2010). Notably, the extended, partially methylated domains (PMDs) observed in cancer genomes overlap with Hi-C B compartment, late-replicating, nuclear lamina-associated domains (Berman et al. 2011, Hon et al. 2012, Zhou et al. 2018). In the remainder of this dissertation we will refer to the Hi-C A and B compartments as euchromatic and heterochromatic compartments, respectively.

Here we use a combination of Hi-C and WGBS data to document the DNA methylation changes associated with TET loss-of-function in diverse TET-deficient cell types. We show that 5hmC in wildtype genomes, and DNA hypermethylation in TET-deficient genomes, are largely confined to the euchromatic Hi-C A compartment. This finding is consistent with the known

functions of TET proteins in DNA demethylation and the known distribution of 5hmC at active enhancers and in the gene bodies of highly transcribed genes. In contrast, we show that the unexpected DNA hypomethylation noted in TET-deficient genomes is primarily present in the heterochromatic Hi-C B compartment. We also show that Dnmt3a relocalizes from the heterochromatic compartment to the euchromatic compartment in *Tet1*-deficient mESC, providing a potential mechanism for the heterochromatin hypomethylation observed in TET-deficient genomes. Our results are consistent with the co-occurrence of *DNMT3A* and *TET2* mutations in human cancers and the more pronounced leukemic phenotype observed in double *Tet2/Dnmt3a*-deficient mice compared to mice with individual disruption of *Tet2* or *Dnmt3a* alone. Taken together, these data point to a functional interaction between TET proteins and DNMTs.

2.2 Results

2.2.1 Widespread DNA hypomethylation in TET-deficient mouse ESC

To understand the impact of TET loss-of-function on genome-wide patterns of DNA modification, we re-analyzed data from several publicly available WGBS datasets across a diverse range of TET-deficient murine cell types: ESC (Hon et al. 2014, Lu et al. 2014, Yin et al. 2017), neural precursor cells differentiated from ESC (NPC) (Hon et al. 2014); pro-B cells (Lio et al. 2016); hematopoietic stem/precursor cells (HSPC) (Zhang et al. 2016) and a mouse model of TET-deficient T cells (Tsagaratou et al. 2017) (**Figs. 2.1-2.5**; and *see below*). Bisulfite sequencing estimates the sum of 5mC and 5hmC (Huang et al. 2010), but 5hmC is always <10% of 5mC in the control cell types considered here, and lower or absent in their TET-deficient counterparts. Thus throughout this dissertation, we refer to the values obtained from WGBS as “DNA methylation”.

We compared mouse embryonic stem cells (mESC) triply deficient in all three TET proteins (Tet1, Tet2 and Tet3; *Tet TKO* mESC) with mESC triply deficient in all three DNMTs

(Dnmt1, Dnmt3a and Dnmt3b; *Dnmt TKO* mESC) (Yin et al. 2017) (**Fig. 2.1A**). We plotted the distribution of average DNA methylation values in 1 and 10 kb windows containing at least 3 and at least 10 CpGs respectively (**Fig. 2.1A**). As expected (Li et al. 2015), *Dnmt TKO* mESC lost all DNA methylation (5mC+5hmC; **Fig. 2.1A**, *left and middle panels*). Unexpectedly, however, *Tet TKO* mESC also showed pronounced and widespread loss of methylation compared to WT ESC, with the shift in the distribution of window methylation percentages visible at both 1 kb and 10 kb resolution (**Fig. 2.1A**, *compare red and black traces*). Hypomethylation in *Tet TKO* mESC could also be visualized in dot plots comparing DNA methylation levels in WT and mutant ESC (**Fig. 2.2A**). Each dot shows DNA methylation levels in 1 kb windows with at least 3 CpGs: 73.7% of the 1 kb windows showed decreased methylation in *Tet TKO* mESC compared to WT, whereas only 26.1% showed increased methylation (**Fig. 2.2A**). Analysis of WGBS data from an independent study of *Tet TKO* mESC, generated using CRISPR/Cas9 technology (Lu et al. 2014), yielded very similar results (**Fig. 2.1B**, **Fig. 2.2B**). The differences in global hypomethylation observed between WT and *Tet TKO* mESC in the two studies may be due to different culture conditions or different lengths of time that the cells were maintained in culture (*compare* **Figs. 2.1A**, **B**).

2.2.2 DNA hypermethylation in euchromatin and hypomethylation in heterochromatin in diverse TET-deficient cells

In genome browser views, DNA methylation changes were most striking when viewed at megabase-scale resolution (**Fig. 2.3**), and were reminiscent of the large, partially-methylated domains noted in cancer genomes (Berman et al. 2011, Hon et al. 2012). Drawing on publicly available Hi-C (Bonev et al. 2017), replication timing (Hiratani et al. 2008), Lamina B association and H3K9me2 ChIP-seq data (Poleshko et al. 2017) on mESC, we found that regions that lost DNA methylation in all *Tet*-mutant mESC (**Fig. 2.3**, *tracks 6-9*) mostly overlapped with the

heterochromatic Hi-C B compartment (*track 1*), late replicating regions (*track 2*), lamina-associated domains (*track 3*), and regions marked by H3K9me2 (track 4). In contrast, H3K27me3 was present primarily in the euchromatic Hi-C A compartment (track 5). Likewise, Tet1 and Tet2 were primarily in the euchromatic compartment (tracks 10 and 11), consistent with the fact that 5hmC in WT mESC was predominantly located in the transcriptionally active, euchromatic Hi-C A compartment of the genome, as shown by two independent methods of mapping 5hmC: TAB-seq (Hon et al. 2014) and CMS-IP (Huang et al. 2014) (**Fig. 2.3**, *tracks 12 and 13*; the background signal for TAB-seq is noisier than for CMS-IP). The preferential presence of 5hmC in the euchromatic Hi-C A compartment was observed in all mouse cell types investigated, including ESC, HSC and pro-B cells (**Fig. 2.1C**; for similar data on mouse T cells please see **Fig. 2.6E**). The majority of hypermethylated genomic regions in TET-deficient cells are located in the euchromatic compartment; this finding, coupled with the presence of Tet1, Tet2 and 5hmC in the same euchromatic compartment in WT cells, confirms the by now well-established connection between DNA demethylation, TET-generated 5hmC, and 5hmC localization in actively transcribed genes and enhancers (Wu and Zhang 2017).

Dnmt TKO mESC showed extensive hypomethylation in both the Hi-C A and B compartments as expected (**Fig. 2.1A,** *right panels*), but the gains and losses of DNA methylation in *Tet TKO* mESC mostly occurred in the euchromatic and heterochromatic compartments respectively (**Fig. 2.1A, B**; **Fig. 2.2A, B**; **Fig. 2.3**, *tracks 6-7*). Even in singly-deficient *Tet1 KO* and *Tet2 KO* mESC (Hon et al. 2014), there is striking global DNA hypomethylation in both the euchromatic and heterochromatic compartments, which far exceeded the extent of the expected DNA hypermethylation (**Fig. 2.1D; Fig. 2.3**, *tracks 8-9*). Notably, hypermethylated genomic regions were clearly apparent in *Tet2 KO* mESC, but there was only marginal hypermethylation in *Tet1 KO* mESC (Hon et al. 2014) (**Fig. 2.1D**, **Fig. 2.2C**). Even when *Tet2 KO* mESC were

differentiated to neural precursor cells (NPC) for 3 days (Hon et al. 2014), the differentiated NPC showed heterochromatin hypomethylation (**Fig. 2.4A, B**).

To ask if our findings were generally applicable to other cell types, we integrated publicly available Hi-C datasets from mouse pro-B cells (Lin et al. 2012) with WGBS data from WT or *Tet2-/- Tet3fl/fl Mb1Cre* pro-B cells (Lio et al. 2016) (**Fig. 2.4C-E**). Deficiency of Tet2 and Tet3 in *Tet2-/- Tet3fl/fl Mb1Cre* mice results in defective immunoglobulin light chain rearrangement and a consequent block of B cell development at the pro-B to pre-B transition (Lio et al. 2016). Again, we observed concurrent increases and decreases of DNA methylation in pro-B cells genome-wide, with hypermethylation again occurring in the euchromatic compartment and hypomethylation in the heterochromatic compartment (**Fig. 2.4C-E**). Very similar results were obtained for mouse T cells and HSPC (*see below*, **Figs. 2.6, 2.8**).

2.2.3 Antigen-driven expansion in TET-deficient T cell leukemia/ lymphoma

To study hypomethylation induced by TET loss-of-function in the context of oncogenic transformation, we used a mouse model in which mice with profound TET loss-of-function (*Tet2-/- Tet3fl/fl CD4Cre (Tet2/3 DKO*) lacking Tet2 and Tet3 in T cells) rapidly developed an aggressive T cell leukemia/ lymphoma with 100% penetrance (Tsagaratou et al. 2017). The disease involves a normally minor subset of T cells (*i*NKT cells, hereafter referred to simply as NKT cells), which recognize lipid antigens presented on a non-classical major histocompatibility complex (MHC) protein known as CD1d (Bendelac, Savage and Teyton 2007, Engel and Kronenberg 2012). *Tet2/3 DKO* mice showed >10-fold expansion of NKT cells in the thymus as early as 20 days after birth and in the spleen by 3-4 weeks, and succumbed to an NKT cell leukemia starting at 5 weeks (Tsagaratou et al. 2017). Transfer of purified NKT cells from young mice, even into fully immunocompetent recipients, resulted in transfer of the leukemia, but transfer to recipient mice lacking CD1d, the MHC protein that presents lipid antigen to NKT cells (Bendelac et al. 2007),

resulted in minimal expansion (Tsagaratou et al. 2017) (**Suppl. Fig. 2.5**), indicating that the leukemia was driven by NKT cell expansion arising from presentation of lipid antigens by CD1d.

2.2.4 TET-deficient T cell lymphomas show euchromatic compartment hypermethylation and heterochromatic compartment hypomethylation

To define DNA modification patterns in the active and inactive genome compartments of TET-deficient NKT cell lymphomas, we generated and analyzed WGBS and Hi-C data for WT, young and transferred/expanded *Tet2/3 DKO* NKT cells. Like TET-deficient mouse ESC and pro-B cells (**Figs. 2.1-2.4**), both young and transferred/expanded *Tet2/3 DKO* NKT cells showed increased methylation in the euchromatic compartment and decreased methylation in the heterochromatic compartment relative to WT NKT cells (**Fig. 2.6A-C; Figs. 2.7A, B**). Moreover, as in the cell types considered above, the euchromatic compartment was gene-rich and contained the majority of expressed genes, accessible chromatin regions and 5hmC (**Figs. 2.6D, E**; **Fig. 2.6F**, *tracks 9-11*; **Fig. 2.7C**). In genome browser views, there were only minor changes in euchromatic and heterochromatic compartments between WT thymus and spleen (**Fig. 2.6F**, *tracks 1, 2*; **Fig. 2.7D**, *top left panel*), but somewhat greater differences between WT and expanded *Tet2/3 DKO* splenic NKT cells (**Fig. 2.6F**, *tracks 2, 3*; **Fig. 2.7D**, *bottom panels*). Windows that were less accessible in *Tet2/3 DKO* NKT cells compared to WT were primarily in the euchromatic compartment; the few windows that were more accessible in *Tet2/3 DKO* cells compared to WT were present in the heterochromatic compartment (**Fig. 2.6D**). Finally, both young and transferred/expanded *Tet2/3 DKO* NKT cells showed extended regions of increased and decreased methylation compared to WT NKT cells, and these largely coincided with the euchromatic and heterochromatic compartments respectively (**Fig. 2.6F**, *tracks 4-8*).

Overall, the data on TET-deficient NKT cell lymphomas were completely concordant with those for the other TET-deficient cell types considered above. Regardless of cell type, TET

deficiency was broadly associated with DNA hypermethylation in the euchromatic compartment, concurrently with DNA hypomethylation in the heterochromatic compartment. In Chapter 3 of this dissertation, we examine other genomic features of *Tet2/3 DKO* NKT cells reported to be associated with DNA hypomethylation, including reactivation of repeat elements and increased mutational load.

2.2.5 Paradoxical increase in heterochromatic DNA hypomethylation in HSPC from *Dnmt3a-Tet2 DKO* mice

DNMT3A and *TET2* mutations are frequently observed, both individually and together, in diverse myeloid and lymphoid malignancies including MDS, AML and PTCL (Couronné, Bastard and Bernard 2012, Ley et al. 2013, Odejide et al. 2014, Palomero et al. 2014, Sakata-Yanagimoto et al. 2014, Papaemmanuil et al. 2016). Based on the biochemical activities of the encoded proteins, we originally expected opposing effects of *DNMT3A* and *TET2* loss-of-function, both in terms of biological outcomes and in genome-wide analyses. Specifically, we expected first that the disease phenotypes of *Dnmt3a* and *Tet2*-deficient mice might be ameliorated in mice deficient for both enzymes, and second that there would be widespread losses of DNA methylation in the case of *Dnmt3a* mutations but widespread gains of DNA methylation in the case of *Tet2* mutations. However, a previous collaborative study (Zhang et al. 2016) unexpectedly showed that mice with dual *Tet2* and *Dnmt3a* deficiency in hematopoietic stem/precursor cells (HSPC) displayed more severe phenotypes than mice with individual *Tet2* or *Dnmt3a* deletions alone.

Given these data, we postulated that one explanation for the paradoxical synergy between *Dnmt3a* and *Tet2* mutations might be heterochromatic hypomethylation in both types of mutant cells. To test this hypothesis, we used publicly available Hi-C data on WT HSPC (Hu et al. 2018) and previously published WGBS data for WT, *Tet2 KO*, *Dnmt3a KO* and *Dnmt3a/Tet2 DKO* HSPC (Jeong et al. 2014, Zhang et al. 2016) to localize DNA methylation changes to the two genomic

compartments defined by Hi-C (**Fig. 2.8A**; **Fig. 2.8B**, *top track*). Indeed, both *Tet2* and *Dnmt3a* deficiency were characterized by widespread losses of DNA methylation in HSPC; moreover, HSPC from doubly *Tet2*/*Dnmt3a*-deficient mice showed greater hypomethylation than HSPC with either mutation alone (**Fig. 2.8**). The synergistic loss of DNA methylation was striking when 10 kb windows with at least 10 CpGs were considered (**Fig. 2.8C, 2.8D**, *right panels*), although the small fraction of fully methylated regions as well as completely or almost completely demethylated regions (e.g. CpG islands) were best observed in 1 kb windows with at least 3 CpGs (**Fig. 2.8C, 2.8D**, *left panels*). Specifically, *Dnmt3a*-deficient (*Dnmt3a KO*) HSPC (*blue traces*) displayed greater loss of methylation compared to *Tet2*-deficient (*Tet2 KO*) HSPC (*green traces*) as expected, but HSPC doubly deficient in *Tet2* and *Dnmt3a* (*Dnmt3a/Tet2 DKO*) showed even greater loss of methylation compared to HSPC singly deficient in either enzyme alone (**Fig. 2.8C, 2.8D**, *red traces*).

In genome browser views, extended domains of hypomethylation were observed in *Tet2 KO* HSC (**Fig. 2.8B**, *track 2*), as in all the other cell types considered above (**Figs. 2.1-2.4, 2.6**). These partially methylated domains mostly overlapped with the heterochromatic compartment, whereas regions in the euchromatic compartment showed a slight gain of methylation (**Fig. 2.8A**, *top panel*; **Fig. 2.8B**, *track 2*). In contrast, in both the *Dnmt3a KO* and *Dnmt3a/Tet2 DKO* HSPCs (**Fig. 2.8A**, *middle and bottom panels*; **Fig. 2.8B**, *tracks 3, 4*), we observed widespread DNA hypomethylation in both the euchromatic and heterochromatic compartments, as expected from the loss of Dnmt3a activity. Thus, both *Tet2* and *Dnmt3a* mutations result in widespread hypomethylation in heterochromatic regions of the genome.

2.2.6 Decreased localization of Dnmt3a in the heterochromatin compartment of TET-deficient ESC

We investigated the mechanisms underlying the loss of methylation in heterochromatin in TET-deficient cells. One potential mechanism stems from the observation that similar hypomethylated domains are observed in rapidly proliferating cells (Zhou et al. 2018). This argument is plausible for the NKT cell lymphoma in which TET deficiency accounts for rapid proliferation, but cannot apply in the case of ESC. PMDs are not observed in WT ESC and induced pluripotent stem cells (iPSC) despite the high proliferation rates of these cells (Lister et al. 2009, Lister et al. 2011); moreover, *Tet1/2/3 TKO* mESC do not proliferate faster than their WT counterparts (Fig. S6B from (Li et al. 2016) reproduced here in **Fig. 2.9A** for the reader's convenience), but clearly show decreased DNA methylation at late-replicating, lamina-associated domains (**Fig. 2.3**). Similarly, 5hmC levels – implying ongoing TET activity – are highest in non-proliferating neurons compared to other cells (Kriaucionis and Heintz 2009), but PMDs in the CpG context do not occur in the brain (Lister et al. 2013). Finally, senescent cells that have stopped proliferating also show partially-methylated domains, with loss of methylation occurring predominantly in the heterochromatic compartment (Cruickshanks et al. 2013).

Since *TET*-deficient ESC showed heterochromatin hypomethylation without increased proliferation, we asked whether hypomethylation in ESC could be attributed to alterations of DNMT localization or function. We focused on the de novo DNMTs, Dnmt3a and Dnmt3b, because of a previous report in which Dnmt3a and Dnmt3b were both shown by immunocytochemistry to co-localize with HP-1a in the heterochromatic compartment of mouse 3T3 cells and ES cells (Bachman, Rountree and Baylin 2001). To infer the contribution of each DNMT to methylation in the euchromatic and heterochromatic compartments, we reanalyzed a dataset in which Dnmt3b1 and the two splice variants Dnmt3a1 and Dnmt3a2 were reconstituted in ES cells lacking all DNMTs (Baubec et al. 2015, Manzo et al. 2017). Mapping of these three Dnmt3 proteins in WT ESC showed that all three were primarily present in the euchromatic compartment but were also significantly represented in the heterochromatic compartment (**Fig.**

2.10A; Fig. 2.10B, *tracks 1-4*). WGBS performed on the DNMT-reconstituted cells showed that all three Dnmt3 proteins contributed to methylation in both the euchromatic and heterochromatic compartments (**Fig. 2.10B**, *tracks 5-8*; **Fig. 2.10C**). Notably, the major contribution to DNA methylation in the heterochromatic compartment was from Dnmt3a1 (**Fig. 2.10C**).

Heterochromatic hypomethylation in *TET*-deficient cells could reflect either altered distribution or function of Dnmt3a1 (these two scenarios are not mutually exclusive). To examine alterations in Dnmt3a localization, we used a dataset from a study in which Dnmt3a1 tagged with the biotin acceptor peptide for *E. coli* BirA was expressed in WT and *Tet1*-deficient mouse ESC (Gu et al. 2018). The data show unambiguously that compared to WT ESC, Dnmt3a1 was enriched in the euchromatic compartment and depleted from the heterochromatic compartment in *Tet1*-deficient ESC (**Fig. 2.11A**). A genome browser view illustrating the relocalization is shown in **Fig. 2.11B**; *tracks 2, 3* show the normalized ChIP-seq coverage and *track 4* shows the difference in Dnmt3a1 binding in WT vs Tet1-deficient ESC. Together these data indicate that Dnmt3a1 relocalizes from the heterochromatic to the euchromatic compartment in *Tet1*-deficient ESC.

We used the same datasets described above (Gu et al. 2018, Manzo et al. 2017, Baubec et al. 2015, Williams et al. 2011) to determine how Dnmt3a1 relocalized within the euchromatin compartment in Tet1-deficient ESC. A zoomed-in view of the *Lefty1/2* locus within the euchromatic compartment revealed strong mutually exclusive localization of TET1/2 and DNMT3A (**Fig 2.11C**). A contour plot illustrating this mutually exclusive localization is shown in **Fig. 2.11D**; Q1 contains the regions in the euchromatic Hi-C A compartment occupied by Dnmt3a1 but not by Tet1, whereas Q3 contains the regions occupied by Tet1 but not by Dnmt3a1. In *Tet1*-deficient ESC, this distribution is altered: the histograms for Q1 show that binding of Dnmt3a1 to its exclusive sites in WT ESC is substantially decreased in *Tet1 KO* ESC, whereas the histograms for Q3 show that Dnmt3a1 shows increased binding in *Tet1 KO* compared to WT

ESC (**Fig. 2.11E**). This mutually exclusive localization seems to be a general feature of the relation between TET proteins and DNMT3s (**Fig. 2.9B**, **C**), as previously noted for Tet1 and Dnmt3a1 by (Gu et al. 2018). Together, the data indicate that loss of TET1 (almost exclusively from the euchromatic compartment) in mouse ES cells results in relocalization of DNMT3A1, on the one hand from the heterochromatic to the euchromatic compartment and on the other hand to regions within euchromatin that were previously occupied by TET1 (**Fig. 2.11F)**. It is plausible that this relocalization contributes at least partly to the paradoxical loss of DNA methylation in the heterochromatic compartment that we observe in *TET*-deficient cells, as well as to the hypermethylation observed in the euchromatic compartment of the same cells.

2.3 Discussion

In this study, we document an unexpected similarity between the DNA methylation patterns of diverse TET-deficient cell types and those of cancer genomes. Cancer genomes show local hypermethylation combined with widespread hypomethylation (Jones and Baylin 2002, Baylin and Jones 2016), and we reproducibly observed both features in TET-deficient cells. As expected from the biochemical activities of TET enzymes in generating oxi-mC bases and their involvement in DNA demethylation (Pastor et al. 2013, Wu and Zhang 2017), local DNA hypermethylation was consistently observed in the euchromatic (Hi-C A) compartment of TET-deficient cells; this compartment contains the vast majority of 5hmC, a stable modification that is most highly enriched in the gene bodies of the most highly expressed genes and at the most active enhancers (Tsagaratou et al. 2014, Hon et al. 2014). Unexpectedly, however, we also observed large domains of DNA hypomethylation in the heterochromatic (Hi-C B) compartment of diverse TET-deficient cell types, including ESC, NPC, HSPC, T cells and pro-B cells. The existence of these hypomethylated domains cannot be explained by our current understanding of TET enzymatic activity, but their presence in TET-deficient cells suggests strongly that TET

proteins are required, directly or indirectly, for optimal DNMT-mediated DNA methylation in heterochromatin.

Our reanalysis of published data suggests a potential mechanism for the synergistic actions of DNMT3A and TET proteins. TET1 and DNMT3A occupy mutually exclusive locations in the euchromatic compartment of mouse embryonic stem cells, and loss of TET proteins from euchromatin results in relocalization of DNMT3A1 to regions previously occupied by TET1 (see model in **Fig. 2.11E**). Broadly, this observation suggests that the DNMT3 enzymes responsible for de novo methylation are recruited to euchromatin through a scaffold complex or other recruitment mechanism in common with TET enzymes, but for which the DNMTs have lower affinities than TETs under normal physiological conditions. Assuming that the DNMT3 enzymes are present at limiting concentrations, loss of TET proteins would cause them to relocalize away from heterochromatin and into euchromatic regions, resulting in the observed dual loss of DNA methylation in heterochromatin and increased DNA methylation in euchromatin. This observation is consistent with the notable finding that every animal genome that encodes a DNMT also harbors at least one functional TET/JBP protein (Iyer et al. 2014). Further studies in specific cancer models will be required to advance our global understanding of the biochemistry underlying the functional interactions between TETs and DNMTs.

2.4 Figures

Figure 2.1. Widespread DNA hypomethylation in *TET*-deficient mouse ESC. (*A*) Density distribution (*left panels*) and cumulative distribution (*middle panels*) of average DNA methylation values within 1 kb and 10 kb windows across the genome in wild type (WT), *Dnmt TKO*, and *Tet TKO* mouse embryonic stem cells (mESC) (data from (Yin et al. 2017)). Only windows with 3 or more CpGs per 1 kb, or 10 or more CpGs per 10 kb, covered by at least 5 WGBS reads per CpG, were considered for the analysis. Two-sample Kolmogorov-Smirnov test comparing the distributions of WT and *Tet TKO* was used to calculate the D_{ks} statistic and *p*-value. (*Right panels*) Correlation between DNA methylation changes (difference in cytosine modification percentage, mutant minus wild type) and euchromatin/heterochromatin compartments (positive versus negative Hi-C PC1 values), in *Tet TKO* (*top*), and *Dnmt TKO* (*bottom*) mESC. Spearman correlation coefficient is shown (r_s value). (*B*), Density distribution (*left panel*) and cumulative distribution (*middle panel*) of average DNA methylation values within 10 kb windows across the genome in wild type (WT) and *Tet TKO* mouse embryonic stem cells (mESC) (data from (Lu et al. 2014)). (*Right panel*) Correlation between DNA methylation changes and euchromatin/heterochromatin compartments. Spearman correlation coefficient is shown (r_s value). (*C*), Relationship between 5hmC distribution (CMS-IP) and euchromatin/heterochromatin compartments (Hi-C PC1 values) in WT mESC, HSC and pro-B cells. CMS-IP enrichment was calculated for 1 kb windows. (*D*), Density distribution (*left panel*) and cumulative distribution (*middle panel*) of average DNA methylation values within 10 kb windows across the genome, in wild type (WT), *Tet1 KO, and Tet2 KO* mESC (data from (Hon et al. 2014)). (*Right panel*) Correlation between DNA methylation changes and euchromatin/heterochromatin compartments. Spearman correlation coefficient is shown (r_s value).

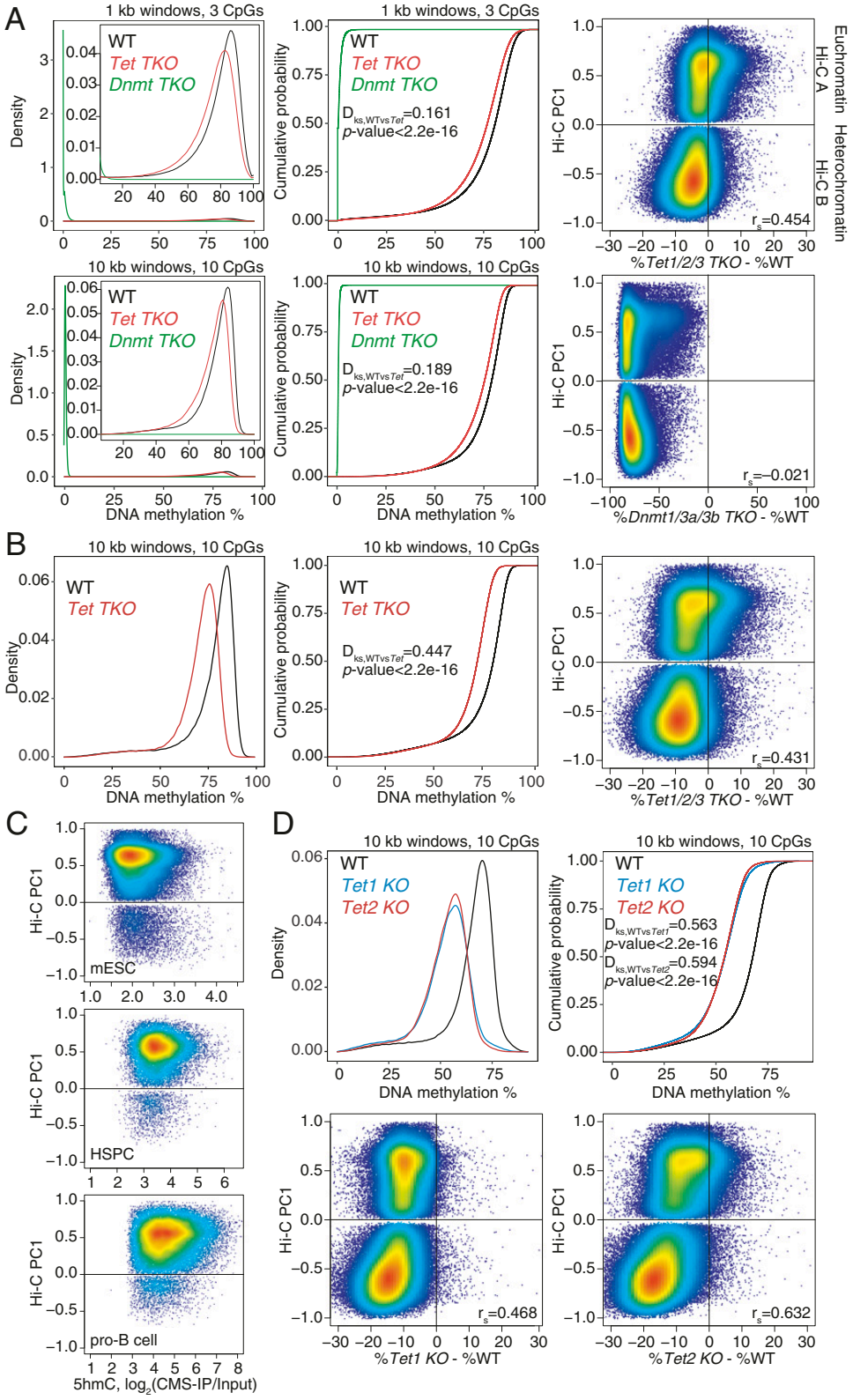

A
1 kb windows, 3 CpGs
WT
Tet TKO
Dnmt TKO
Density
Cumulative probability
D ks,WTvsTet=0.161
p-value<2.2e-16
Hi-C PC1
rs=0.454
%Tet1/2/3 TKO - %WT
Euchromatin
Hi-C A
Heterochromatin
Hi-C B
10 kb windows, 10 CpGs
D ks,WTvsTet=0.189
p-value<2.2e-16
rs=−0.021
%Dnmt1/3a/3b TKO - %WT
DNA methylation %
B
D ks,WTvsTet=0.447
p-value<2.2e-16
rs=0.431
C
mESC
HSPC
pro-B cell
5hmC, log2(CMS-IP/Input)
D
Tet1 KO
Tet2 KO
D ks,WTvsTet1=0.563
p-value<2.2e-16
D ks,WTvsTet2=0.594
p-value<2.2e-16
rs=0.468
%Tet1 KO - %WT
rs=0.632
%Tet2 KO - %WT

Figure 2.2. B compartment hypomethylation in *TET*-mutant mESC. (*A*), Smoothed scatterplot of the average DNA methylation values within 1 kb windows across the genome, comparing WT cytosine modification values (x-axis) to the ones in *Tet TKO* (*top*), and *Dnmt TKO* (*bottom*) (y-axis). LOESS regression (dashed line) is displayed for each panel. (*B*), Smoothed scatterplot of the average DNA methylation values within 1 kb windows across the genome, comparing WT cytosine modification values (x-axis) to the ones in *Tet TKO* (y-axis). LOESS regression (dashed line) is displayed for each panel. (*C*), Smoothed scatterplot of the average DNA methylation within 1 kb windows across the genome, comparing cytosine modification values between WT (x-axis) and *Tet1 KO* or *Tet2 KO* (y-axis) mESC. LOESS regression (dashed line) is displayed for each panel. *Tet1 KO* mESC (*top*) show global loss of methylation whereas *Tet2 KO* mESC (*bottom*) show hypermethylation in regions with low-intermediate methylation in WT mESC and hypomethylation at regions with high methylation in WT mESC.

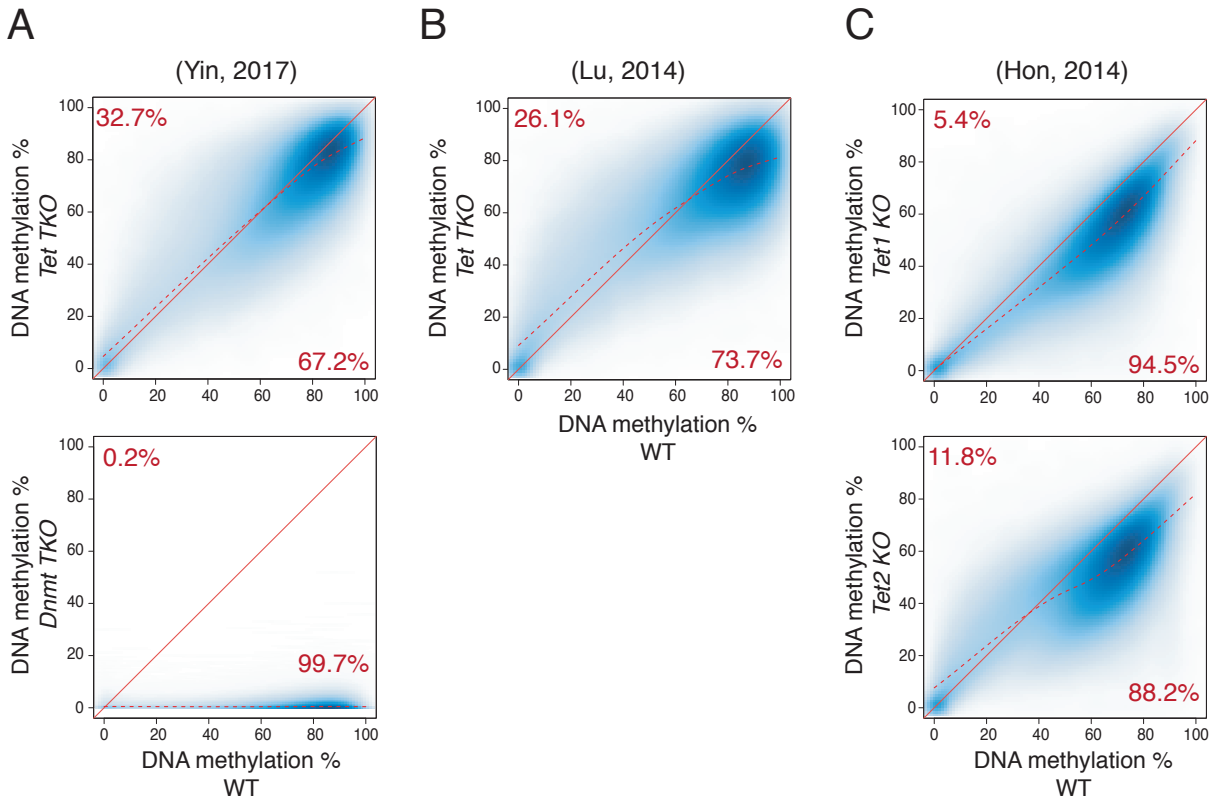

A
(Yin, 2017)
DNA methylation %
Tet TKO
32.7%
67.2%
DNA methylation %
Dnmt TKO
0.2%
99.7%
DNA methylation %
WT
B
(Lu, 2014)
DNA methylation %
Tet TKO
26.1%
73.7%
DNA methylation %
WT
C
(Hon, 2014)
DNA methylation %
Tet1 KO
5.4%
94.5%
DNA methylation %
Tet2 KO
11.8%
88.2%
DNA methylation %
WT

Figure 2.3. DNA hypomethylation in heterochromatin compartment in *TET*-deficient mouse ESC. Genome tracks showing an overlap between Hi-C-defined A/B compartments (*track 1*), early/late replicating sites (early replicating: positive values; *track 2*), lamina-associated domains (*track 3*), regions marked by histone marks H3K9me2 (*track 4*) and H3K27me3 (*track 5*), and large hypomethylated domains in different *Tet KO* mESC (shown as subtraction of DNA methylation percentage, mutant minus wild type; *tracks 6-9*). TET1 and TET2 binding (ChIP-seq; *tracks 10-11*), 5hmC distribution (TAB-seq and CMS-IP; *tracks 12-13*) and gene expression (RNA-seq; *tracks 14*) in WT mESC are shown for reference. Heterochromatic Hi-C B compartment regions are highlighted.

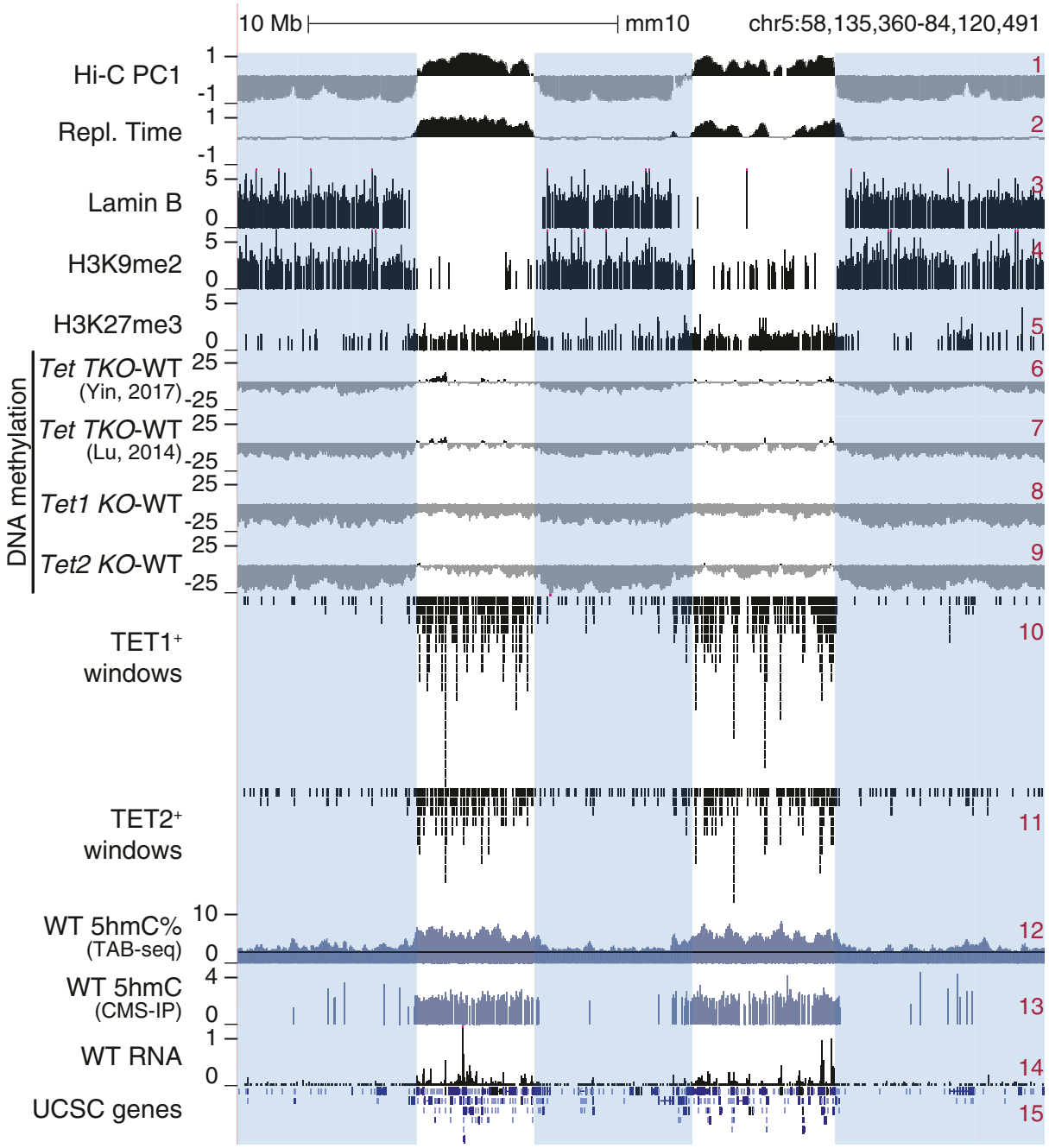

10 Mb
mm10
chr5:58,135,360-84,120,491
Hi-C PC1
Repl. Time
Lamin B
H3K9me2
H3K27me3
DNA methylation
Tet TKO-WT (Yin, 2017)
Tet TKO-WT (Lu, 2014)
Tet1 KO-WT
Tet2 KO-WT
TET1+ windows
TET2+ windows
WT 5hmC% (TAB-seq)
WT 5hmC (CMS-IP)
WT RNA
UCSC genes

Figure 2.4. B compartment hypomethylation in *TET*-mutant NPC and pro-B cells. (*A*), Distribution of average DNA methylation values within 1 kb and 10 kb windows across the genome in wild type (WT) and *Tet2 KO* mESC differentiated to neural precursor cells (NPC) at day 3 post-differentiation. (*B*), Correlation between DNA methylation changes (difference in cytosine modification percentage, *Tet2 KO* minus WT) and A/B compartments (Hi-C PC1 values) in mESC differentiated to NPCs. The Spearman correlation coefficient is shown (r_s value). (*C*), Distribution of average DNA methylation values within 1 kb and 10 kb windows across the genome, in wild type (WT) and *Tet2/3 DKO* pro-B cells. (*D*), Correlation between DNA methylation changes (difference in cytosine modification percentage, *Tet2/3 DKO* minus WT) and A/B compartments (Hi-C PC1 values) in mouse pro-B cells. The Spearman correlation coefficient is shown (r_s value). (*E*), Genome tracks showing a correspondence between heterochromatic Hi-C B compartment and large hypomethylated domains in pro-B cells (shown as subtraction of DNA methylation percentage, *Tet2/3 DKO* minus WT). RNA transcription (GRO-seq track) and 5hmC distribution (CMS-IP track) in WT pro-B cells are shown for reference.

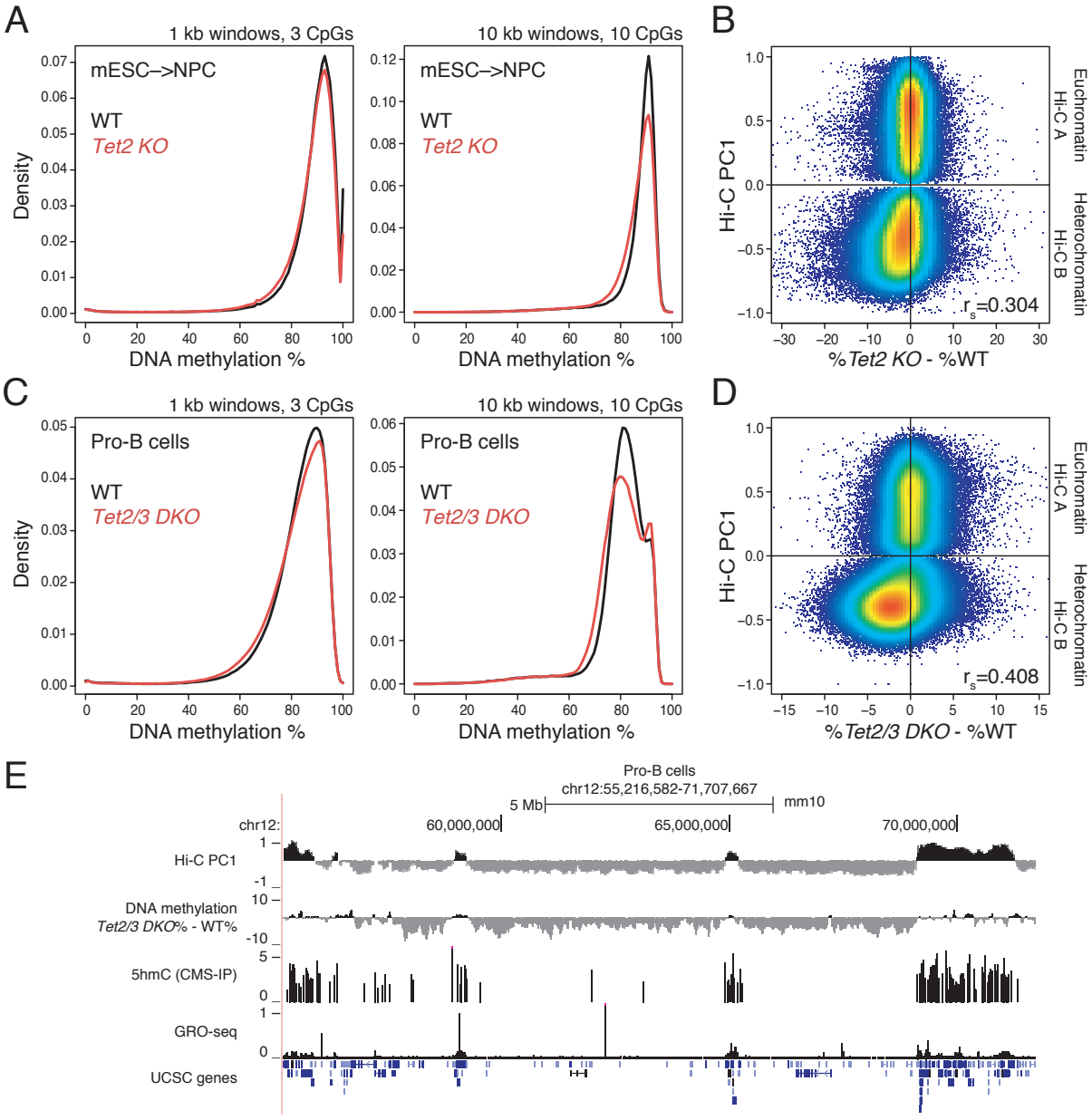

A
1 kb windows, 3 CpGs
10 kb windows, 10 CpGs
mESC–>NPC
WT
Tet2 KO
Density
DNA methylation %
B
Hi-C PC1
Euchromatin
Hi-C A
Heterochromatin
Hi-C B
r_s=0.304
%Tet2 KO - %WT
C
Pro-B cells
WT
Tet2/3 DKO
D
r_s=0.408
%Tet2/3 DKO - %WT
E
Pro-B cells
chr12:55,216,582-71,707,667
5 Mb
mm10
chr12:
60,000,000
65,000,000
70,000,000
Hi-C PC1
DNA methylation
Tet2/3 DKO% - WT%
5hmC (CMS-IP)
GRO-seq
UCSC genes

Figure 2.5. Expansion of *Tet2/3 DKO* NKT cells. (*A*), *Left,* experimental workflow. *Middle,* the picture depicts splenomegaly and enlarged lymph nodes in wild type (WT) but not CD1d knockout recipients of *Tet2/3 DKO* NKT cells. *Right*, times of disease emergence.

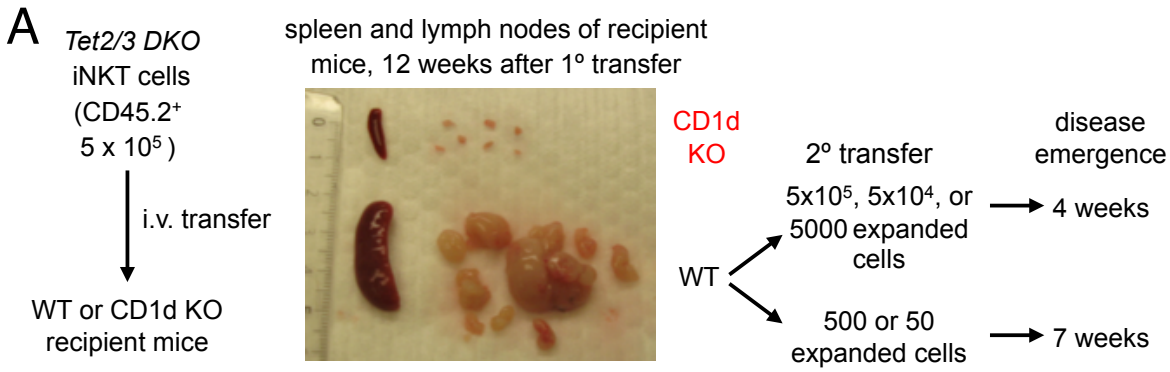

A
Tet2/3 DKO
iNKT cells
(CD45.2+
5 x 10^5)
i.v. transfer
WT or CD1d KO
recipient mice
spleen and lymph nodes of recipient
mice, 12 weeks after 1º transfer
CD1d
KO
WT
2º transfer
5x10^5, 5x10^4, or
5000 expanded
cells
500 or 50
expanded cells
disease
emergence
4 weeks
7 weeks

Figure 2.6. DNA hypermethylation in euchromatin and hypomethylation in heterochromatin in NKT cell lymphoma from *Tet2/3 DKO* mice. (*A, B*) Cumulative (*A*) and density distribution (*B*) of average DNA methylation values within 1 kb and 10 kb windows across the genome, in wild type (WT), young, and transferred and expanded *Tet2/3 DKO* NKT cells. In (*A*), two-sample Kolmogorov-Smirnov test comparing the distributions of WT and *Tet2/3 DKO* (young or expanded, as indicated) was used to calculate the D_{ks} statistic and *p*-value. (*C*), Correlation between DNA methylation changes (mutant minus wild type) and euchromatin/heterochromatin compartments in *Tet2/3 DKO* NKT cells at two stages of expansion (young, and transferred and expanded). Spearman correlation coefficient is shown (r_s value). (*D*), Correlation between changes in chromatin accessibility (ATAC-seq) in WT versus *Tet2/3 DKO* cells, and euchromatin/heterochromatin compartments, in thymic NKT cells from young mice. Accessibility differences were calculated at 1 kb resolution. (*E*), Correlation between 5hmC distribution (CMS-IP) and euchromatin/heterochromatin compartments in WT thymic NKT cells. CMS-IP enrichment was calculated for 1 kb windows. (*F*), Genome tracks showing a correspondence between Hi-C-defined B compartments in thymic and splenic WT NKT cells (*tracks 1-2*) and expanded splenic *Tet2/3 DKO* NKT cells (*track 3*), and large hypomethylated domains in WT, young and expanded *Tet2/3 DKO* NKT cells (*tracks 4-8*). Notice progressive hypomethylation of the heterochromatic compartment. 5hmC distribution (CMS-IP, *track 9*), chromatin accessibility (ATAC-seq, *track 10*), and gene expression (RNA-seq, *track 11*) in WT NKT cells are shown to contextualize methylation changes. The heterochromatic regions that lose DNA methylation in *Tet2/3 DKO* T cells are highlighted.

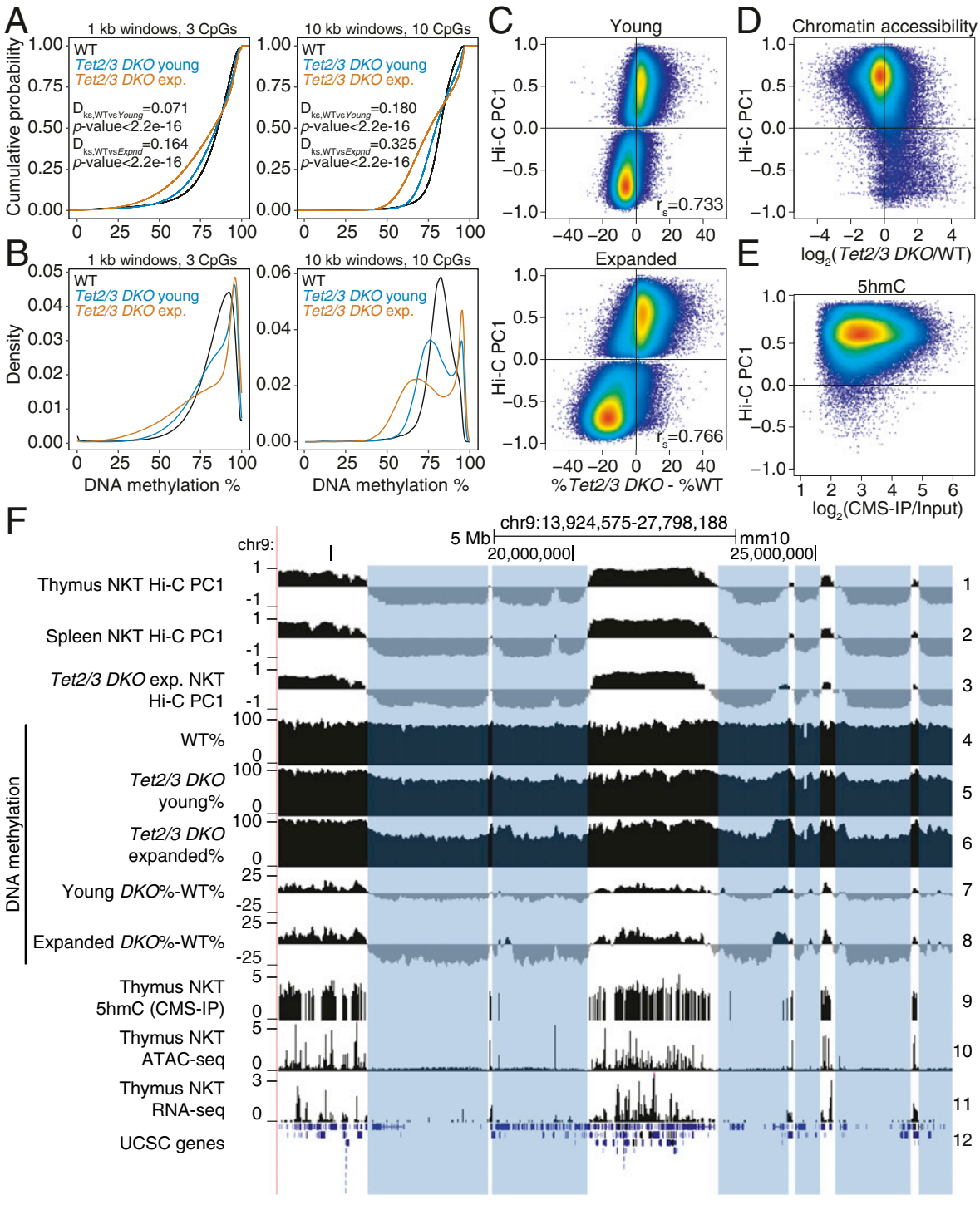

A
1 kb windows, 3 CpGs
10 kb windows, 10 CpGs
Cumulative probability
WT
Tet2/3 DKO young
Tet2/3 DKO exp.
D ks,WTvsYoung =0.071
p-value<2.2e-16
D ks,WTvsExpnd =0.164
p-value<2.2e-16
D ks,WTvsYoung =0.180
p-value<2.2e-16
D ks,WTvsExpnd =0.325
p-value<2.2e-16
B
Density
DNA methylation %
C
Young
Expanded
Hi-C PC1
rs=0.733
rs=0.766
%Tet2/3 DKO - %WT
D
Chromatin accessibility
log2(Tet2/3 DKO/WT)
E
5hmC
log2(CMS-IP/Input)
F
chr9:13,924,575-27,798,188
chr9:
5 Mb
mm10
20,000,000
25,000,000
Thymus NKT Hi-C PC1
Spleen NKT Hi-C PC1
Tet2/3 DKO exp. NKT Hi-C PC1
DNA methylation
WT%
Tet2/3 DKO young%
Tet2/3 DKO expanded%
Young DKO%-WT%
Expanded DKO%-WT%
Thymus NKT 5hmC (CMS-IP)
Thymus NKT ATAC-seq
Thymus NKT RNA-seq
UCSC genes

Figure 2.7. *Tet2/3*-deficient NKT cell lymphoma displays progressive hypomethylation in the heterochromatic Hi-C B compartment. (*A*), Smoothed scatterplot of the average DNA methylation within 1 kb windows across the genome, comparing WT (x-axis) to young (*left panel*) or transferred and expanded (*right panel*) *Tet2/3 DKO* NKT cells (y-axis). LOESS regression (dashed line) is displayed for each panel. (*B*), Distribution of DNA methylation (WGBS signal) at cytosines within the CpG context covered by at least 5 WGBS reads. DNA methylation values are shown for WT, *Tet2/3 DKO* young (Yng), and expanded (Exp) *Tet2/3 DKO* NKT cells. (*C*), Comparison of gene expression levels (RNA-seq, $\log_2$(TPM+1)) in the euchromatic Hi-C A and heterochromatic Hi-C B compartments in young (*left panel*) and expanded (*right panel*) *Tet2/3 DKO* NKT cells. (*D*), Pairwise comparison of Hi-C PC1 values between independent samples. Note the good correlation between the WT thymus and spleen NKT cell samples and between *Tet2/3 DKO* replicates 1 and 2 which were expanded from the same donor mouse, compared to the slightly greater differences between *Tet2/3 DKO* replicates 1 and 2 and replicate 3 which was from a different donor mouse. Similarly, there are slight differences in the Hi-C compartment between NKT cells from WT spleen and all three *Tet2/3 DKO* NKT cells taken from the spleen.

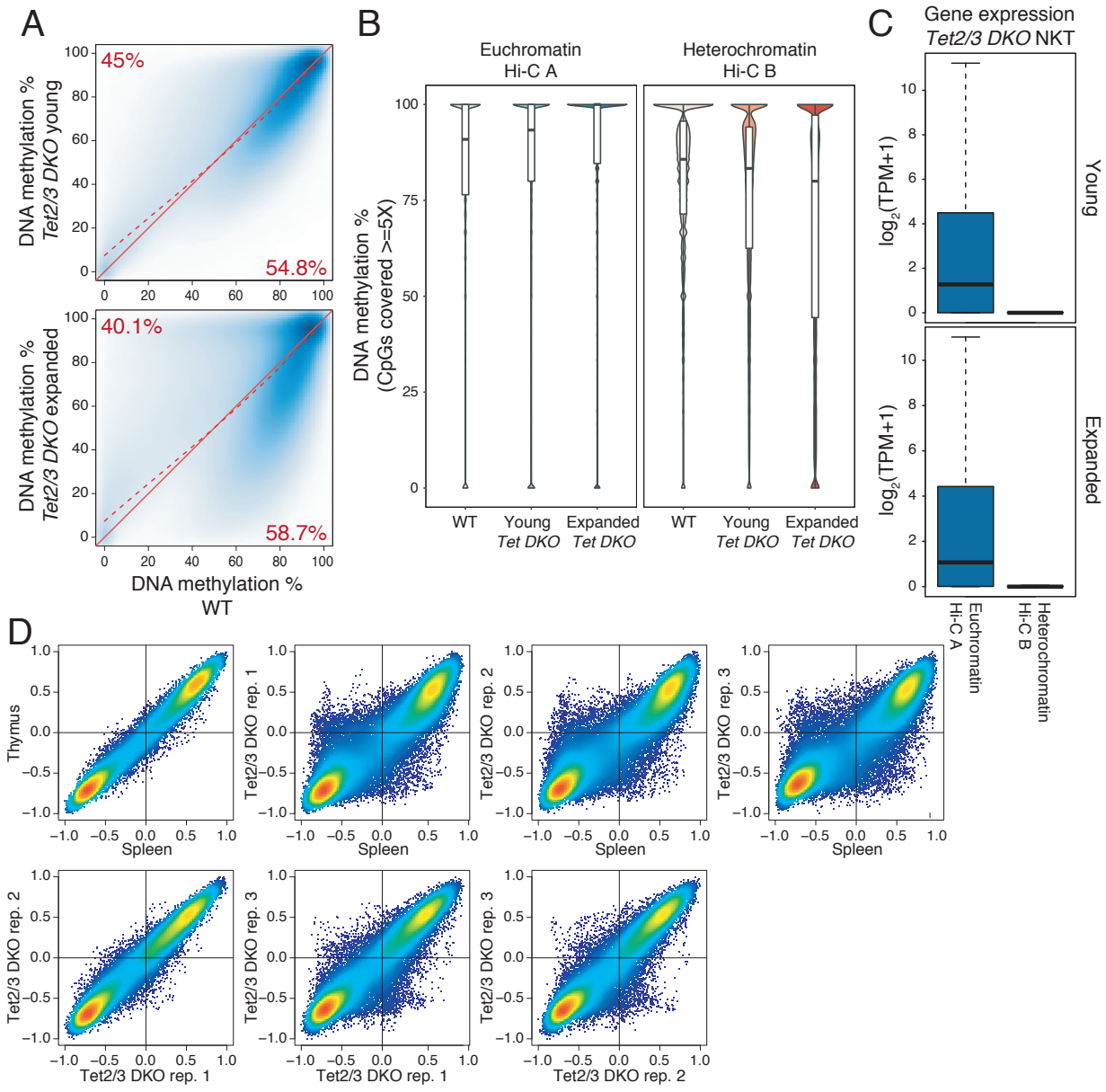

A
45%
54.8%
40.1%
58.7%
DNA methylation %
Tet2/3 DKO young
DNA methylation %
Tet2/3 DKO expanded
DNA methylation %
WT
B
Euchromatin
Hi-C A
Heterochromatin
Hi-C B
DNA methylation %
(CpGs covered >=5X)
WT
Young
Tet DKO
Expanded
Tet DKO
C
Gene expression
Tet2/3 DKO NKT
log2(TPM+1)
Young
Expanded
Euchromatin
Hi-C A
Heterochromatin
Hi-C B
D
Thymus
Spleen
Tet2/3 DKO rep. 1
Tet2/3 DKO rep. 2
Tet2/3 DKO rep. 3

Figure 2.8. Increased DNA hypomethylation in double *DNMT3A* and *TET2* mutant mice. (*A*), Correlation between DNA methylation changes (mutant minus wild type) and euchromatin/heterochromatin compartments, in *Tet2 KO* (*top*), *Dnmt3a KO* (*middle*), and *Dnmt3a/Tet2 DKO* (*bottom*) hematopoietic stem cells (HSPC). Spearman correlation coefficient is shown (r_s value). DNA methylation data from (Zhang et al. 2016, Jeong et al. 2014). (*B*), Genome tracks showing a correspondence between the Hi-C B compartment (negative Hi-C PC1 values, *track 1*) and large hypomethylated domains in *Tet2 KO* (*track 2*). In contrast, *Dnmt3a KO* and *Dnmt3a/Tet2 DKO* show a global, compartment-independent DNA hypomethylation (*tracks 3 and 4*). Gene expression (*track 5*) and 5hmC distribution (*track 6*) in WT HSPC are shown for reference. (*C, D*), Density (*C*) and cumulative distribution (*D*) of average DNA methylation values within 1 kb (*left*) and 10 kb (*right*) windows across the genome, in wild type (WT), *Dnmt3a KO*, *Tet2 KO*, and *Dnmt3a/Tet2 DKO* HSPC.

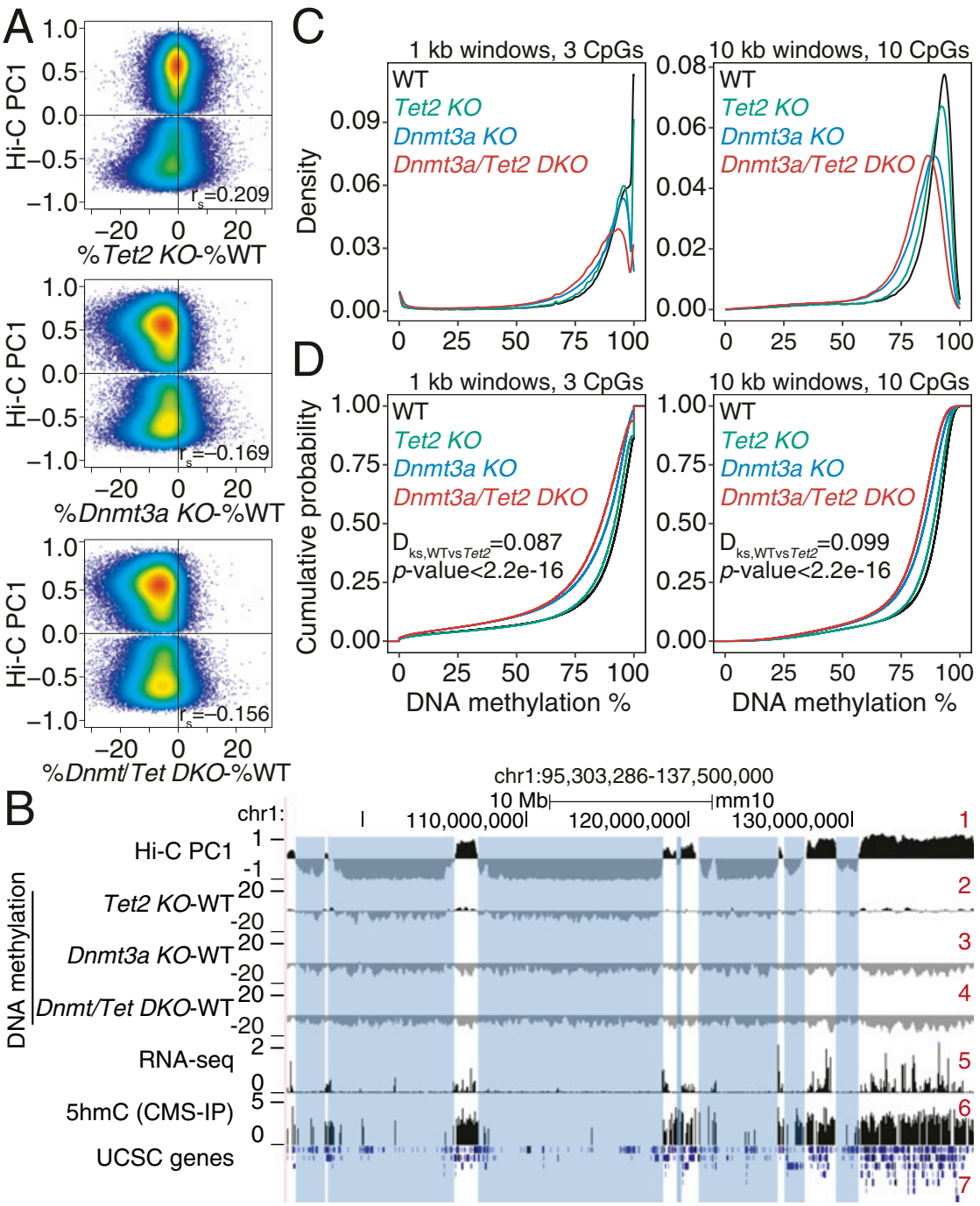

A
Hi-C PC1
%Tet2 KO-%WT
r_s=0.209
%Dnmt3a KO-%WT
r_s=−0.169
%Dnmt/Tet DKO-%WT
r_s=−0.156
C
1 kb windows, 3 CpGs
10 kb windows, 10 CpGs
WT
Tet2 KO
Dnmt3a KO
Dnmt3a/Tet2 DKO
Density
D
Cumulative probability
D_ks,WTvsTet2=0.087
p-value<2.2e-16
D_ks,WTvsTet2=0.099
p-value<2.2e-16
DNA methylation %
B
chr1:95,303,286-137,500,000
10 Mb
mm10
chr1:
110,000,000
120,000,000
130,000,000
DNA methylation
Hi-C PC1
Tet2 KO-WT
Dnmt3a KO-WT
Dnmt/Tet DKO-WT
RNA-seq
5hmC (CMS-IP)
UCSC genes

Figure 2.9. TET TKO mESC proliferate more slowly than their WT counterparts, and genome-wide distribution of DNMT3 enzymes in relation to TET1 and TET2 in mESC. (*A*), Growth curves of WT and Tet1/2/3 TKO mESCs. Cells were split every 3 days, and cells were counted. Reprinted from (Li et al. 2016). (*B-C*), Comparison of the localization of DNMT3 proteins versus TET1 (*B*) and TET2 (*C*) in WT mESC. ChIP-seq enrichment (log_2FoldChange) was calculated for 1 kb windows. A different dataset (Manzo et al. 2017, Baubec et al. 2015) was used from that of **Fig. 2.11**.

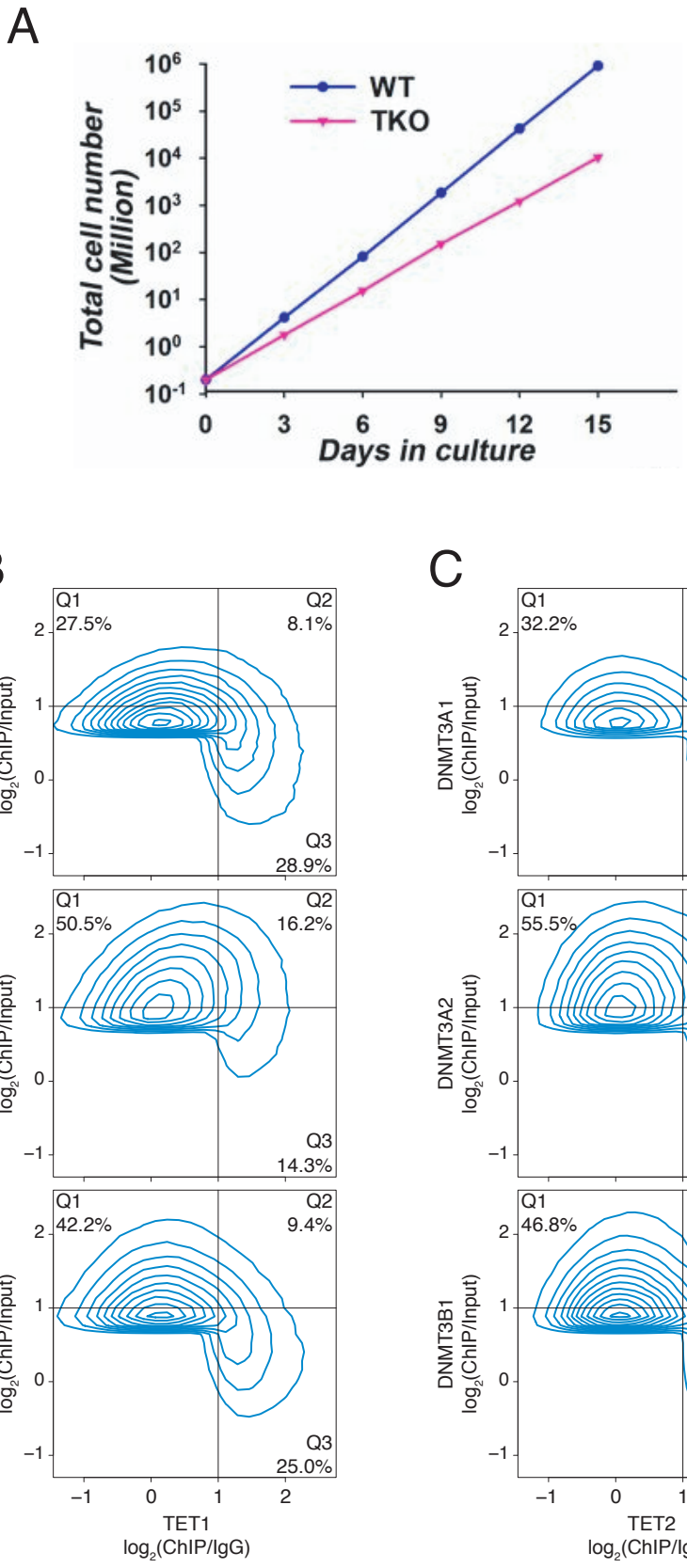

A
WT
TKO
Total cell number (Million)
Days in culture
B
C
Q1
Q2
Q3
27.5%
8.1%
28.9%
50.5%
16.2%
14.3%
42.2%
9.4%
25.0%
32.2%
4.9%
26.4%
55.5%
11.9%
13.7%
46.8%
7.5%
22.1%
DNMT3A1
DNMT3A2
DNMT3B1
log2(ChIP/Input)
TET1
TET2
log2(ChIP/IgG)

Figure 2.10. Contributions of DNMT3 enzymes to DNA methylation in euchromatin and heterochromatin. (*A*), Genome-wide distribution of the *de novo* DNA methyltransferases DNMT3A1, DNMT3A2 and DNMT3B1 within euchromatin and heterochromatin compartments in WT mESC. DNMT ChIP-seq enrichment (log_2 fold change) was calculated at 1 kb resolution. Data from (Manzo et al. 2017, Baubec et al. 2015). (*B*), Genome tracks showing the distribution of DNMT3A1, DNMT3A2 and DNMT3B1 (*tracks 2-4*) within Hi-C defined compartments (*track 1*), and the contribution to DNA methylation of the *de novo* DNMT proteins after reconstitution in a *Dnmt TKO* background in mESC (*tracks 5-7*). *Dnmt TKO* methylation is shown for reference (*track 8*). (*C*), Density distribution of average DNA methylation values within 10 kb windows across the euchromatin and heterochromatin compartments, in *Dnmt TKO* mESC reconstituted with DNMT3A1, DNMT3A2, and DNMT3B1.

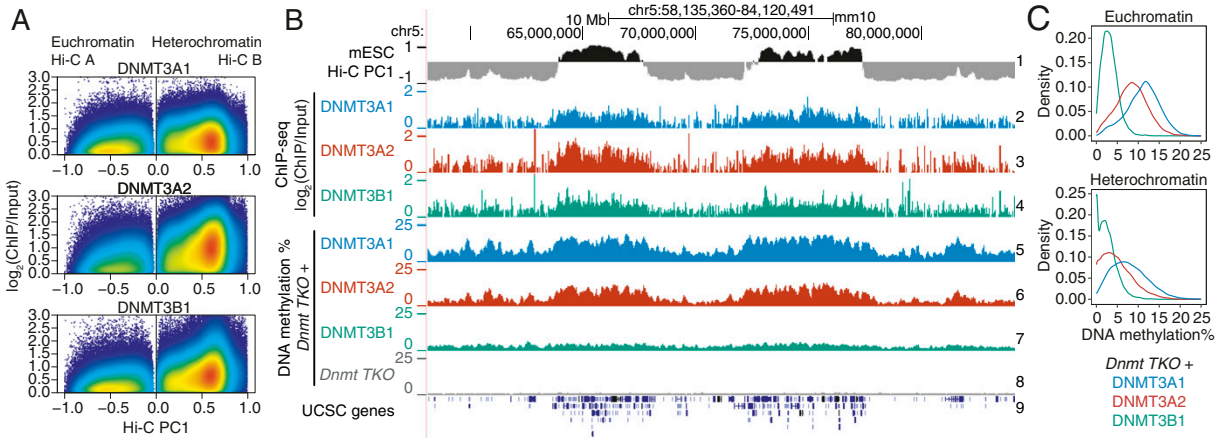
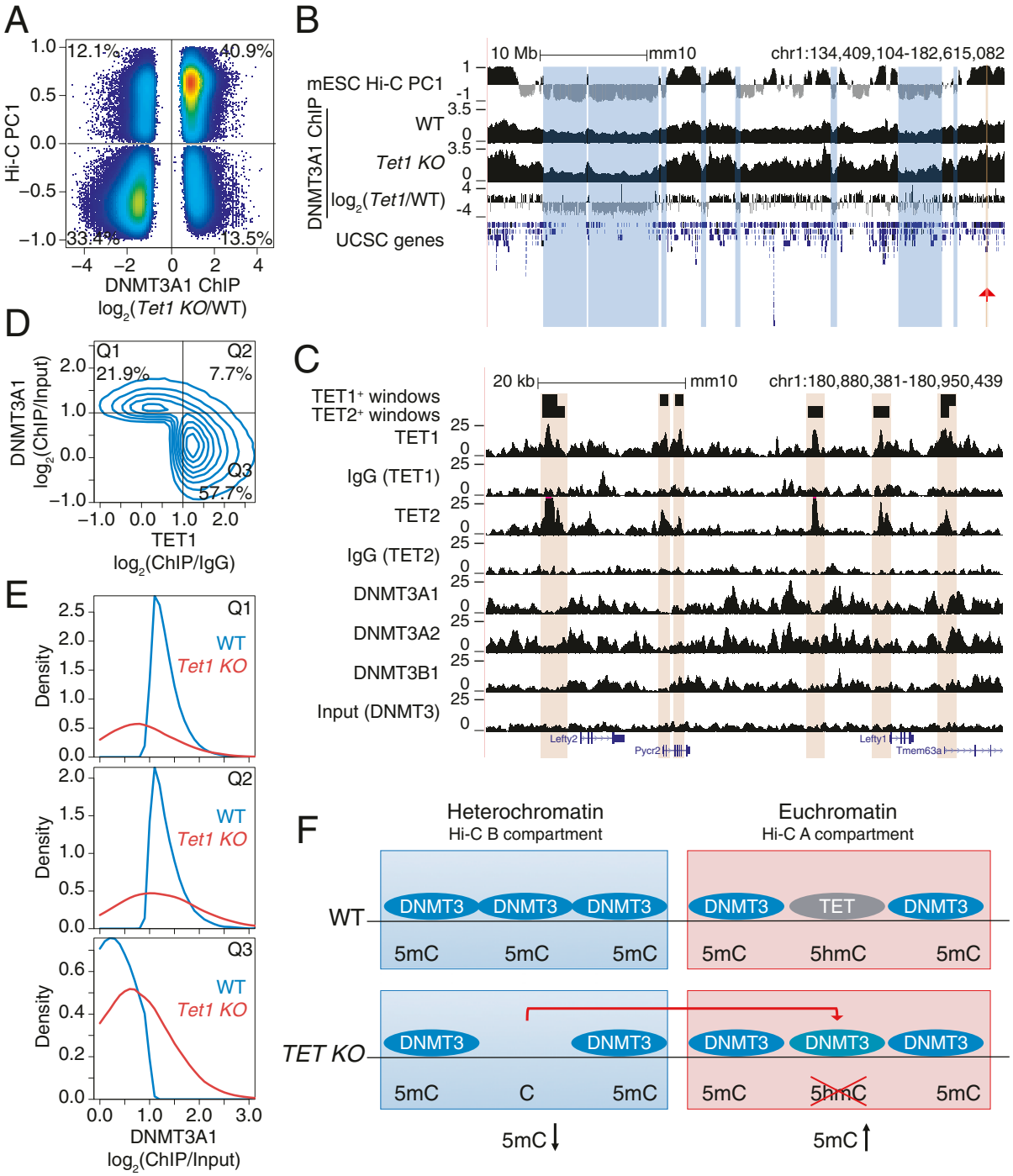

A
Euchromatin
Hi-C A
Heterochromatin
Hi-C B
DNMT3A1
DNMT3A2
DNMT3B1
log2(ChIP/Input)
Hi-C PC1
B
chr5:
10 Mb
chr5:58,135,360-84,120,491
mm10
65,000,000
70,000,000
75,000,000
80,000,000
mESC
Hi-C PC1
ChIP-seq
log2(ChIP/Input)
DNMT3A1
DNMT3A2
DNMT3B1
DNA methylation %
Dnmt TKO +
Dnmt TKO
UCSC genes
C
Euchromatin
Heterochromatin
Density
DNA methylation%
Dnmt TKO +
DNMT3A1
DNMT3A2
DNMT3B1

Figure 2.11. Relocalization of DNMT3A away from the heterochromatin compartment of *TET*-deficient mESC. (*A*), Comparison of changes in DNMT3A1 binding in *Tet1 KO* versus WT mESC, and euchromatin versus heterochromatin compartments (Hi-C PC1 values) in WT mESC. Differential binding was calculated at 1 kb windows and displayed are those with *p*-value<0.05. Percentage of windows within each quadrant is indicated. (*B*), Genome tracks showing a correspondence between the heterochromatic Hi-C B compartment (*track 1*) and decrease in DNMT3A1 occupancy in *Tet1 KO* with respect to WT *(tracks 2-4)*. Tracks 2 and 3 show DNMT3A1 occupancy in WT and *Tet1 KO* mESC, respectively (shown as reads per kilobase per million reads). Track 4 shows the difference in DNMT3A1 occupancy (visualized as $\log_2$ fold-change *Tet1 KO*/WT). A region containing the *Lefty1/2* locus is highlighted by the red arrowhead and is shown in (*C*). (*C*), Zoomed-in view of the *Lefty1/2* locus within the euchromatic compartment. Genome tracks showing mutually exclusive localization of TET1 and TET2 (*tracks 1-6*) and *de novo* DNMT proteins (*tracks 7-10*). (*D*), Mutually exclusive localization between DNMT3A1 and TET1 binding in the euchromatic Hi-C A compartment in WT mESC. Percentage of windows within each quadrant is indicated. ChIP-seq enrichment ($\log_2$ fold change) was calculated for 1 kb windows. DNMT3A1 data from (Gu et al. 2018); TET1 data from (Williams et al. 2011). (*E*), Distribution of DNMT3A1 occupancy in WT and *Tet1 KO* mESC (displayed as enrichment over input) within each of the quadrants defined in (*D*). (*F*), Schematic model illustrating that loss of TET1 in mouse ES cells results in the relocalization of DNMT3A1, on the one hand from the heterochromatic to the euchromatic compartment as shown in Fig. 2.11B, and on the other hand to regions within euchromatin that were previously occupied by TET1 as shown in Fig. 2.11D, E.


A
12.1%
40.9%
33.4%
13.5%
Hi-C PC1
DNMT3A1 ChIP
log2(Tet1 KO/WT)
B
10 Mb
mm10
chr1:134,409,104-182,615,082
mESC Hi-C PC1
DNMT3A1 ChIP
WT
Tet1 KO
log2(Tet1/WT)
UCSC genes
C
20 kb
mm10
chr1:180,880,381-180,950,439
TET1+ windows
TET2+ windows
TET1
IgG (TET1)
TET2
IgG (TET2)
DNMT3A1
DNMT3A2
DNMT3B1
Input (DNMT3)
Lefty2
Pycr2
Lefty1
Tmem63a
D
Q1
21.9%
Q2
7.7%
Q3
57.7%
DNMT3A1
log2(ChIP/Input)
TET1
log2(ChIP/IgG)
E
Q1
Q2
Q3
WT
Tet1 KO
Density
DNMT3A1
log2(ChIP/Input)
F
Heterochromatin
Hi-C B compartment
Euchromatin
Hi-C A compartment
WT
TET KO
DNMT3
TET
5mC
5hmC
C
5mC
5mC
5mC

2.5 Acknowledgements

Chapter 2, in full, is a reprint with modifications as it appears in Paradoxical association of TET loss of function with genome-wide hypomethylation, *PNAS* (2019), published online August 1st, 2019. doi: 10.1073/pnas.1903059116. The dissertation author was the primary author of this paper. Other authors include Ageliki Tsagaratou, Hiroshi Yuita, Hyungseok Seo, Benjamin Delatte, Sven Heinz, Chris Benner & Anjana Rao.

2.6 References

An, J., E. González-Avalos, A. Chawla, M. Jeong, I. F. López-Moyado, W. Li, M. A. Goodell, L. Chavez, M. Ko & A. Rao (2015) Acute loss of TET function results in aggressive myeloid cancer in mice. *Nat Commun,* 6**,** 10071.

Bachman, K. E., M. R. Rountree & S. B. Baylin (2001) Dnmt3a and Dnmt3b are transcriptional repressors that exhibit unique localization properties to heterochromatin. *J Biol Chem,* 276**,** 32282-7.

Baubec, T., D. F. Colombo, C. Wirbelauer, J. Schmidt, L. Burger, A. R. Krebs, A. Akalin & D. Schübeler (2015) Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature,* 520**,** 243-7.

Baylin, S. B. & P. A. Jones (2016) Epigenetic Determinants of Cancer. *Cold Spring Harb Perspect Biol,* 8.

Bendelac, A., P. B. Savage & L. Teyton (2007) The biology of NKT cells. *Annu Rev Immunol,* 25**,** 297-336.

Berman, B. P., D. J. Weisenberger, J. F. Aman, T. Hinoue, Z. Ramjan, Y. Liu, H. Noushmehr, C. P. Lange, C. M. van Dijk, R. A. Tollenaar, D. Van Den Berg & P. W. Laird (2011) Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat Genet,* 44**,** 40-6.

Bonev, B., N. Mendelson Cohen, Q. Szabo, L. Fritsch, G. L. Papadopoulos, Y. Lubling, X. Xu, X. Lv, J. P. Hugnot, A. Tanay & G. Cavalli (2017) Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell,* 171**,** 557-572.e24.

Cimmino, L., M. M. Dawlaty, D. Ndiaye-Lobry, Y. S. Yap, S. Bakogianni, Y. Yu, S. Bhattacharyya, R. Shaknovich, H. Geng, C. Lobry, J. Mullenders, B. King, T. Trimarchi, B. Aranda-Orgilles, C. Liu, S. Shen, A. K. Verma, R. Jaenisch & I. Aifantis (2015) TET1 is a tumor suppressor of hematopoietic malignancy. *Nat Immunol,* 16**,** 653-62.

Couronné, L., C. Bastard & O. A. Bernard (2012) TET2 and DNMT3A mutations in human T-cell lymphoma. *N Engl J Med,* 366**,** 95-6.

Cruickshanks, H. A., T. McBryan, D. M. Nelson, N. D. Vanderkraats, P. P. Shah, J. van Tuyn, T. Singh Rai, C. Brock, G. Donahue, D. S. Dunican, M. E. Drotar, R. R. Meehan, J. R. Edwards, S. L. Berger & P. D. Adams (2013) Senescent cells harbour features of the cancer epigenome. *Nat Cell Biol,* 15**,** 1495-506.

Ehrlich, M. (2009) DNA hypomethylation in cancer cells. *Epigenomics,* 1**,** 239-59.

Engel, I. & M. Kronenberg (2012) Making memory at birth: understanding the differentiation of natural killer T cells. *Curr Opin Immunol,* 24**,** 184-90.

Feinberg, A. P. & B. Vogelstein (1983) Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature,* 301**,** 89-92.

Flavahan, W. A., Y. Drier, B. B. Liau, S. M. Gillespie, A. S. Venteicher, A. O. Stemmer-Rachamimov, M. L. Suvà & B. E. Bernstein (2016) Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature,* 529**,** 110-4.

Gu, T., X. Lin, S. M. Cullen, M. Luo, M. Jeong, M. Estecio, J. Shen, S. Hardikar, D. Sun, J. Su, D. Rux, A. Guzman, M. Lee, L. S. Qi, J. J. Chen, M. Kyba, Y. Huang, T. Chen, W. Li & M. A. Goodell (2018) DNMT3A and TET1 cooperate to regulate promoter epigenetic landscapes in mouse embryonic stem cells. *Genome Biol,* 19**,** 88.

He, Y. F., B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He & G. L. Xu (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science,* 333**,** 1303-7.

Hiratani, I., T. Ryba, M. Itoh, T. Yokochi, M. Schwaiger, C. W. Chang, Y. Lyou, T. M. Townes, D. Schübeler & D. M. Gilbert (2008) Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol,* 6**,** e245.

Hon, G. C., R. D. Hawkins, O. L. Caballero, C. Lo, R. Lister, M. Pelizzola, A. Valsesia, Z. Ye, S. Kuan, L. E. Edsall, A. A. Camargo, B. J. Stevenson, J. R. Ecker, V. Bafna, R. L. Strausberg, A. J. Simpson & B. Ren (2012) Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res,* 22**,** 246-58.

Hon, G. C., C. X. Song, T. Du, F. Jin, S. Selvaraj, A. Y. Lee, C. A. Yen, Z. Ye, S. Q. Mao, B. A. Wang, S. Kuan, L. E. Edsall, B. S. Zhao, G. L. Xu, C. He & B. Ren (2014) 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Mol Cell,* 56**,** 286-297.

Hu, G., K. Cui, D. Fang, S. Hirose, X. Wang, D. Wangsa, W. Jin, T. Ried, P. Liu, J. Zhu, E. V. Rothenberg & K. Zhao (2018) Transformation of Accessible Chromatin and 3D Nucleome Underlies Lineage Commitment of Early T Cells. *Immunity,* 48**,** 227-242.e8.

Huang, Y., L. Chavez, X. Chang, X. Wang, W. A. Pastor, J. Kang, J. A. Zepeda-Martínez, U. J. Pape, S. E. Jacobsen, B. Peters & A. Rao (2014) Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proc Natl Acad Sci U S A,* 111**,** 1361-6.

Huang, Y., W. A. Pastor, Y. Shen, M. Tahiliani, D. R. Liu & A. Rao (2010) The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS One,* 5**,** e8888.

Ito, S., L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He & Y. Zhang (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science,* 333**,** 1300-3.

Iyer, L. M., D. Zhang, R. F. de Souza, P. J. Pukkila, A. Rao & L. Aravind (2014) Lineage-specific expansions of TET/JBP genes and a new class of DNA transposons shape fungal genomic and epigenetic landscapes. *Proc Natl Acad Sci U S A,* 111**,** 1676-83.

Jeong, M., D. Sun, M. Luo, Y. Huang, G. A. Challen, B. Rodriguez, X. Zhang, L. Chavez, H. Wang, R. Hannah, S. B. Kim, L. Yang, M. Ko, R. Chen, B. Göttgens, J. S. Lee, P. Gunaratne, L. A.

Godley, G. J. Darlington, A. Rao, W. Li & M. A. Goodell (2014) Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nat Genet,* 46**,** 17-23.

Jones, P. A. & S. B. Baylin (2002) The fundamental role of epigenetic events in cancer. *Nat Rev Genet,* 3**,** 415-28.

Kriaucionis, S. & N. Heintz (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science,* 324**,** 929-30.

Ley, T. J., C. Miller, L. Ding, B. J. Raphael, A. J. Mungall, A. Robertson, K. Hoadley, T. J. Triche, P. W. Laird, J. D. Baty, L. L. Fulton, R. Fulton, S. E. Heath, J. Kalicki-Veizer, C. Kandoth, J. M. Klco, D. C. Koboldt, K. L. Kanchi, S. Kulkarni, T. L. Lamprecht, D. E. Larson, L. Lin, C. Lu, M. D. McLellan, J. F. McMichael, J. Payton, H. Schmidt, D. H. Spencer, M. H. Tomasson, J. W. Wallis, L. D. Wartman, M. A. Watson, J. Welch, M. C. Wendl, A. Ally, M. Balasundaram, I. Birol, Y. Butterfield, R. Chiu, A. Chu, E. Chuah, H. J. Chun, R. Corbett, N. Dhalla, R. Guin, A. He, C. Hirst, M. Hirst, R. A. Holt, S. Jones, A. Karsan, D. Lee, H. I. Li, M. A. Marra, M. Mayo, R. A. Moore, K. Mungall, J. Parker, E. Pleasance, P. Plettner, J. Schein, D. Stoll, L. Swanson, A. Tam, N. Thiessen, R. Varhol, N. Wye, Y. Zhao, S. Gabriel, G. Getz, C. Sougnez, L. Zou, M. D. Leiserson, F. Vandin, H. T. Wu, F. Applebaum, S. B. Baylin, R. Akbani, B. M. Broom, K. Chen, T. C. Motter, K. Nguyen, J. N. Weinstein, N. Zhang, M. L. Ferguson, C. Adams, A. Black, J. Bowen, J. Gastier-Foster, T. Grossman, T. Lichtenberg, L. Wise, T. Davidsen, J. A. Demchok, K. R. Shaw, M. Sheth, H. J. Sofia, L. Yang, J. R. Downing, G. Eley, et al. (2013) Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med,* 368**,** 2059-74.

Li, X., X. Yue, W. A. Pastor, L. Lin, R. Georges, L. Chavez, S. M. Evans & A. Rao (2016) Tet proteins influence the balance between neuroectodermal and mesodermal fate choice by inhibiting Wnt signaling. *Proc Natl Acad Sci U S A,* 113**,** E8267-E8276.

Li, Z., H. Dai, S. N. Martos, B. Xu, Y. Gao, T. Li, G. Zhu, D. E. Schones & Z. Wang (2015) Distinct roles of DNMT1-dependent and DNMT1-independent methylation patterns in the genome of mouse embryonic stem cells. *Genome Biol,* 16**,** 115.

Lieberman-Aiden, E., N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander & J. Dekker (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science,* 326**,** 289-93.

Lin, Y. C., C. Benner, R. Mansson, S. Heinz, K. Miyazaki, M. Miyazaki, V. Chandra, C. Bossen, C. K. Glass & C. Murre (2012) Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. *Nat Immunol,* 13**,** 1196-204.

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

Lister, R., E. A. Mukamel, J. R. Nery, M. Urich, C. A. Puddifoot, N. D. Johnson, J. Lucero, Y. Huang, A. J. Dwork, M. D. Schultz, M. Yu, J. Tonti-Filippini, H. Heyn, S. Hu, J. C. Wu, A. Rao,

M. Esteller, C. He, F. G. Haghighi, T. J. Sejnowski, M. M. Behrens & J. R. Ecker (2013) Global epigenomic reconfiguration during mammalian brain development. *Science,* 341**,** 1237905.

Lister, R., M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q. M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren & J. R. Ecker (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature,* 462**,** 315-22.

Lister, R., M. Pelizzola, Y. S. Kida, R. D. Hawkins, J. R. Nery, G. Hon, J. Antosiewicz-Bourget, R. O'Malley, R. Castanon, S. Klugman, M. Downes, R. Yu, R. Stewart, B. Ren, J. A. Thomson, R. M. Evans & J. R. Ecker (2011) Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature,* 471**,** 68-73.

Lu, F., Y. Liu, L. Jiang, S. Yamaguchi & Y. Zhang (2014) Role of Tet proteins in enhancer activity and telomere elongation. *Genes Dev,* 28**,** 2103-19.

Manzo, M., J. Wirz, C. Ambrosi, R. Villaseñor, B. Roschitzki & T. Baubec (2017) Isoform-specific localization of DNMT3A regulates DNA methylation fidelity at bivalent CpG islands. *EMBO J,* 36**,** 3421-3434.

Odejide, O., O. Weigert, A. A. Lane, D. Toscano, M. A. Lunning, N. Kopp, S. Kim, D. van Bodegom, S. Bolla, J. H. Schatz, J. Teruya-Feldstein, E. Hochberg, A. Louissaint, D. Dorfman, K. Stevenson, S. J. Rodig, P. P. Piccaluga, E. Jacobsen, S. A. Pileri, N. L. Harris, S. Ferrero, G. Inghirami, S. M. Horwitz & D. M. Weinstock (2014) A targeted mutational landscape of angioimmunoblastic T-cell lymphoma. *Blood,* 123**,** 1293-6.

Palomero, T., L. Couronné, H. Khiabanian, M. Y. Kim, A. Ambesi-Impiombato, A. Perez-Garcia, Z. Carpenter, F. Abate, M. Allegretta, J. E. Haydu, X. Jiang, I. S. Lossos, C. Nicolas, M. Balbin, C. Bastard, G. Bhagat, M. A. Piris, E. Campo, O. A. Bernard, R. Rabadan & A. A. Ferrando (2014) Recurrent mutations in epigenetic regulators, RHOA and FYN kinase in peripheral T cell lymphomas. *Nat Genet,* 46**,** 166-70.

Papaemmanuil, E., M. Gerstung, L. Bullinger, V. I. Gaidzik, P. Paschka, N. D. Roberts, N. E. Potter, M. Heuser, F. Thol, N. Bolli, G. Gundem, P. Van Loo, I. Martincorena, P. Ganly, L. Mudie, S. McLaren, S. O'Meara, K. Raine, D. R. Jones, J. W. Teague, A. P. Butler, M. F. Greaves, A. Ganser, K. Döhner, R. F. Schlenk, H. Döhner & P. J. Campbell (2016) Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med,* 374**,** 2209-2221.

Pastor, W. A., L. Aravind & A. Rao (2013) TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat Rev Mol Cell Biol,* 14**,** 341-56.

Peric-Hupkes, D., W. Meuleman, L. Pagie, S. W. Bruggeman, I. Solovei, W. Brugman, S. Gräf, P. Flicek, R. M. Kerkhoven, M. van Lohuizen, M. Reinders, L. Wessels & B. van Steensel (2010) Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol Cell,* 38**,** 603-13.

Poleshko, A., P. P. Shah, M. Gupta, A. Babu, M. P. Morley, L. J. Manderfield, J. L. Ifkovits, D. Calderon, H. Aghajanian, J. E. Sierra-Pagán, Z. Sun, Q. Wang, L. Li, N. C. Dubois, E. E. Morrisey, M. A. Lazar, C. L. Smith, J. A. Epstein & R. Jain (2017) Genome-Nuclear Lamina Interactions Regulate Cardiac Stem Cell Lineage Restriction. *Cell,* 171**,** 573-587.e14.

Rasmussen, K. D., G. Jia, J. V. Johansen, M. T. Pedersen, N. Rapin, F. O. Bagger, B. T. Porse, O. A. Bernard, J. Christensen & K. Helin (2015) Loss of TET2 in hematopoietic cells leads to DNA hypermethylation of active enhancers and induction of leukemogenesis. *Genes Dev,* 29**,** 910-22.

Sakata-Yanagimoto, M., T. Enami, K. Yoshida, Y. Shiraishi, R. Ishii, Y. Miyake, H. Muto, N. Tsuyama, A. Sato-Otsubo, Y. Okuno, S. Sakata, Y. Kamada, R. Nakamoto-Matsubara, N. B. Tran, K. Izutsu, Y. Sato, Y. Ohta, J. Furuta, S. Shimizu, T. Komeno, T. Ito, M. Noguchi, E. Noguchi, M. Sanada, K. Chiba, H. Tanaka, K. Suzukawa, T. Nanmoku, Y. Hasegawa, O. Nureki, S. Miyano, N. Nakamura, K. Takeuchi, S. Ogawa & S. Chiba (2014) Somatic RHOA mutation in angioimmunoblastic T cell lymphoma. *Nat Genet,* 46**,** 171-5.

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

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

Tsagaratou, A., T. Äijö, C. W. Lio, X. Yue, Y. Huang, S. E. Jacobsen, H. Lähdesmäki & A. Rao (2014) Dissecting the dynamic changes of 5-hydroxymethylcytosine in T-cell development and differentiation. *Proc Natl Acad Sci U S A,* 111**,** E3306-15.

van Steensel, B. & A. S. Belmont (2017) Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell,* 169**,** 780-791.

Williams, K., J. Christensen, M. T. Pedersen, J. V. Johansen, P. A. Cloos, J. Rappsilber & K. Helin (2011) TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature,* 473**,** 343-8.

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

Yin, Y., E. Morgunova, A. Jolma, E. Kaasinen, B. Sahu, S. Khund-Sayeed, P. K. Das, T. Kivioja, K. Dave, F. Zhong, K. R. Nitta, M. Taipale, A. Popov, P. A. Ginno, S. Domcke, J. Yan, D. Schübeler, C. Vinson & J. Taipale (2017) Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science,* 356.

Yue, X., S. Trifari, T. Äijö, A. Tsagaratou, W. A. Pastor, J. A. Zepeda-Martínez, C. W. Lio, X. Li, Y. Huang, P. Vijayanand, H. Lähdesmäki & A. Rao (2016) Control of Foxp3 stability through modulation of TET activity. *J Exp Med,* 213**,** 377-97.

Zhang, X., J. Su, M. Jeong, M. Ko, Y. Huang, H. J. Park, A. Guzman, Y. Lei, Y. H. Huang, A. Rao, W. Li & M. A. Goodell (2016) DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells. *Nat Genet,* 48**,** 1014-23.

Zhou, W., H. Q. Dinh, Z. Ramjan, D. J. Weisenberger, C. M. Nicolet, H. Shen, P. W. Laird & B. P. Berman (2018) DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat Genet,* 50**,** 591-602.

Chapter 3

TET loss of function is associated with genome instability and aneuploidies

Abstract

Cancer genomes are characterized by focal increases in DNA methylation, co-occurring with widespread hypomethylation. TET loss-of-function results in a similar genomic footprint, and in a mouse model of T cell lymphoma driven by TET deficiency (*Tet2/3 DKO* T cells), in addition to the expected DNA hypermethylation largely confined to the active euchromatic compartment, malignant T cells display DNA hypomethylation in the heterochromatic genomic compartment, as well as reactivation of repeat elements and enrichment for single nucleotide alterations, primarily in heterochromatic regions of the genome. Moreover, like DNMT-deficient cells, expanded *Tet2/3 DKO* T cells displayed recurrent aneuploidies and increased accumulation of DNA double-strand breaks. Our data suggest heterochromatin DNA hypomethylation may at least partially explain some of the shared features between TET and DNMT mutants, highlight the potential contribution of heterochromatin hypomethylation to oncogenesis, and suggest that heterochromatin dysfunction contributes to the recurrent association of TET loss-of-function with cancer.

3.1 Introduction

Even in the absence of TET coding region mutations, TET loss-of-function and low 5hmC levels are strongly associated with cancer (Ko et al. 2010, Marçais et al. 2017, Lemonnier et al. 2018, Huang and Rao 2014). *TET2* mutations are frequent in diverse hematopoietic malignancies, including myelodysplastic syndromes (MDS), acute myeloid leukemias (AML) and peripheral T cell lymphomas (PTCL) (Delhommeau et al. 2009, Langemeijer et al. 2009, Lemonnier et al. 2012, Odejide et al. 2014). However, both solid tumors and hematopoietic malignancies display TET loss-of-function without TET coding region mutations, as a result of TET promoter methylation, increased degradation of TET proteins, or aberrant microRNA expression (Cimmino et al. 2015, Ko, An and Rao 2015, Raffel et al. 2017, Wu et al. 2018).

Cancer genomes are characterized by two opposing patterns of aberrant DNA methylation: focal or regional hypermethylation (Jones and Baylin 2002, Baylin and Jones 2016) and widespread DNA hypomethylation (Feinberg and Vogelstein 1983, Ehrlich 2009). DNA hypermethylation at promoters and enhancers contributes to oncogenesis through transcriptional silencing of genes involved in DNA damage repair and tumor suppressors (Jones and Baylin 2002, Baylin and Jones 2016). Local hypermethylation has been shown to reflect the impaired expression or activity of TET proteins. Despite our understanding of the biochemical and biological consequences of local hypermethylation, however, the causes and consequences of DNA hypomethylation in cancer genomes are less well understood.

We previously reported the consequences of profound TET loss-of-function in T cells (Tsagaratou et al. 2017). Mice with disruption of both *Tet2* and *Tet3* genes in T cells (*Tet2/3 DKO* mice) displayed a dramatic expansion of a normally minor T cell subtype known as *i*NKT cells (Tsagaratou et al. 2017), which express an invariant T cell receptor alpha (TCRα) chain, recognize lipid antigens presented on a non-classical major histocompatibility protein, CD1d, and are not deleted in the thymus (Bendelac, Savage and Teyton 2007, Engel and Kronenberg 2012).

In this Chapter, we examine other genomic features of *Tet2/3 DKO* NKT cells reported to be associated with hypomethylation, including reactivation of repeat elements and chromosome segregation defects. Here we report that the predominant genetic abnormalities observed in expanding *i*NKT cells from *Tet2/3 DKO* mice are increased copy number alterations (CNA) and aneuploidies, particularly a recurrent trisomy of mouse chromosome 17, concomitantly with increased accumulation of DNA double-strand breaks (DSBs).

3.2 Results

3.2.1 Antigen-driven expansion, increased clonality and DNA damage in TET-deficient T cell leukemia/ lymphoma

To study hypomethylation induced by TET loss-of-function in the context of oncogenic transformation, we used a mouse model in which mice with profound TET loss-of-function (*Tet2-/- Tet3fl/fl CD4Cre (Tet2/3 DKO*) lacking Tet2 and Tet3 in T cells) rapidly developed an aggressive T cell leukemia/ lymphoma with 100% penetrance (Tsagaratou et al. 2017). The disease involves a normally minor subset of T cells (*i*NKT cells, hereafter referred to simply as NKT cells), which recognize lipid antigens presented on a non-classical major histocompatibility complex (MHC) protein known as CD1d (Bendelac et al. 2007, Engel and Kronenberg 2012). *Tet2/3 DKO* mice showed >10-fold expansion of NKT cells in the thymus as early as 20 days after birth and in the spleen by 3-4 weeks, and succumbed to an NKT cell leukemia starting at 5 weeks (Tsagaratou et al. 2017). Transfer of purified NKT cells from young mice, even into fully immunocompetent recipients, resulted in transfer of the leukemia, but transfer to recipient mice lacking CD1d, the MHC protein that presents lipid antigen to NKT cells (Bendelac et al. 2007), resulted in minimal expansion (Tsagaratou et al. 2017) (**Fig. 3.1A**), indicating that the leukemia was driven by NKT cell expansion arising from presentation of lipid antigens by CD1d.

The ability to transfer disease suggested that NKT cells from young mice had already undergone some degree of oncogenic transformation. To gauge the extent of transformation, we transferred 5 x 10^5 *Tet2/3 DKO* NKT (CD45.2$^+$) cells to recipient CD45.1$^+$ mice, allowed the transferred cells to expand for 12 weeks, then isolated the expanded CD45.2$^+$ NKT cells and transferred different cell numbers into secondary recipient mice. Recipients of 5000, 50,000 or 500,000 cells developed disease simultaneously within 4 weeks (**Fig. 3.1A**), at which time the transferred CD45.2$^+$ NKT cells had expanded to be ~70%, ~55% and ~35% of the splenocytes of the recipient mice (**Fig. 3.1B**). Recipients of 50 or 500 expanded NKT cells showed obvious signs of disease by 7 weeks (**Fig. 3.1A**), at which time the transferred CD45.2$^+$ NKT cells had expanded to be around 60% and 50% of splenocytes of the recipient mice respectively (**Fig. 3.1B**).

Since as few as 50 transferred and expanded *Tet2/3 DKO* NKT cells were sufficient to transfer disease to secondary recipient mice, we examined the clonality of the expanding cells. Sequencing of T cell receptor Vβ chains, using primers for the junctional regions of rearranged TCR Vβ7, Vβ8.1 and Vβ8.2 chains, revealed that wild type NKT cells were extremely diverse, with ~37,000 Vβ sequences identified overall and no single sequence represented at greater than 0.06% (**Fig. 3.1C**, *left*), whereas NKT cells from young *Tet2/3 DKO* mice were oligoclonal (**Fig. 3.1C**, *middle*; each color represents a single TCR Vβ sequence). In contrast, cells that had been transferred from young mice into wild type recipients and allowed to expand for 9-12 weeks were effectively monoclonal (>95%; **Fig. 3.1C**, *right*). When purified NKT cells from the same donor young *Tet2/3 DKO* mouse were transferred into multiple recipient mice, we observed predominant expansion of the same NKT cell clone as judged by the TCR Vβ sequence utilized in each case (**Fig. 3.1C**, *right*), indicating a strong proliferative advantage for that specific clone, likely due to due to strong selective expansion of cells with specific genomic characteristics (*see below*). Given that the expanded cells were effectively monoclonal and constituted >95% of lymphocytes in the

peripheral lymphoid organs of recipient mice, we were able to perform a variety of whole-genome analyses – whole-genome sequencing (WGS), together with WGBS and Hi-C, at relatively low cost.

Overall, the data on TET-deficient NKT cell lymphomas were completely concordant with those for the other TET-deficient cell types (as described in Chapter 2). TET-deficient T cell lymphomas show euchromatic compartment hypermethylation and heterochromatic compartment hypomethylation. Regardless of cell type, TET deficiency was broadly associated with DNA hypermethylation in the euchromatic compartment, concurrently with DNA hypomethylation in the heterochromatic compartment. In the remainder of this study, we examine other genomic features of *Tet2/3 DKO* NKT cells reported to be associated with hypomethylation, including reactivation of repeat elements and increased mutational load.

3.2.2 Mutational signatures in TET-deficient T cell lymphomas

Hypomethylation has been previously associated with increased mutation rates (Chen et al. 1998) and genome instability (Eden et al. 2003, Gaudet et al. 2003), and we and others have observed increased levels of DNA damage after TET deletion (An et al. 2015, Cimmino et al. 2015). Expansion of *Tet2/3 DKO* NKT cells after transfer was accompanied by a striking increase in DNA double-strand breaks (DSBs): the entire population of expanded *Tet2/3 DKO* NKT cells showed increased staining for phospho-H2AX (γH2AX, a marker for DSBs (Fernandez-Capetillo et al. 2004)), compared to WT NKT cells (**Fig. 3.1D, E**). In contrast, *Tet2/3 DKO* NKT cells transferred to *CD1d KO* recipient mice, which undergo only minimal expansion (Tsagaratou et al., 2017) (**Fig. 3.1A**), displayed only a slight increase in DSBs compared to WT NKT cells (**Fig. 3.1D, E**). These results are consistent with the reports of increased DNA DSBs in *Tet1*-deficient B cells (Cimmino et al. 2015) and in acute myeloid leukemias resulting from inducible deletion of *Tet2* and *Tet3* (An et al. 2015).

We performed whole-genome sequencing (WGS) at >20x coverage. We examined our WGS data to identify single-nucleotide variations (SNVs) in WT versus *Tet2/3 DKO* NKT cells. WGS on expanded NKT cells showed that most SNVs occurred in the largely heterochromatic compartment, which constitutes 54% of the genome but contains 77% of the SNVs (**Fig. 3.2A**). Furthermore, the SNVs identified in gene coding regions in the five different samples of transferred and expanded NKT cells from independent recipient mice (**Fig. 3.2B**) were not recurrently observed, suggesting that the selective advantage potentially conferred by any given SNV is limited to individual mice. Thus most SNVs observed in *Tet2/3 DKO* NKT cells after transfer and expansion arise through a stochastic process occurring primarily in the heterochromatic compartment, as also observed for H3K9me3-marked heterochromatic genome regions in human cancers (Schuster-Böckler and Lehner 2012).

Among the mutations which affected coding regions, many of them occurred in cancer-associated genes with functions in signal transduction, cell cycle, DNA damage responses and gene/chromatin regulation. Mutations that involved a substantial fraction of alleles (19-56%) and were likely to affect function because they introduced stop codons or led to frameshifts were observed in *Adamts4*, *Mfap4*, *Zfp608* (a zinc finger transcription factor expressed in thymocytes), *Rad51ap2* (RAD51-associated protein 2; RAD51 is involved in homologous recombination and DNA repair), *Dlgap2* (discs large-associated protein 2), *Cct6b* and *Smarca4* (encoding Brg1, the catalytic ATPase subunit of the SWI/SNF chromatin remodeling complex). Nonsynonymous substitutions involving 9-64% of alleles were observed in *Apc*, *Map2k2* (encoding MEK2), *Kras, Arhgef12* (a guanine nucleotide exchange factor for RhoA, a small GTPase frequently mutated in T cell lymphomas), *Rnf144a* (a ubiquitin ligase for the catalytic subunit of DNA-PK, a pivotal kinase in the DNA repair pathway), *Ddias* (DNA damage-induced suppressor of apotosis) *Gemin5* (an RNA-binding protein involved in mRNA splicing and translation), *Alk* (a receptor tyrosine kinase that is frequently amplified or rearranged in cancers including non-small cell lung cancer

and anaplastic large cell lymphomas), *Mtor, Ctcf*, *Pik3r1, Rhov* (a Ras homolog) and others. While it was not clear whether these nonsynonymous mutations compromised the biological function of the encoded gene, the observed large values for variant allele percentages suggest that the mutations conferred a selective advantage on the expanding *Tet2/3 DKO* NKT cells.

The mutational signature of the SNVs, based on nucleotide substitutions and sequence context at the 5' and 3' ends (Alexandrov et al. 2013), clustered separately between the euchromatic and heterochromatic compartments (**Fig. 3.3A, B**). In both compartments, the signature was predominantly characterized by transitions (C>T, T>C; includes G>A and A>G). Even though SNVs in general were more prevalent in the heterochromatic compartment, SNVs at cytosines in the CpG context were more prevalent in the hypermethylated euchromatic compartment (14%) compared to the hypomethylated heterochromatic compartment (8.6%) (**Fig. 3.3A**, *compare C>T red bars in top and bottom panels*), as expected from the tendency of 5mC to undergo spontaneous deamination (Coulondre et al. 1978). Indeed, for CpGs for which DNA methylation data were available from WGBS analysis, we observed that C>T mutations in the CpG context occurred at CpG sites that were largely methylated (**Fig. 3.3C**). Rainfall plots of intermutational distance against genomic location for three individual *Tet2/3 DKO* mice showed that mutations were often clustered in similar chromosomal locations but did not occur at the same nucleotides (**Fig. 3.3D**).

3.2.3 Reactivation of repeat elements in TET-deficient T cell lymphomas

DNA hypomethylation has been widely associated with reactivation of repeat elements (Walsh, Chaillet and Bestor 1998). In light of the global hypomethylation in heterochromatin of *Tet2/3 DKO* NKT cells, we analyzed the expression levels of distinct families of repeat elements in young *Tet2/3 DKO* NKT cells by RNA-seq (**Fig. 3.2C-D**), keeping in mind that long interspersed elements (LINEs) are primarily located in heterochromatin while short interspersed elements

(SINEs) are found in euchromatin (van Steensel and Belmont 2017). Indeed, using RNA-seq datasets from our previous study (Tsagaratou et al. 2017) as well as newly generated RNA-seq data from total ribodepleted RNA (**Fig. 3.2C, D**; **Fig. 3.3E**), we found that for those repeat elements for which we were able to detect transcripts reliably in at least one biological replicate, LTR and LINE elements were more highly expressed in *Tet2/3 DKO* NKT cells with respect to WT, whereas SINE elements remained largely unchanged. Furthermore, a substantial fraction of the identified mutations fell within repetitive elements, such as LINEs and LTRs, but appeared underrepresented in SINE elements with respect to the genome average (**Fig. 3.2E**). These data support the hypothesis that the reactivation of LINEs and LTRs result from TET-associated hypomethylation occurring in heterochromatin, whereas the euchromatic compartment undergoes TET-associated hypermethylation and therefore most SINEs remain silent.

Reactivation and spurious transcription of repeat elements has been associated with formation of R-loops and genome instability (Zeller et al. 2016, Zhu et al. 2018), linked to DNA damage and DNA double-strand breaks (Crossley, Bocek and Cimprich 2019). Indeed, we found an increase of R-loops in expanded *Tet2/3 DKO* compared to WT NKT cells, as detected by flow cytometry and DNA dot blots using the S9.6 antibody against RNA:DNA hybrids (Boguslawski et al. 1986) (**Figs. 3.1F, G**). Further analysis will require methods for mapping DSBs and R-loops that can be applied to the very small numbers of NKT cells that can be isolated from normal WT mice, so as to determine whether these features show differential distribution in the euchromatic and heterochromatic compartments in WT versus TET-deficient cells.

3.2.4 Widespread copy number variations and aneuploidies in TET-deficient T cell lymphomas

Genome-wide analyses of NKT cells from 10 different *Tet2/3 DKO* mice after transfer and expansion in recipient mice using several different genome-wide techniques – WGS (5 mice),

WGBS (2 mice) and HiC (3 mice) – revealed striking copy number variations (CNVs) and aneuploidies (**Fig. 3.4**, **Fig. 3.5**). All *Tet2/3 DKO* NKT cells showed chromosome 17 trisomy as well as partial, often recurrent, trisomies of other chromosomes and numerous focal gains and losses across all chromosomes (**Fig. 3.4A**). Consistent with chromosome 17 trisomy, the expression levels of genes located on chromosome 17 were increased relative to the expression levels of genes located on other chromosomes (**Fig. 3.4B**). We confirmed the presence of recurrent aneuploidies by metaphase spreads on two additional samples of transferred and expanded *Tet2/3 DKO* NKT cells (**Fig. 3.4C**). Low-coverage whole-genome sequencing of one of these samples (MCS A), in which ~60% of the cells showed predominantly 43 chromosomes (3 more than the normal number of 40 chromosomes in the mouse), confirmed the presence of trisomies of chromosome 2, 14 and 17 (**Fig. 3.4D**); of these, chromosome 17 trisomy and partial trisomy of chromosome 2 had been observed in other samples using other techniques (**Fig. 3.4A**).

Notably, chromosome 17 trisomy was not observed in bulk NKT cells from 6/6 young *Tet2/3 DKO* mice analyzed, although a partial trisomy of Chr. 9 was observed in NKT cells from one of the young mice (**Fig. 3.5**). Single-cell analyses – whole-genome sequencing (Knouse et al., 2014) or RNA-seq (Patel et al., 2014) – will be needed to establish whether young *Tet2/3 DKO* NKT cells are prone to stochastic aneuploidies and other copy-number variations that are not discernible in bulk populations because they are not selected for during expansion. Conversely, the invariable appearance of chromosome 17 trisomy and other recurrent copy number variations in expanded NKT cells suggests strong selection for these features during *Tet2/3 DKO* NKT cell transfer and expansion. We are investigating the underlying mechanisms in a separate project.

3.2.5 Inducible TET deletion in mESC results in aneuploidies and chromosome segregation defects

The difference in aneuploidies between young and expanded NKT cells led us to ask whether aneuploidies and chromosome segregation defects were early or late consequences of TET loss-of-function. To examine this point, we set up a system for acute deletion of all three TET proteins in mESC. Briefly, we generated ESC in-house from *Tet1fl/fl Tet2fl/fl Tet3fl/fl* mice that had been bred to contain tamoxifen-inducible *Cre-ERT2* and Cre-inducible *Rosa26 H2B-GFP*^{LSL} alleles (**Fig. 3.6A**). When these ESC were treated with 4-hydroxytamoxifen (4-OHT) in culture for 2.5-3 days and then rested for 2 days (**Fig. 3.6A**), the Cre-ERT2 fusion protein translocates to the nucleus and deletes both the floxed *Tet* exons and the LSL (*loxp*-STOP-*loxP*) cassette, inducing TET gene disruption and simultaneous expression of the H2B-GFP fusion protein. Control ESC, containing only the *Cre-ERT2* and *Rosa26 H2B-GFP*^{LSL} alleles, were treated in parallel with 4-OHT. *Tet* gene disruption was confirmed both by acute, cell division-dependent loss of 5hmC (**Fig. 3.6B**), and by the absence of reads over the deleted TET exons in RNA-seq performed at 6 days (*not shown*). Metaphase spreads performed on day 6, after 4 days of 4-OHT treatment followed by 2 days of rest (**Fig. 3.6A**), showed that acutely *Tet*-deleted ESC (*TKO*) displayed considerably more aneuploidy (42.5%) compared to WT ESC (19.5%) (**Fig. 3.6C**). Additionally, live-cell imaging conducted over a 48-hour period showed that micronuclei and all chromosome segregation defects including chromosomal bridges and lagging chromosomes were significantly more frequent in *Tet*-deleted compared to WT ESC (**Fig. 3.6D**). These data are consistent with previous observations from a different lab (Kafer et al., 2016).

3.3 Discussion

To explore the biological consequences of TET loss-of-function in vivo, particularly in the context of heterochromatin DNA hypomethylation, we used a mouse model of profound TET deficiency in T cells. Mice with deletion of *Tet2* and *Tet3* genes in T cells showed early signal-dependent expansion and increased clonality, which rapidly progressed to an aggressive NKT

cell lymphoma. The expanded *Tet2/3 DKO* NKT cells developed the same aberrations in DNA methylation – hypermethylation in the euchromatic compartment and hypomethylation in the heterochromatic compartment – that occur in cancer genomes and that we have noted above for multiple TET-deficient cell types. The cells accumulated SNVs, largely in the hypomethylated heterochromatic compartment but also in the coding regions of potential cancer-associated genes, through an apparently stochastic process that differed in each individual mouse. We also observed reactivation of transposable elements, particularly LINE and LTR elements that are primarily located in heterochromatin, recalling the genome instability produced by spurious transcription of repeat elements (Zhu et al. 2018, Zeller et al. 2016); these repetitive elements were also more prone to mutations compared to the genome in general. As described in more detail below, DNA hypomethylation in heterochromatin may at least partly explain the oncogenic transformation, genome instability and DNA damage observed in diverse mouse models of partial or profound TET deficiency (Cimmino et al. 2015, An et al. 2015). The latency and penetrance of oncogenic transformation in these models likely depends on the extent of TET loss-of-function.

DNA hypomethylation has been associated with many biological consequences, including reactivation of transposable elements (Walsh et al. 1998), sharply increased mutation rates (Chen et al. 1998) and genome instability with chromosome segregation defects and aneuploidies (Eden et al. 2003, Gaudet et al. 2003). Mice with a hypomorphic mutation in *Dnmt1* displayed genome-wide hypomethylation in all tissues and developed T cell lymphomas that occurred in 80% of mice and were characterized by recurrent aneuploidies (Gaudet et al. 2003). Reactivation of transposable elements is prevalent in cancer genomes, and is associated with the formation of RNA-DNA hybrids and R-loops (Zeller et al. 2016, Zhu et al. 2018), which in turn have been linked to DNA damage and the appearance of DNA double-strand breaks (Crossley et al. 2019). Each of these features was observed, together with heterochromatin hypomethylation, in expanded *Tet2/3 DKO* NKT cells. Thus in addition to their well-established role in promoting and maintaining

DNA demethylation at promoters, gene bodies and enhancers, TET proteins participate in maintaining physiological levels of DNA methylation in heterochromatic compartments of the genome.

Our data suggest that loss of DNA methylation in heterochromatin results in "heterochromatin dysfunction" (Janssen, Colmenares and Karpen 2018). This phenomenon has many manifestations, including aneuploidies resulting from chromosome instability related to centromere dysfunction, as also observed in ICF (immunodeficiency, centromere instability, facial abnormalities) patients with germline DNMT3B mutations (Xu et al. 1999), as well as reactivation of transposable elements and increased R-loops. These features are all observed in *Tet2/3 DKO* NKT cells, as well as in cells with hypomorphic mutations in DNMT1 (Gaudet et al. 2003). Based on these considerations, we speculate that cancers related to TET loss-of-function are initiated at least partly through defects in the maintenance of heterochromatin function. Thus loss of DNA methylation in the heterochromatic compartment, and the consequent development of heterochromatin dysfunction could be the first steps in the development of many cancers characterized by TET loss-of-function. Moreover, mutations in proteins associated with the maintenance of heterochromatin integrity are frequent in cancer and many of them (e.g. *NPM1*) co-occur with TET2 mutations, leading to the postulate that heterochromatin dysfunction is not only a common feature of sporadic (non-hereditary) human cancers but also potentially an initiating event in oncogenic transformation (Janssen et al. 2018).

3.4 Figures

Figure 3.1. Expansion of *Tet2/3 DKO* NKT cells is accompanied by increased clonality, and accumulation of DNA double-strand breaks and R-loops. (*A*), *Left,* experimental workflow. *Middle,* the picture depicts splenomegaly and enlarged lymph nodes in wild type (WT) but not CD1d knockout recipients of *Tet2/3 DKO* NKT cells. *Right*, times of disease emergence (*see* (*B*)). (*B*), Percentage of WT or expanded *Tet2/3 DKO* NKT cells in splenocytes of congenic WT recipient mice, injected with the indicated numbers of *Tet2/3 DKO* NKT cells. 2 mice were used per condition. Mice injected with as few as 50 *Tet2/3 DKO* cells develop NKT cell lymphoma. (*C*), Evaluation of TCR β chain CDR3 variable region sequences by whole genome sequencing (WGS) of DNA from NKT cells. 1 WT mouse, 5 3-4 week old *Tet2/3* DKO mice, and 7 WT recipients of 5 x 10^5 *Tet2/3* DKO NKT cells were evaluated. Each color represents a single TCR Vβ sequence. Mice that received NKT cells from a single *Tet2/3 DKO* donor showed expansion of the same NKT cell clone. Asterisks indicate independent mice WGS D and WGS E for which SNV data are shown in Fig. 4B. (*D*), Evaluation of phospho-H2AX staining as a marker for DNA DSBs in WT and *Tet2/3 DKO* NKT cells. NKT cells (αGalCer-CD1d$^+$, TCRβ$^+$) were isolated from healthy WT mice or after transfer of *Tet2/3 DKO* NKT cells to non-irradiated WT or CD1dKO recipients as outlined in (*A*). A representative flow cytometric analysis is shown. (*E*), (*Top panel*) Percentage of phospho-H2AX$^+$ NKT cells isolated from WT mice or from *Tet2/3* DKO NKT cells transferred to and recovered from non-irradiated WT or CD1dKO recipients as outlined in (*A*). (*Bottom panel*) Median fluorescence intensity of phospho-H2AX staining in NKT cells isolated from WT or from *Tet2/3* DKO mice after transfer to and recovery from WT or CD1dKO recipients as outlined in (*A*). Data are mean ± SEM, n=2 (WT NKT cells), n=3 (NKT cells from CD1dKO recipients) and n=4 (NKT cells from WT recipients). ns, not significant. At least 2 independent experiments were performed per condition. (*F*), Flow cytometric analysis evaluating R-loops in NKT cells isolated from healthy WT mice (*blue histogram*) or from *Tet2/3 DKO* NKT cells transferred and expanded in nonirradiated immunocompetent WT recipients (*red histogram*). The S9.6 antibody recognizes RNA:DNA hybrids. (*G*), Dot blot of genomic DNA from *Tet2/3 DKO* NKT cells transferred to and expanded in non-irradiated WT recipients shows increased R-loop formation compared to WT NKT cells (*top panel, right*). Specificity for RNA:DNA hybrids was confirmed by RNase H treatment of genomic DNA prior to spotting, which results in elimination of the signal (*top panel, left*). Equivalent DNA loading was confirmed by methylene blue staining (*bottom panel*).

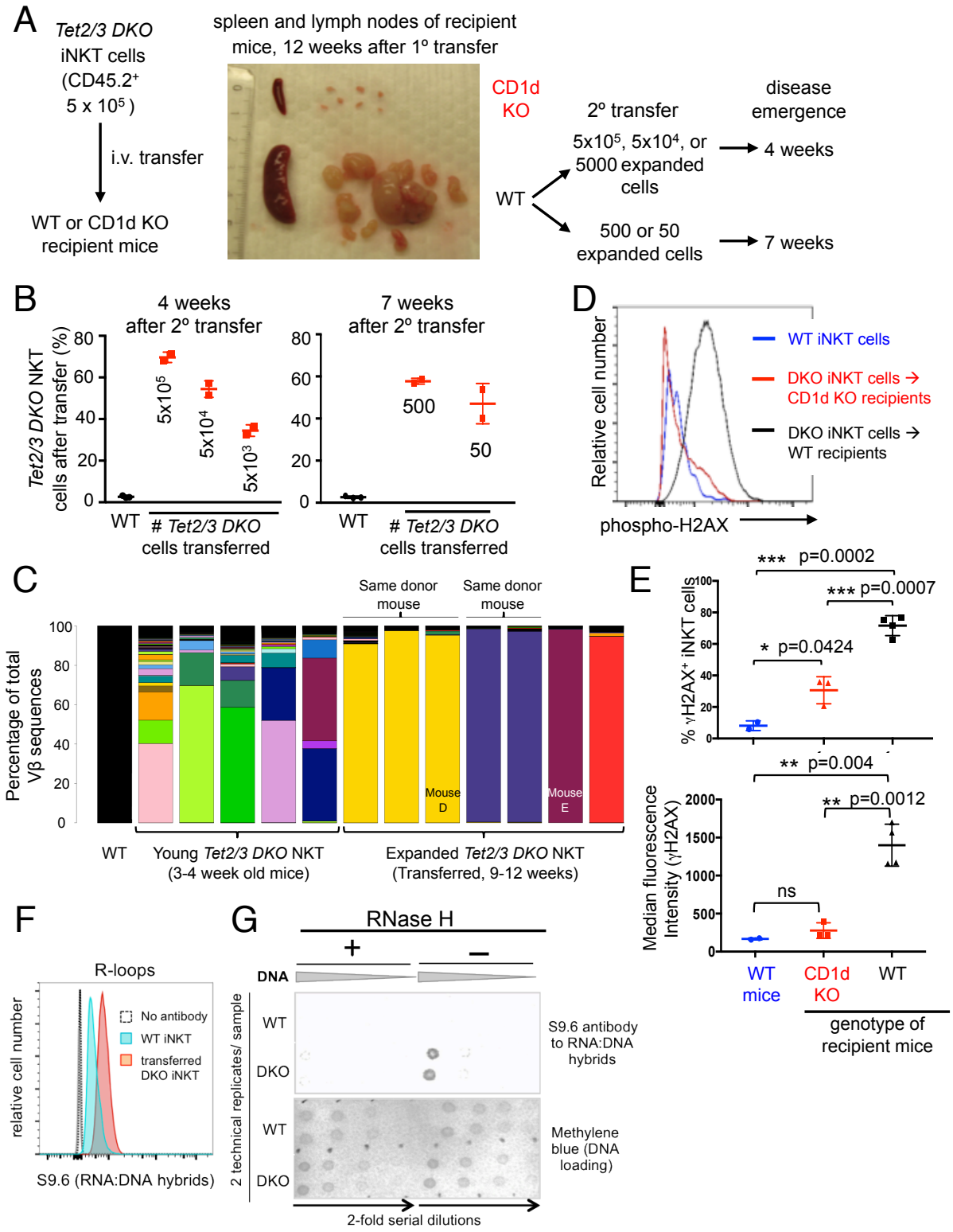

A
Tet2/3 DKO iNKT cells (CD45.2+ 5 x 10^5)
i.v. transfer
WT or CD1d KO recipient mice
spleen and lymph nodes of recipient mice, 12 weeks after 1° transfer
CD1d KO
WT
2° transfer
5x10^5, 5x10^4, or 5000 expanded cells
500 or 50 expanded cells
disease emergence
4 weeks
7 weeks
B
4 weeks after 2° transfer
7 weeks after 2° transfer
Tet2/3 DKO NKT cells after transfer (%)
5x10^5
5x10^4
5x10^3
500
50
WT
Tet2/3 DKO cells transferred
D
Relative cell number
phospho-H2AX
WT iNKT cells
DKO iNKT cells → CD1d KO recipients
DKO iNKT cells → WT recipients
C
Same donor mouse
Same donor mouse
Percentage of total Vβ sequences
Mouse D
Mouse E
WT
Young Tet2/3 DKO NKT (3-4 week old mice)
Expanded Tet2/3 DKO NKT (Transferred, 9-12 weeks)
E
*** p=0.0002
*** p=0.0007
* p=0.0424
% γH2AX+ iNKT cells
** p=0.004
** p=0.0012
ns
Median fluorescence Intensity (γH2AX)
WT mice
CD1d KO
WT
genotype of recipient mice
F
R-loops
relative cell number
No antibody
WT iNKT
transferred DKO iNKT
S9.6 (RNA:DNA hybrids)
G
RNase H
+
–
DNA
WT
DKO
2 technical replicates/ sample
S9.6 antibody to RNA:DNA hybrids
Methylene blue (DNA loading)
2-fold serial dilutions

Figure 3.2. Transposable element reactivation and increased mutations predominate in heterochromatin of *TET*-deficient NKT cell lymphoma. (*A*), Genome-wide distribution (*top*), and percentage of SNVs (*bottom*) located within euchromatin and heterochromatin compartments in *Tet2/3 DKO* expanded NKT cells. (*B*), Coding mutations resulting in change in the amino acid sequence in five independent WGS samples from *Tet2/3 DKO* expanded NKT cells. Mutant allele frequencies are shown in parentheses. (*C*), Distribution of the changes in expression (log_2 fold change) of transposable elements (TEs) belonging to the LTR, LINE and SINE families in *Tet2/3 DKO* young NKT cells compared to WT, obtained from analysis of total RNA-seq data. Fold change differences for all genes in the genome are shown for reference. (*D*), Volcano plots of expression of TEs belonging to the LTR, LINE and SINE families, in *Tet2/3 DKO* young NKT cells compared to WT. Differentially expressed (DE) TEs (adj. *p*-value<0.1) are highlighted. (*E*), Percentage of SNVs in *Tet2/3 DKO* expanded NKT cells that overlap with LINEs, LTR, and SINEs, compared to the mm10 genome distribution of each TE families. *p*-values were calculated by one sample t-test.

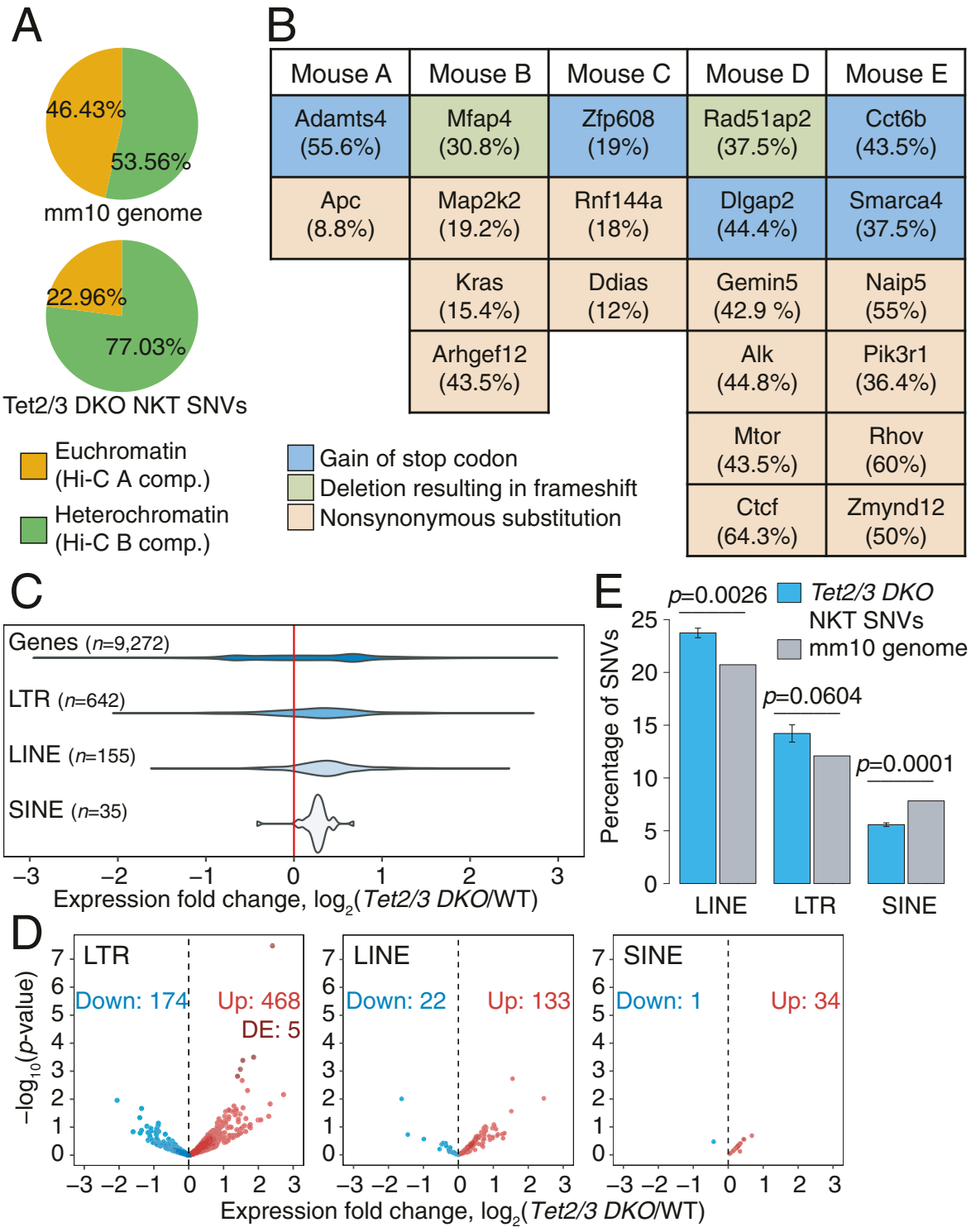
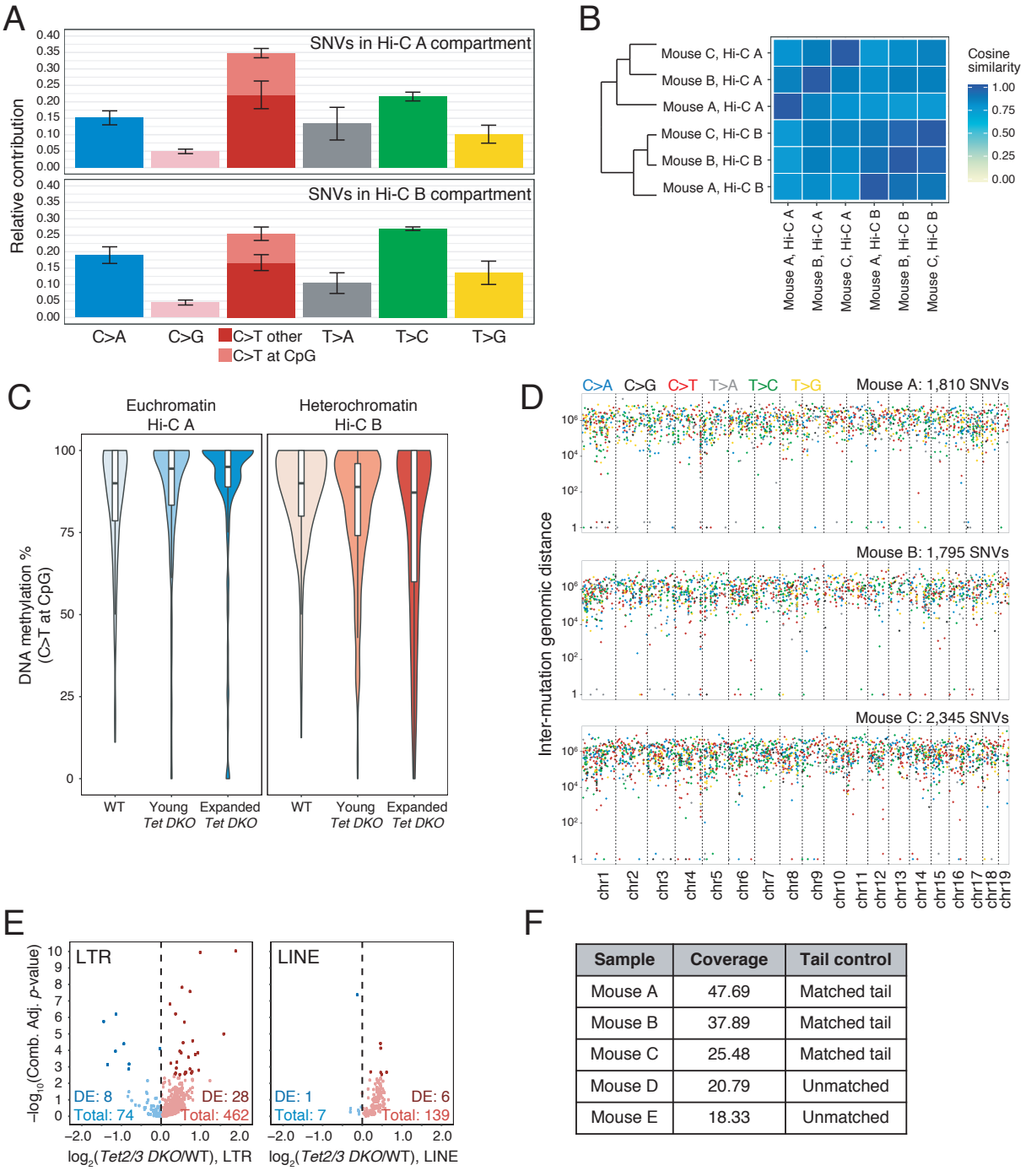

A
46.43%
53.56%
mm10 genome
22.96%
77.03%
Tet2/3 DKO NKT SNVs
Euchromatin (Hi-C A comp.)
Heterochromatin (Hi-C B comp.)
B
Mouse A
Mouse B
Mouse C
Mouse D
Mouse E
Adamts4 (55.6%)
Mfap4 (30.8%)
Zfp608 (19%)
Rad51ap2 (37.5%)
Cct6b (43.5%)
Apc (8.8%)
Map2k2 (19.2%)
Rnf144a (18%)
Dlgap2 (44.4%)
Smarca4 (37.5%)
Kras (15.4%)
Ddias (12%)
Gemin5 (42.9 %)
Naip5 (55%)
Arhgef12 (43.5%)
Alk (44.8%)
Pik3r1 (36.4%)
Mtor (43.5%)
Rhov (60%)
Ctcf (64.3%)
Zmynd12 (50%)
Gain of stop codon
Deletion resulting in frameshift
Nonsynonymous substitution
C
Genes (n=9,272)
LTR (n=642)
LINE (n=155)
SINE (n=35)
Expression fold change, log2(Tet2/3 DKO/WT)
D
LTR
Down: 174
Up: 468
DE: 5
LINE
Down: 22
Up: 133
SINE
Down: 1
Up: 34
−log10(p-value)
Expression fold change, log2(Tet2/3 DKO/WT)
E
p=0.0026
Tet2/3 DKO NKT SNVs
mm10 genome
p=0.0604
p=0.0001
Percentage of SNVs
LINE
LTR
SINE

Figure 3.3. Mutational spectrum of transferred and expanded *Tet2/3 DKO* T cells. (*A*), Mutational spectrum associated with euchromatic Hi-C A and heterochromatic Hi-C B compartments in *Tet2/3 DKO* expanded NKT cells. (*B*), Cosine similarity between mutational profiles obtained from three independent *Tet2/3 DKO* expanded T cells WGS samples (WGS A, WGS B, WGS C), separating the mutations by their location within Hi-C compartments (A vs B) in *Tet2/3 DKO* NKT cells. Notice how mutational profiles cluster by Hi-C compartment and not by sample of origin. (*C*), DNA methylation at cytosines within the CpG context with C>T substitution type. DNA methylation values shown for WT, *Tet2/3 DKO* young, and *Tet2/3 DKO* expanded T cells. (*D*), Rainfall plots representing the inter-mutational genomic distance (y-axis) for all single nucleotide variants (SNV) (x-axis) encountered in samples WGS A, B and C. Substitutions are color-coded as indicated in the top of the figure. While mutations cluster at certain regions, each mice exhibits a unique spectrum of mutations. (*E*), Distribution of the changes in expression (log_2 fold-change difference) of different transposable elements in *Tet2/3 DKO* young NKT cells compared to WT, belonging to the LTR, LINE and SINE families, obtained from analysis of total (ribodepleted) RNA-seq data. Fold-change differences for all the genes in the genome are shown for reference. The highest expressed genes were used as control genes for size factor estimation in DESeq2. P-values of two independent experiments (same biological conditions, different library preparation methods, TruSeq and SMARTseq) were combined using the Fisher method. (*F*), Table with coverage values of WGS samples.



F

Sample	Coverage	Tail control
Mouse A	47.69	Matched tail
Mouse B	37.89	Matched tail
Mouse C	25.48	Matched tail
Mouse D	20.79	Unmatched
Mouse E	18.33	Unmatched

Figure 3.4. Recurrent aneuploidies in in *TET*-deficient NKT cell lymphoma. (*A*), Copy number analysis in somatic chromosomes from different samples originating from WGS, WGBS and Hi-C experiments. Each line inside a chromosome represents a 500 kb window. Red, copy number gain; blue, copy number loss. Notice the recurrence of chromosome 17 gain. (*B*), Gene expression per chromosome in transferred and expanded *Tet2/3 DKO NKT* cells compared to WT splenic NKT cells. Each boxplot represents the distribution of $\log_2$ fold changes at the RNA level in *Tet2/3 DKO* vs WT, for all expressed genes at a given chromosome. (*C*), Histograms summarizing the number of chromosomes per cell in metaphase chromosome spreads (MCS) for two samples of transferred and expanded *Tet2/3 DKO* NKT cells. (*D*), Copy number analysis by WGS on sample MCS A confirms that most cells have 43 chromosomes, and reveals trisomy of Chr. 2, 14 and 17 (sample in **Fig. 3.4C**). Each dot represents a 500 kb window.

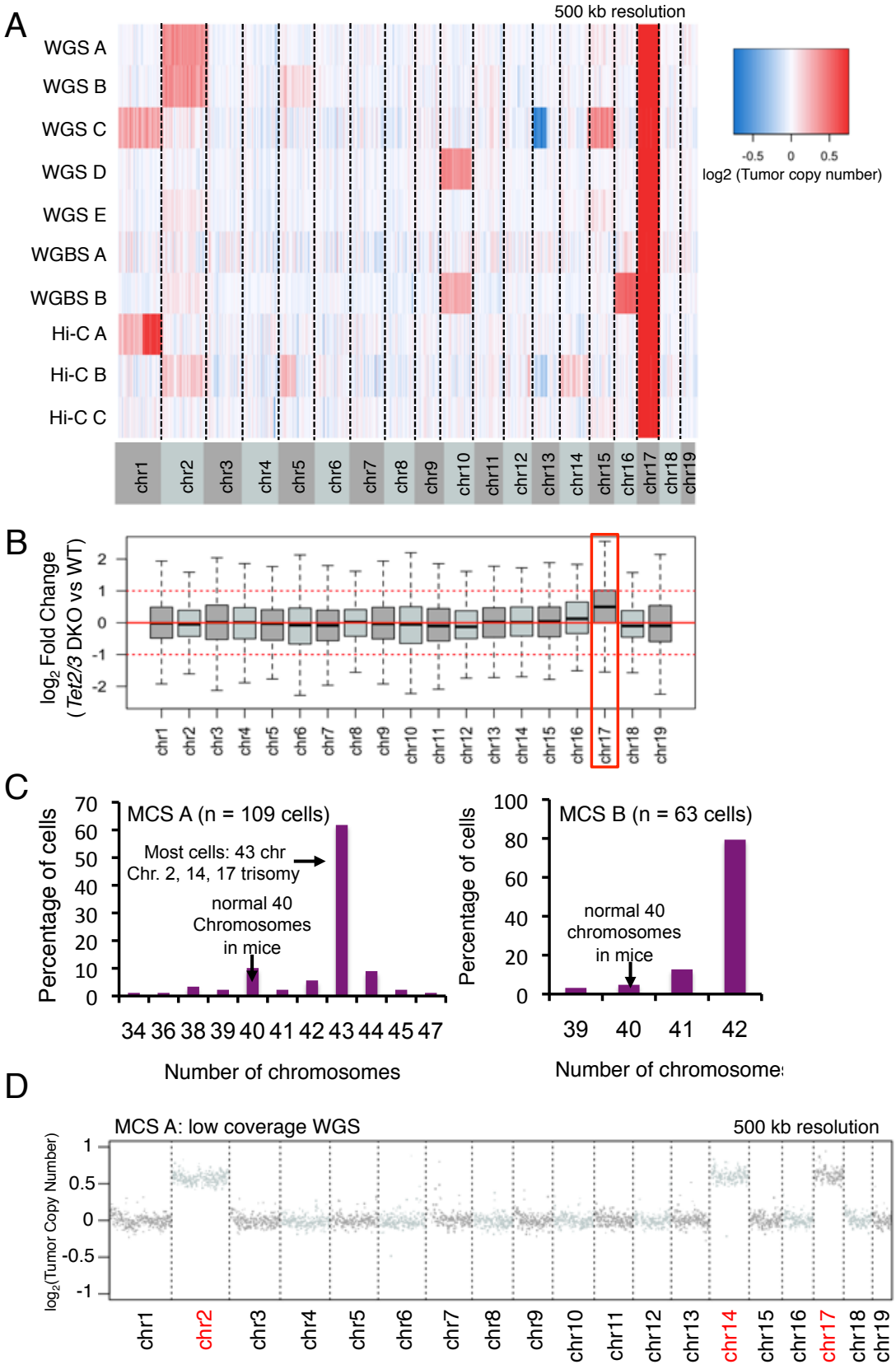

A
500 kb resolution
WGS A
WGS B
WGS C
WGS D
WGS E
WGBS A
WGBS B
Hi-C A
Hi-C B
Hi-C C
-0.5 0 0.5
log2 (Tumor copy number)
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19
B
log2 Fold Change (Tet2/3 DKO vs WT)
C
MCS A (n = 109 cells)
Most cells: 43 chr
Chr. 2, 14, 17 trisomy
normal 40 Chromosomes in mice
Percentage of cells
Number of chromosomes
MCS B (n = 63 cells)
normal 40 chromosomes in mice
D
MCS A: low coverage WGS
500 kb resolution
log2(Tumor Copy Number)

Figure 3.5. Chromosome 9 aneuploidy found in 1 out of 6 NKT cell lymphoma samples originating from young *Tet2/3 DKO* mice. (*A*), Copy number analysis using WGS (low coverage) information from four samples of young *Tet2/3 DKO* NKT cells. Each dot represents a 500 kb window. This assay would only detect aneuploidies if they are shared by many cells within a population. (*B*), Copy number analysis using WGBS information from two samples of young *Tet2/3 DKO* NKT cells. Each dot represents a 500 kb window. This assay would only detect aneuploidies if they are shared by many cells within a population.

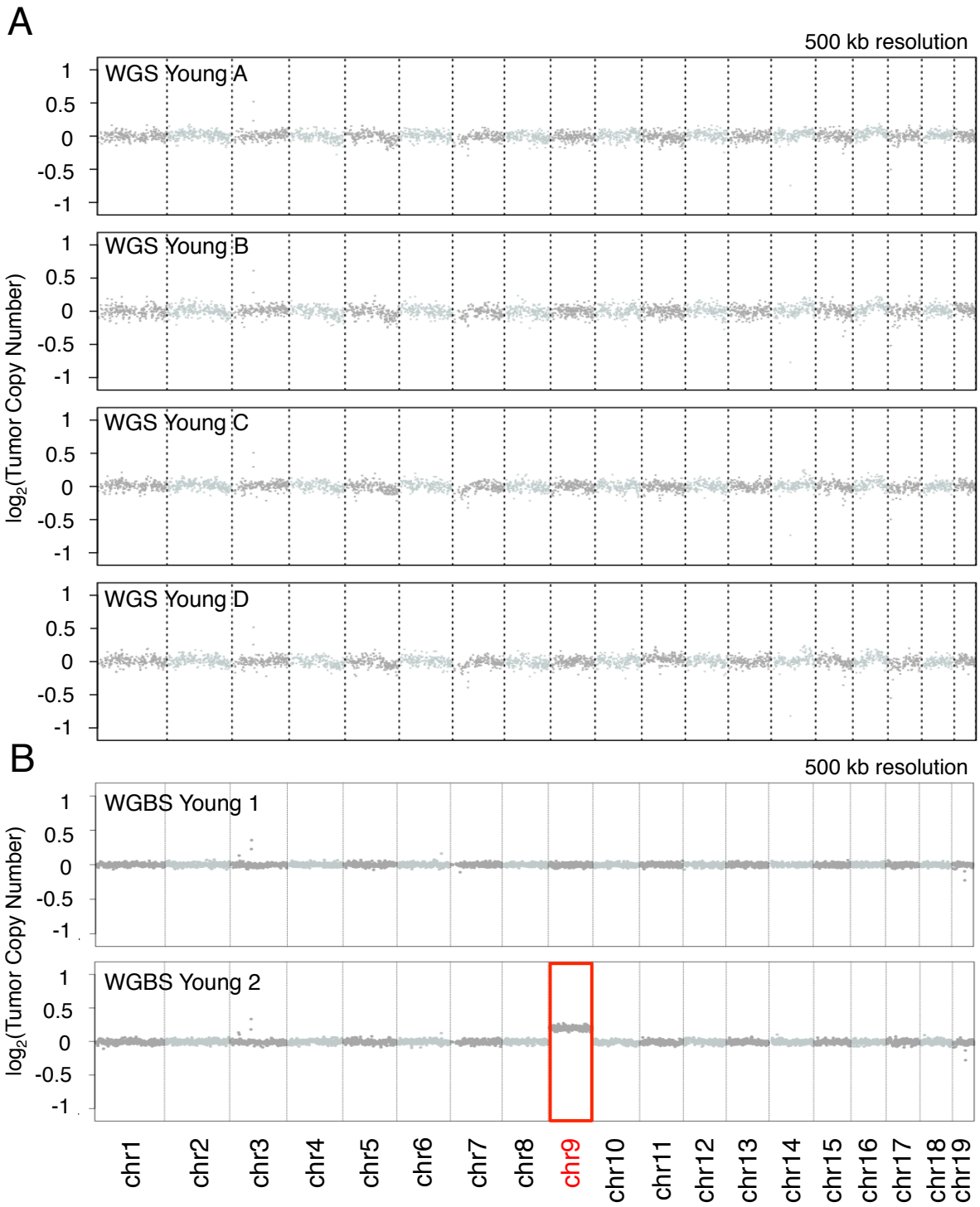
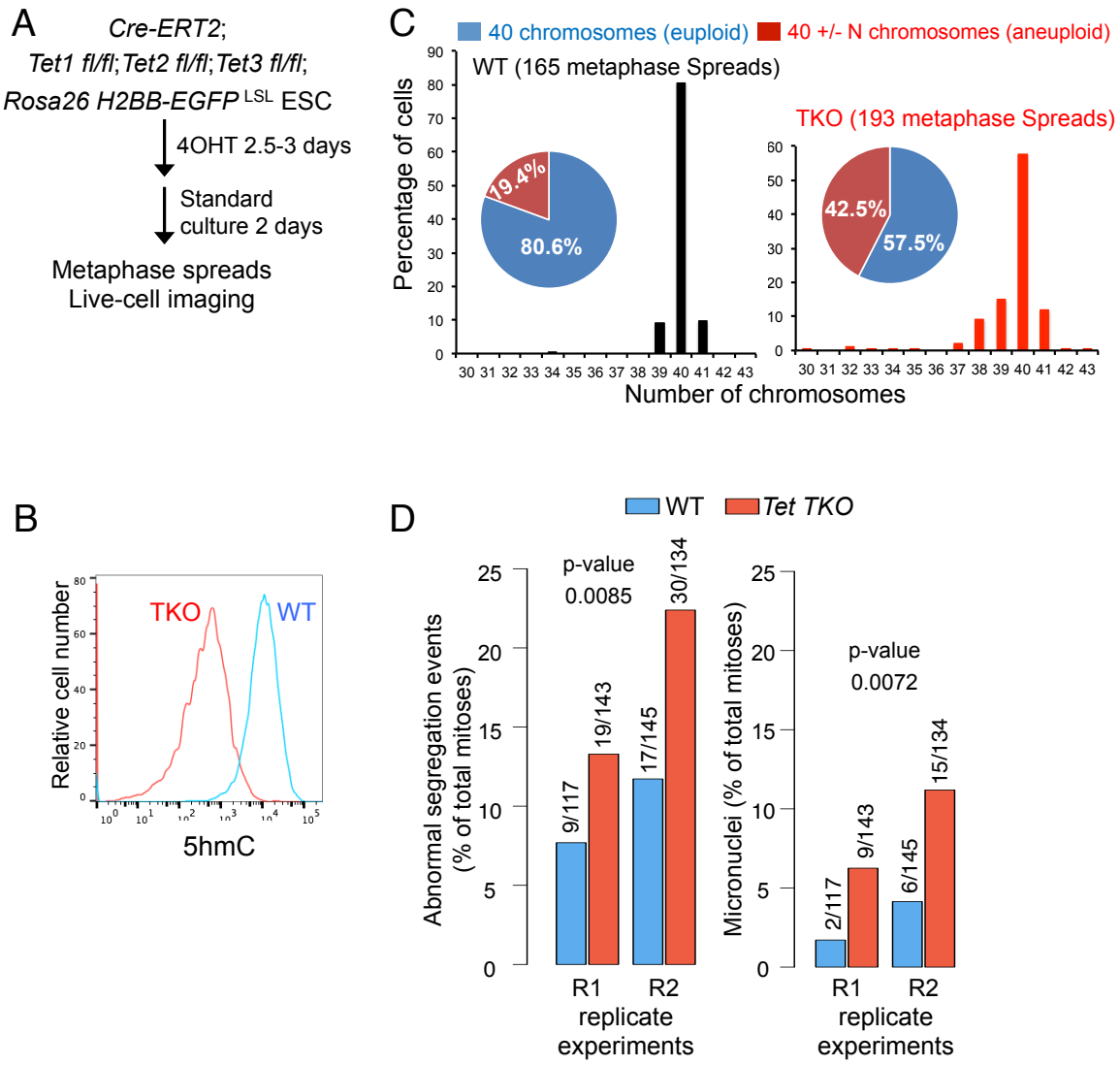

A
500 kb resolution
WGS Young A
WGS Young B
WGS Young C
WGS Young D
log2(Tumor Copy Number)
1
0.5
0
-0.5
-1
B
500 kb resolution
WGBS Young 1
WGBS Young 2
log2(Tumor Copy Number)
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19

Figure 3.6. Inducible acute TET triple deletion in mESC results in aneuploidies and chromosome segregation defects. (*A*), Outline of the procedure used to generate acute deletion of the 3 TET enzymes in mESCs. (*B*), Quantification of 5-hmC levels on day 6 post-4OH-TAM treatment in mESCs harboring an acute loss-of-function of the three TET enzymes. Data are representative of 5 experiments. (*C*), Ploidy analysis by metaphase spreads of WT versus *Tet1/2/3 TKO* mESCs 6 days after 4OH-TAM treatment. (*D*), Abnormal chromosome segregation and micronuclei occurrence in WT versus *Tet1/2/3 TKO* mESCs observed along the 48 hours of live-cell imaging (Day 4.5 to 6.5 post-4OH-TAM treatment). Cochran-Mantel-Haenszel test was performed to calculate pvalues. Data are representative from 2 independent experiments.


A
Cre-ERT2;
Tet1 fl/fl;Tet2 fl/fl;Tet3 fl/fl;
Rosa26 H2BB-EGFP LSL ESC
4OHT 2.5-3 days
Standard culture 2 days
Metaphase spreads
Live-cell imaging
C
40 chromosomes (euploid)
40 +/- N chromosomes (aneuploid)
WT (165 metaphase Spreads)
TKO (193 metaphase Spreads)
19.4%
80.6%
42.5%
57.5%
Percentage of cells
Number of chromosomes
B
TKO
WT
Relative cell number
5hmC
D
WT
Tet TKO
Abnormal segregation events (% of total mitoses)
p-value 0.0085
9/117
19/143
17/145
30/134
R1
R2
replicate experiments
Micronuclei (% of total mitoses)
p-value 0.0072
2/117
9/143
6/145
15/134
R1
R2
replicate experiments

3.5 Acknowledgements

Chapter 3, sections 3.2.1-3.2.3, is a reprint with modifications as it appears in Paradoxical association of TET loss of function with genome-wide hypomethylation, *PNAS* (2019), published online August 1st, 2019. doi: 10.1073/pnas.1903059116. The dissertation author was the primary author of this paper. Other authors include Ageliki Tsagaratou, Hiroshi Yuita, Hyungseok Seo, Benjamin Delatte, Sven Heinz, Chris Benner & Anjana Rao.

Chapter 3, sections 3.2.4-3.2.5, is part of a manuscript under preparation. The dissertation author is the primary author of this paper along with Romain O. Georges. Other authors include Romain O. Georges, Ageliki Tsagaratou, Benjamin Delatte & Anjana Rao.

3.6 References

Alexandrov, L. B., S. Nik-Zainal, D. C. Wedge, S. A. Aparicio, S. Behjati, A. V. Biankin, G. R. Bignell, N. Bolli, A. Borg, A. L. Børresen-Dale, S. Boyault, B. Burkhardt, A. P. Butler, C. Caldas, H. R. Davies, C. Desmedt, R. Eils, J. E. Eyfjörd, J. A. Foekens, M. Greaves, F. Hosoda, B. Hutter, T. Ilicic, S. Imbeaud, M. Imielinski, M. Imielinsk, N. Jäger, D. T. Jones, D. Jones, S. Knappskog, M. Kool, S. R. Lakhani, C. López-Otín, S. Martin, N. C. Munshi, H. Nakamura, P. A. Northcott, M. Pajic, E. Papaemmanuil, A. Paradiso, J. V. Pearson, X. S. Puente, K. Raine, M. Ramakrishna, A. L. Richardson, J. Richter, P. Rosenstiel, M. Schlesner, T. N. Schumacher, P. N. Span, J. W. Teague, Y. Totoki, A. N. Tutt, R. Valdés-Mas, M. M. van Buuren, L. van 't Veer, A. Vincent-Salomon, N. Waddell, L. R. Yates, J. Zucman-Rossi, P. A. Futreal, U. McDermott, P. Lichter, M. Meyerson, S. M. Grimmond, R. Siebert, E. Campo, T. Shibata, S. M. Pfister, P. J. Campbell, M. R. Stratton, A. P. C. G. Initiative, I. B. C. Consortium, I. M.-S. Consortium & I. PedBrain (2013) Signatures of mutational processes in human cancer. *Nature,* 500**,** 415-21.

An, J., E. González-Avalos, A. Chawla, M. Jeong, I. F. López-Moyado, W. Li, M. A. Goodell, L. Chavez, M. Ko & A. Rao (2015) Acute loss of TET function results in aggressive myeloid cancer in mice. *Nat Commun,* 6**,** 10071.

Baylin, S. B. & P. A. Jones (2016) Epigenetic Determinants of Cancer. *Cold Spring Harb Perspect Biol,* 8.

Bendelac, A., P. B. Savage & L. Teyton (2007) The biology of NKT cells. *Annu Rev Immunol,* 25**,** 297-336.

Boguslawski, S. J., D. E. Smith, M. A. Michalak, K. E. Mickelson, C. O. Yehle, W. L. Patterson & R. J. Carrico (1986) Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. *J Immunol Methods,* 89**,** 123-30.

Chen, R. Z., U. Pettersson, C. Beard, L. Jackson-Grusby & R. Jaenisch (1998) DNA hypomethylation leads to elevated mutation rates. *Nature,* 395**,** 89-93.

Cimmino, L., M. M. Dawlaty, D. Ndiaye-Lobry, Y. S. Yap, S. Bakogianni, Y. Yu, S. Bhattacharyya, R. Shaknovich, H. Geng, C. Lobry, J. Mullenders, B. King, T. Trimarchi, B. Aranda-Orgilles, C. Liu, S. Shen, A. K. Verma, R. Jaenisch & I. Aifantis (2015) TET1 is a tumor suppressor of hematopoietic malignancy. *Nat Immunol,* 16**,** 653-62.

Coulondre, C., J. H. Miller, P. J. Farabaugh & W. Gilbert (1978) Molecular basis of base substitution hotspots in Escherichia coli. *Nature,* 274**,** 775-80.

Crossley, M. P., M. Bocek & K. A. Cimprich (2019) R-Loops as Cellular Regulators and Genomic Threats. *Mol Cell,* 73**,** 398-411.

Delhommeau, F., S. Dupont, V. Della Valle, C. James, S. Trannoy, A. Massé, O. Kosmider, J. P. Le Couedic, F. Robert, A. Alberdi, Y. Lécluse, I. Plo, F. J. Dreyfus, C. Marzac, N. Casadevall, C. Lacombe, S. P. Romana, P. Dessen, J. Soulier, F. Viguié, M. Fontenay, W. Vainchenker & O. A. Bernard (2009) Mutation in TET2 in myeloid cancers. *N Engl J Med,* 360**,** 2289-301.

Eden, A., F. Gaudet, A. Waghmare & R. Jaenisch (2003) Chromosomal instability and tumors promoted by DNA hypomethylation. *Science,* 300**,** 455.

Ehrlich, M. (2009) DNA hypomethylation in cancer cells. *Epigenomics,* 1**,** 239-59.

Engel, I. & M. Kronenberg (2012) Making memory at birth: understanding the differentiation of natural killer T cells. *Curr Opin Immunol,* 24**,** 184-90.

Feinberg, A. P. & B. Vogelstein (1983) Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature,* 301**,** 89-92.

Fernandez-Capetillo, O., A. Lee, M. Nussenzweig & A. Nussenzweig (2004) H2AX: the histone guardian of the genome. *DNA Repair (Amst),* 3**,** 959-67.

Gaudet, F., J. G. Hodgson, A. Eden, L. Jackson-Grusby, J. Dausman, J. W. Gray, H. Leonhardt & R. Jaenisch (2003) Induction of tumors in mice by genomic hypomethylation. *Science,* 300**,** 489-92.

Huang, Y. & A. Rao (2014) Connections between TET proteins and aberrant DNA modification in cancer. *Trends Genet,* 30**,** 464-74.

Janssen, A., S. U. Colmenares & G. H. Karpen (2018) Heterochromatin: Guardian of the Genome. *Annu Rev Cell Dev Biol,* 34**,** 265-288.

Jones, P. A. & S. B. Baylin (2002) The fundamental role of epigenetic events in cancer. *Nat Rev Genet,* 3**,** 415-28.

Ko, M., J. An & A. Rao (2015) DNA methylation and hydroxymethylation in hematologic differentiation and transformation. *Curr Opin Cell Biol,* 37**,** 91-101.

Ko, M., Y. Huang, A. M. Jankowska, U. J. Pape, M. Tahiliani, H. S. Bandukwala, J. An, E. D. Lamperti, K. P. Koh, R. Ganetzky, X. S. Liu, L. Aravind, S. Agarwal, J. P. Maciejewski & A. Rao (2010) Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature,* 468**,** 839-43.

Langemeijer, S. M., R. P. Kuiper, M. Berends, R. Knops, M. G. Aslanyan, M. Massop, E. Stevens-Linders, P. van Hoogen, A. G. van Kessel, R. A. Raymakers, E. J. Kamping, G. E. Verhoef, E. Verburgh, A. Hagemeijer, P. Vandenberghe, T. de Witte, B. A. van der Reijden & J. H. Jansen (2009) Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat Genet,* 41**,** 838-42.

Lemonnier, F., L. Couronné, M. Parrens, J. P. Jaïs, M. Travert, L. Lamant, O. Tournillac, T. Rousset, B. Fabiani, R. A. Cairns, T. Mak, C. Bastard, O. A. Bernard, L. de Leval & P. Gaulard (2012) Recurrent TET2 mutations in peripheral T-cell lymphomas correlate with TFH-like features and adverse clinical parameters. *Blood,* 120**,** 1466-9.

Lemonnier, F., E. Poullot, A. Dupuy, L. Couronné, N. Martin, L. Scourzic, V. Fataccioli, J. Bruneau, R. A. Cairns, T. W. Mak, O. A. Bernard, L. de Leval & P. Gaulard (2018) Loss of 5-hydroxymethylcytosine is a frequent event in peripheral T-cell lymphomas. *Haematologica,* 103**,** e115-e118.

Marçais, A., L. Waast, J. Bruneau, K. Hanssens, V. Asnafi, P. Gaulard, F. Suarez, P. Dubreuil, A. Gessain, O. Hermine & C. Pique (2017) Adult T cell leukemia aggressivenness correlates with loss of both 5-hydroxymethylcytosine and TET2 expression. *Oncotarget,* 8**,** 52256-52268.

Odejide, O., O. Weigert, A. A. Lane, D. Toscano, M. A. Lunning, N. Kopp, S. Kim, D. van Bodegom, S. Bolla, J. H. Schatz, J. Teruya-Feldstein, E. Hochberg, A. Louissaint, D. Dorfman, K. Stevenson, S. J. Rodig, P. P. Piccaluga, E. Jacobsen, S. A. Pileri, N. L. Harris, S. Ferrero, G. Inghirami, S. M. Horwitz & D. M. Weinstock (2014) A targeted mutational landscape of angioimmunoblastic T-cell lymphoma. *Blood,* 123**,** 1293-6.

Raffel, S., M. Falcone, N. Kneisel, J. Hansson, W. Wang, C. Lutz, L. Bullinger, G. Poschet, Y. Nonnenmacher, A. Barnert, C. Bahr, P. Zeisberger, A. Przybylla, M. Sohn, M. Tönjes, A. Erez, L. Adler, P. Jensen, C. Scholl, S. Fröhling, S. Cocciardi, P. Wuchter, C. Thiede, A. Flörcken, J. Westermann, G. Ehninger, P. Lichter, K. Hiller, R. Hell, C. Herrmann, A. D. Ho, J. Krijgsveld, B. Radlwimmer & A. Trumpp (2017) BCAT1 restricts αKG levels in AML stem cells leading to IDHmut-like DNA hypermethylation. *Nature,* 551**,** 384-388.

Schuster-Böckler, B. & B. Lehner (2012) Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature,* 488**,** 504-7.

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

van Steensel, B. & A. S. Belmont (2017) Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell,* 169**,** 780-791.

Walsh, C. P., J. R. Chaillet & T. H. Bestor (1998) Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet,* 20**,** 116-7.

Wu, D., D. Hu, H. Chen, G. Shi, I. S. Fetahu, F. Wu, K. Rabidou, R. Fang, L. Tan, S. Xu, H. Liu, C. Argueta, L. Zhang, F. Mao, G. Yan, J. Chen, Z. Dong, R. Lv, Y. Xu, M. Wang, Y. Ye, S. Zhang, D. Duquette, S. Geng, C. Yin, C. G. Lian, G. F. Murphy, G. K. Adler, R. Garg, L. Lynch, P. Yang, Y. Li, F. Lan, J. Fan, Y. Shi & Y. G. Shi (2018) Glucose-regulated phosphorylation of TET2 by AMPK reveals a pathway linking diabetes to cancer. *Nature,* 559**,** 637-641.

Xu, G. L., T. H. Bestor, D. Bourc'his, C. L. Hsieh, N. Tommerup, M. Bugge, M. Hulten, X. Qu, J. J. Russo & E. Viegas-Péquignot (1999) Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature,* 402**,** 187-91.

Zeller, P., J. Padeken, R. van Schendel, V. Kalck, M. Tijsterman & S. M. Gasser (2016) Histone H3K9 methylation is dispensable for Caenorhabditis elegans development but suppresses RNA:DNA hybrid-associated repeat instability. *Nat Genet,* 48**,** 1385-1395.

Zhu, Q., N. Hoong, A. Aslanian, T. Hara, C. Benner, S. Heinz, K. H. Miga, E. Ke, S. Verma, J. Soroczynski, J. R. Yates, T. Hunter & I. M. Verma (2018) Heterochromatin-Encoded Satellite RNAs Induce Breast Cancer. *Mol Cell,* 70**,** 842-853.e7.

Chapter 4

Concluding remarks

In this dissertation, we document an unexpected similarity between the DNA methylation patterns of diverse TET-deficient cell types and those of cancer genomes. Cancer genomes show local hypermethylation combined with widespread hypomethylation (Jones and Baylin 2002, Baylin and Jones 2016), and we reproducibly observed both features in TET-deficient cells. As expected from the biochemical activities of TET enzymes in generating oxi-mC bases and their involvement in DNA demethylation (Pastor, Aravind and Rao 2013, Wu and Zhang 2017), local DNA hypermethylation was consistently observed in the euchromatic (Hi-C A) compartment of TET-deficient cells; this compartment contains the vast majority of 5hmC, a stable modification that is most highly enriched in the gene bodies of the most highly expressed genes and at the most active enhancers (Tsagaratou et al. 2014, Hon et al. 2014). Unexpectedly, however, we also observed large domains of DNA hypomethylation in the heterochromatic (Hi-C B) compartment of diverse TET-deficient cell types, including ESC, NPC, HSPC, T cells and pro-B cells. The existence of these hypomethylated domains cannot be explained by our current understanding of TET enzymatic activity, but their presence in TET-deficient cells suggests strongly that TET proteins are required, directly or indirectly, for optimal DNMT-mediated DNA methylation in heterochromatin.

To explore the biological consequences of TET loss-of-function in vivo, we used a mouse model of profound TET deficiency in T cells. Mice with deletion of *Tet2* and *Tet3* genes in T cells showed early signal-dependent expansion and increased clonality, which rapidly progressed to an aggressive NKT cell lymphoma. The expanded *Tet2/3 DKO* NKT cells developed the same aberrations in DNA methylation – hypermethylation in the euchromatic compartment and hypomethylation in the heterochromatic compartment – that occur in cancer genomes and that we have noted above for multiple TET-deficient cell types.

Importantly, the aggressive NKT cell lymphoma develops features that are common to many types of cancers and some which have been attributed to heterochromatin dysfunction

(Janssen et al., 2018). These include (i) the accumulation of SNVs, largely in the hypomethylated heterochromatic compartment but also in the coding regions of potential cancer-associated genes, through an apparently stochastic process that differed in each individual mouse; (ii) widespread copy number variations and aneuploidies that point to chromosome instability and chromosome missgregation, which we also observe in *Tet TKO* mouse ES cells within 6 days of acute *TET* gene deletion; (iii) reactivation of transposable elements, particularly LINE and LTR elements that are primarily located in heterochromatin, recalling the genome instability produced by spurious transcription of repeat elements (Zhu et al. 2018, Zeller et al. 2016), since these repetitive elements were also more prone to mutations compared to the genome in general; (iv) rapid tumor progression characterized by stochastic cancer-specific mutations as well as recurrent events, such as mouse chromosome 17 trisomy; and (v) aberrations in DNA methylation that resemble those observed in cancer genomes. Based on these findings, we suggest that cancers related to TET loss-of-function are initiated at least partly through defects in the maintenance of heterochromatin function, with the corollary that TET proteins participate in maintaining heterochromatin integrity, either directly or indirectly.

As described in more detail below, DNA hypomethylation in heterochromatin may at least partly explain the oncogenic transformation, genome instability and DNA damage observed in diverse mouse models of partial or profound TET deficiency (Cimmino et al. 2015, An et al. 2015). The latency and penetrance of oncogenic transformation in these models likely depends on the extent of TET loss-of-function. Loss-of-function mutations in *DNMT3A* or *TET2* are associated with clonal hematopoiesis in humans (Jaiswal et al. 2014); similarly, TET deficiency in mouse models promotes the clonal expansion of TET-deficient cells. In both cases, full-blown oncogenesis requires the stochastic appearance of second hit mutations that vary from cell to cell but are subject to selection, driving clonal expansion and cancer evolution and explaining cancer heterogeneity.

The large hypomethylated domains we observe in the heterochromatic compartment of TET-deficient cells are very reminiscent of the extended, partially methylated domains (PMDs) observed in cancer genomes. Based on their overlap with nuclear lamina-associated, late-replicating domains, cancer-associated PMDs occur in the heterochromatic compartment; their presence has been attributed to ineffective DNMT-mediated remethylation of late-replicating genomic regions in rapidly-proliferating cells (Berman et al. 2011, Hon et al. 2012, Zhou et al. 2018). PMDs have also been documented in $CD4^+$ T cells from a 103-year-old individual compared to those from a newborn human (Zhou et al. 2018), suggesting that DNA methylation is also progressively lost in the heterochromatin of cells undergoing sustained long-term proliferation. While the presence of hypomethylated domains in heterochromatin of *Tet2/3 DKO* compared to WT NKT cells may indeed be a consequence of more rapid proliferation, especially since expanded *Tet2/3 DKO* NKT cells that have undergone many more cell divisions show more extensive hypomethylation than *Tet2/3 DKO* NKT cells from young mice (**Fig. 2.6F**), the PMDs observed in TET-deficient mESC cannot be explained by increased proliferation (**Fig. 2.9A**). Moreover, PMDs have also been observed in senescent IMR90 cells, which are no longer engaged in active proliferation (Cruickshanks et al. 2013). Thus, increased proliferation might contribute to, but is not the only mechanism underlying, the widespread losses of DNA methylation in heterochromatin of TET-deficient cells.

DNA hypomethylation has been associated with many biological consequences, including reactivation of transposable elements (Walsh, Chaillet and Bestor 1998), sharply increased mutation rates (Chen et al. 1998) and genome instability with chromosome segregation defects and aneuploidies (Eden et al. 2003, Gaudet et al. 2003). Mice with a hypomorphic mutation in *Dnmt1* displayed genome-wide hypomethylation in all tissues and developed T cell lymphomas that occurred in 80% of mice and were characterized by recurrent aneuploidies (Gaudet et al. 2003). Reactivation of transposable elements is prevalent in cancer genomes, and is associated

with the formation of RNA-DNA hybrids and R-loops (Zeller et al. 2016, Zhu et al. 2018), which in turn have been linked to DNA damage and the appearance of DNA double-strand breaks (Crossley, Bocek and Cimprich 2019). Each of these features was observed, together with heterochromatin hypomethylation, in expanded *Tet2/3 DKO* NKT cells. Thus in addition to their well-established role in promoting and maintaining DNA demethylation at promoters, gene bodies and enhancers, TET proteins participate in maintaining physiological levels of DNA methylation in heterochromatic compartments of the genome.

Our findings may explain the unexpected synergy between TET2 and DNMT3A mutations in humans as well as mice. TET2 and DNMT3A are recurrently co-mutated in a diverse range of myeloid and lymphoid malignancies (Ley et al. 2013, Couronné, Bastard and Bernard 2012, Odejide et al. 2014, Palomero et al. 2014, Sakata-Yanagimoto et al. 2014, Papaemmanuil et al. 2016). In a previous collaborative study, we found that the phenotypes of mice with dual *Tet2* and *Dnmt3a* deficiency in HSPC were considerably more severe than those of mice with individual *Tet2* or *Dnmt3a* deletions alone (Zhang et al. 2016). Dnmt3a and Tet2 deficiency would both result in loss of oxi-mC at specific genomic regions, through a direct decrease in DNA methylation in the case of Dnmt3a deficiency and through loss of the 5hmC substrate in the case of Tet2 deficiency. Thus the stronger defects (e.g. in erythrocyte differentiation) in *Tet2/Dnmt3a DKO* mice compared to mice with *Tet2* or *Dnmt3a* deficiency alone (Zhang et al. 2016) could potentially arise from loss of “cooperation” between DNMT3A and TET2, leading to decreased 5hmC and increased 5mC at specific euchromatic locations (promoters, gene bodies, enhancers) in both humans and mice. Based on our data, however, we speculate that pronounced DNA hypomethylation in the heterochromatic compartment of *Tet2/Dnmt3a DKO* HSPC (**Fig. 2.8**) could also be a major contributor to the observed synergy of oncogenic transformation resulting from loss-of-function mutations of both *Dnmt3a* and *Tet2* (Zhang et al. 2016).

Our reanalysis of published data suggests a potential mechanism for the synergistic actions of DNMT3A and TET proteins. TET1 and DNMT3A occupy mutually exclusive locations in the euchromatic compartment of mouse embryonic stem cells, and loss of TET proteins from euchromatin results in relocalization of DNMT3A1 to regions previously occupied by TET1 (see model in **Fig. 2.11E**). Broadly, this observation suggests that the DNMT3 enzymes responsible for de novo methylation are recruited to euchromatin through a scaffold complex or other recruitment mechanism in common with TET enzymes, but for which the DNMTs have lower affinities than TETs under normal physiological conditions. Assuming that the DNMT3 enzymes are present at limiting concentrations, loss of TET proteins would cause them to relocalize away from heterochromatin and into euchromatic regions, resulting in the observed dual loss of DNA methylation in heterochromatin and increased DNA methylation in euchromatin. This observation is consistent with the notable finding that every animal genome that encodes a DNMT also harbors at least one functional TET/JBP protein (Iyer et al. 2014). Further studies in specific cancer models will be required to advance our global understanding of the biochemistry underlying the functional interactions between TETs and DNMTs.

Our data suggest that loss of DNA methylation in heterochromatin results in "heterochromatin dysfunction" (Janssen, Colmenares and Karpen 2018). This phenomenon has many manifestations, including aneuploidies resulting from chromosome instability related to centromere dysfunction, as also observed in ICF (immunodeficiency, centromere instability, facial abnormalities) patients with germline DNMT3B mutations (Xu et al. 1999), as well as reactivation of transposable elements and increased R-loops. These features are all observed in *Tet2/3 DKO* NKT cells, as well as in cells with hypomorphic mutations in DNMT1 (Gaudet et al. 2003) (the increased copy number variations and aneuploidies observed in *Tet2/3 DKO* NKT cells will be described in a separate study). Based on these considerations, we speculate that cancers related to TET loss-of-function are initiated at least partly through defects in the maintenance of

heterochromatin function. By inference, the functional interactions between DNMT and TET proteins that we document here are likely to be important for maintaining heterochromatin integrity.

In many hematopoietic (Ko et al. 2010) and most solid cancers, TET loss-of-function is observed without coding region mutations in TET genes (Losman and Kaelin 2013, Kaelin and McKnight 2013, Huang and Rao 2014, Ko, An and Rao 2015). Early studies suggested that TET loss-of-function was secondary to TET promoter methylation, increased degradation of TET proteins, or aberrant microRNA expression (Cimmino et al. 2015, Ko et al. 2015, Raffel et al. 2017, Wu et al. 2018). More recently, however, TET loss-of-function in solid cancers has been increasingly attributed to hypoxia, or to a variety of metabolic alterations that decrease α-ketoglutarate levels or increase the levels of the D (a.k.a. *R(-)*) or L (a.k.a. *S(+)*) stereoisomers of 2-hydroxyglutarate (2HG)—increased expression of Branched chain amino acid transaminase 1 (BCAT1), dominant mutations of the isocitrate dehydrogenases *IDH1* or *IDH2*, and loss-of-function mutations of the genes encoding succinate dehydrogenase (SDH) subunits, fumarate hydratase (FH), and the L-2HG and D-2HG dehydrogenases L2HGDH and D2HGDH respectively (Thienpont et al. 2016, Raffel et al. 2017, Kaelin and McKnight 2013). It will be important to determine in the future if such mutations (e.g. *IDH1/2*) also result in heterochromatin DNA hypomethylation, and if so, if it their effect is mediated through their inhibition of TET proteins, or through their inhibition a different member of the 2-oxoglutarate Fe(II)-dependent dioxygenase superfamily, for example, JmjC-domain containing histone demethylase KDM2A, which has been previously linked to heterochromatin silencing (Frescas et al. 2008).

Similarly, it will be relevant to determine in the future if the heterochromatin DNA hypomethylation observed after TET deletion is dependent or independent of TET proteins catalytic activity. Based on our model for this DNA hypomethylation (**Fig. 2.11F**), proposing a competition for binding sites between TET and DNMT3 proteins and relocalization of DNMT3

proteins in the absence of TET proteins, one could hypothesize that TETs role in maintaining heterochromatin DNA methylation will be independent of their catalytic activity. It will be important to determine if TET-catalytic mutants also display the heterochromatin DNA hypomethylation reported in several model of TET deletion throughout this dissertation, and this question may be addressed in the coming years with the use of new catalytically inactive *Tet2* mutant mouse models (see latest work by Meelad M. Dawlaty and Keisuke Ito, September 2019).

We postulate that loss of DNA methylation in the heterochromatic compartment, and the consequent development of heterochromatin dysfunction could be the first steps in the development of many cancers characterized by TET loss-of-function. Moreover, mutations in proteins associated with the maintenance of heterochromatin integrity are frequent in cancer and many of them (e.g. *NPM1*) co-occur with TET2 mutations, leading to the postulate that heterochromatin dysfunction is not only a common feature of sporadic (non-hereditary) human cancers but also potentially an initiating event in oncogenic transformation (Janssen et al. 2018).

The methylation losses that we observe are fractional, only around 25% in our T cell lymphoma model, meaning that only a quarter of the total alleles in the transformed *Tet2/3 DKO* T cell population have lost the methyl mark at any given CpG. This heterogeneity of DNA methylation could affect the reactivation of transposable elements, the binding of methyl-sensitive proteins and transcription factors (Mazor et al. 2016, Burns 2017), thus potentially contributing to the initiating events of transformation. An interesting question that cannot be addressed with current (short reads) sequencing methods is whether DNA demethylation occurs concordantly across long regions of the genome, and if so, whether cells with broadly demethylated alleles are more subject to oncogenic transformation. In either case, however, heterogeneity of DNA methylation could make a substantial contribution to genome diversity, population heterogeneity and clonal evolution in cancer genomes.

In the last ten years, the studies on the function of TET proteins have mostly focused on their DNA demethylation activity through the production of oxi-mC. According to the work presented in this dissertation, TET loss of function can also lead to heterochromatin dysfunction, and this process could be compromise genome stability and start cells on the road to oncogenic transformation. I anticipate that future studies, some of which are in progress in our laboratory, will elucidate the mechanisms involved.

4.2 References

An, J., E. González-Avalos, A. Chawla, M. Jeong, I. F. López-Moyado, W. Li, M. A. Goodell, L. Chavez, M. Ko & A. Rao (2015) Acute loss of TET function results in aggressive myeloid cancer in mice. *Nat Commun,* 6**,** 10071.

Baylin, S. B. & P. A. Jones (2016) Epigenetic Determinants of Cancer. *Cold Spring Harb Perspect Biol,* 8.

Berman, B. P., D. J. Weisenberger, J. F. Aman, T. Hinoue, Z. Ramjan, Y. Liu, H. Noushmehr, C. P. Lange, C. M. van Dijk, R. A. Tollenaar, D. Van Den Berg & P. W. Laird (2011) Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat Genet,* 44**,** 40-6.

Burns, K. H. (2017) Transposable elements in cancer. *Nat Rev Cancer,* 17**,** 415-424.

Chen, R. Z., U. Pettersson, C. Beard, L. Jackson-Grusby & R. Jaenisch (1998) DNA hypomethylation leads to elevated mutation rates. *Nature,* 395**,** 89-93.

Cimmino, L., M. M. Dawlaty, D. Ndiaye-Lobry, Y. S. Yap, S. Bakogianni, Y. Yu, S. Bhattacharyya, R. Shaknovich, H. Geng, C. Lobry, J. Mullenders, B. King, T. Trimarchi, B. Aranda-Orgilles, C. Liu, S. Shen, A. K. Verma, R. Jaenisch & I. Aifantis (2015) TET1 is a tumor suppressor of hematopoietic malignancy. *Nat Immunol,* 16**,** 653-62.

Couronné, L., C. Bastard & O. A. Bernard (2012) TET2 and DNMT3A mutations in human T-cell lymphoma. *N Engl J Med,* 366**,** 95-6.

Crossley, M. P., M. Bocek & K. A. Cimprich (2019) R-Loops as Cellular Regulators and Genomic Threats. *Mol Cell,* 73**,** 398-411.

Cruickshanks, H. A., T. McBryan, D. M. Nelson, N. D. Vanderkraats, P. P. Shah, J. van Tuyn, T. Singh Rai, C. Brock, G. Donahue, D. S. Dunican, M. E. Drotar, R. R. Meehan, J. R. Edwards, S. L. Berger & P. D. Adams (2013) Senescent cells harbour features of the cancer epigenome. *Nat Cell Biol,* 15**,** 1495-506.

Eden, A., F. Gaudet, A. Waghmare & R. Jaenisch (2003) Chromosomal instability and tumors promoted by DNA hypomethylation. *Science,* 300**,** 455.

Frescas, D., D. Guardavaccaro, S. M. Kuchay, H. Kato, A. Poleshko, V. Basrur, K. S. Elenitoba-Johnson, R. A. Katz & M. Pagano (2008) KDM2A represses transcription of centromeric satellite repeats and maintains the heterochromatic state. *Cell Cycle,* 7**,** 3539-47.

Gaudet, F., J. G. Hodgson, A. Eden, L. Jackson-Grusby, J. Dausman, J. W. Gray, H. Leonhardt & R. Jaenisch (2003) Induction of tumors in mice by genomic hypomethylation. *Science,* 300**,** 489-92.

Hon, G. C., R. D. Hawkins, O. L. Caballero, C. Lo, R. Lister, M. Pelizzola, A. Valsesia, Z. Ye, S. Kuan, L. E. Edsall, A. A. Camargo, B. J. Stevenson, J. R. Ecker, V. Bafna, R. L. Strausberg, A. J. Simpson & B. Ren (2012) Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res,* 22**,** 246-58.

Hon, G. C., C. X. Song, T. Du, F. Jin, S. Selvaraj, A. Y. Lee, C. A. Yen, Z. Ye, S. Q. Mao, B. A. Wang, S. Kuan, L. E. Edsall, B. S. Zhao, G. L. Xu, C. He & B. Ren (2014) 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Mol Cell,* 56**,** 286-297.

Huang, Y. & A. Rao (2014) Connections between TET proteins and aberrant DNA modification in cancer. *Trends Genet,* 30**,** 464-74.

Iyer, L. M., D. Zhang, R. F. de Souza, P. J. Pukkila, A. Rao & L. Aravind (2014) Lineage-specific expansions of TET/JBP genes and a new class of DNA transposons shape fungal genomic and epigenetic landscapes. *Proc Natl Acad Sci U S A,* 111**,** 1676-83.

Jaiswal, S., P. Fontanillas, J. Flannick, A. Manning, P. V. Grauman, B. G. Mar, R. C. Lindsley, C. H. Mermel, N. Burtt, A. Chavez, J. M. Higgins, V. Moltchanov, F. C. Kuo, M. J. Kluk, B. Henderson, L. Kinnunen, H. A. Koistinen, C. Ladenvall, G. Getz, A. Correa, B. F. Banahan, S. Gabriel, S. Kathiresan, H. M. Stringham, M. I. McCarthy, M. Boehnke, J. Tuomilehto, C. Haiman, L. Groop, G. Atzmon, J. G. Wilson, D. Neuberg, D. Altshuler & B. L. Ebert (2014) Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med,* 371**,** 2488-98.

Janssen, A., S. U. Colmenares & G. H. Karpen (2018) Heterochromatin: Guardian of the Genome. *Annu Rev Cell Dev Biol,* 34**,** 265-288.

Jones, P. A. & S. B. Baylin (2002) The fundamental role of epigenetic events in cancer. *Nat Rev Genet,* 3**,** 415-28.

Kaelin, W. G. & S. L. McKnight (2013) Influence of metabolism on epigenetics and disease. *Cell,* 153**,** 56-69.

Ko, M., J. An & A. Rao (2015) DNA methylation and hydroxymethylation in hematologic differentiation and transformation. *Curr Opin Cell Biol,* 37**,** 91-101.

Ko, M., Y. Huang, A. M. Jankowska, U. J. Pape, M. Tahiliani, H. S. Bandukwala, J. An, E. D. Lamperti, K. P. Koh, R. Ganetzky, X. S. Liu, L. Aravind, S. Agarwal, J. P. Maciejewski & A. Rao (2010) Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature,* 468**,** 839-43.

Ley, T. J., C. Miller, L. Ding, B. J. Raphael, A. J. Mungall, A. Robertson, K. Hoadley, T. J. Triche, P. W. Laird, J. D. Baty, L. L. Fulton, R. Fulton, S. E. Heath, J. Kalicki-Veizer, C. Kandoth, J. M. Klco, D. C. Koboldt, K. L. Kanchi, S. Kulkarni, T. L. Lamprecht, D. E. Larson, L. Lin, C. Lu, M. D. McLellan, J. F. McMichael, J. Payton, H. Schmidt, D. H. Spencer, M. H. Tomasson, J. W. Wallis, L. D. Wartman, M. A. Watson, J. Welch, M. C. Wendl, A. Ally, M. Balasundaram, I. Birol, Y. Butterfield, R. Chiu, A. Chu, E. Chuah, H. J. Chun, R. Corbett, N. Dhalla, R. Guin, A. He, C. Hirst, M. Hirst, R. A. Holt, S. Jones, A. Karsan, D. Lee, H. I. Li, M. A. Marra, M. Mayo, R. A. Moore, K. Mungall, J. Parker, E. Pleasance, P. Plettner, J. Schein, D. Stoll, L. Swanson, A. Tam, N. Thiessen, R. Varhol, N. Wye, Y. Zhao, S. Gabriel, G. Getz, C. Sougnez, L. Zou, M. D. Leiserson, F. Vandin, H. T. Wu, F. Applebaum, S. B. Baylin, R. Akbani, B. M. Broom, K. Chen, T. C. Motter, K. Nguyen, J. N. Weinstein, N. Zhang, M. L. Ferguson, C. Adams, A. Black, J. Bowen, J. Gastier-Foster, T. Grossman, T. Lichtenberg, L. Wise, T. Davidsen, J. A. Demchok, K. R. Shaw, M. Sheth, H. J. Sofia, L. Yang, J. R. Downing, G. Eley,

et al. (2013) Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med,* 368**,** 2059-74.

Losman, J. A. & W. G. Kaelin (2013) What a difference a hydroxyl makes: mutant IDH, (R)-2-hydroxyglutarate, and cancer. *Genes Dev,* 27**,** 836-52.

Mazor, T., A. Pankov, J. S. Song & J. F. Costello (2016) Intratumoral Heterogeneity of the Epigenome. *Cancer Cell,* 29**,** 440-451.

Odejide, O., O. Weigert, A. A. Lane, D. Toscano, M. A. Lunning, N. Kopp, S. Kim, D. van Bodegom, S. Bolla, J. H. Schatz, J. Teruya-Feldstein, E. Hochberg, A. Louissaint, D. Dorfman, K. Stevenson, S. J. Rodig, P. P. Piccaluga, E. Jacobsen, S. A. Pileri, N. L. Harris, S. Ferrero, G. Inghirami, S. M. Horwitz & D. M. Weinstock (2014) A targeted mutational landscape of angioimmunoblastic T-cell lymphoma. *Blood,* 123**,** 1293-6.

Palomero, T., L. Couronné, H. Khiabanian, M. Y. Kim, A. Ambesi-Impiombato, A. Perez-Garcia, Z. Carpenter, F. Abate, M. Allegretta, J. E. Haydu, X. Jiang, I. S. Lossos, C. Nicolas, M. Balbin, C. Bastard, G. Bhagat, M. A. Piris, E. Campo, O. A. Bernard, R. Rabadan & A. A. Ferrando (2014) Recurrent mutations in epigenetic regulators, RHOA and FYN kinase in peripheral T cell lymphomas. *Nat Genet,* 46**,** 166-70.

Papaemmanuil, E., M. Gerstung, L. Bullinger, V. I. Gaidzik, P. Paschka, N. D. Roberts, N. E. Potter, M. Heuser, F. Thol, N. Bolli, G. Gundem, P. Van Loo, I. Martincorena, P. Ganly, L. Mudie, S. McLaren, S. O'Meara, K. Raine, D. R. Jones, J. W. Teague, A. P. Butler, M. F. Greaves, A. Ganser, K. Döhner, R. F. Schlenk, H. Döhner & P. J. Campbell (2016) Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med,* 374**,** 2209-2221.

Pastor, W. A., L. Aravind & A. Rao (2013) TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat Rev Mol Cell Biol,* 14**,** 341-56.

Raffel, S., M. Falcone, N. Kneisel, J. Hansson, W. Wang, C. Lutz, L. Bullinger, G. Poschet, Y. Nonnenmacher, A. Barnert, C. Bahr, P. Zeisberger, A. Przybylla, M. Sohn, M. Tönjes, A. Erez, L. Adler, P. Jensen, C. Scholl, S. Fröhling, S. Cocciardi, P. Wuchter, C. Thiede, A. Flörcken, J. Westermann, G. Ehninger, P. Lichter, K. Hiller, R. Hell, C. Herrmann, A. D. Ho, J. Krijgsveld, B. Radlwimmer & A. Trumpp (2017) BCAT1 restricts αKG levels in AML stem cells leading to IDHmut-like DNA hypermethylation. *Nature,* 551**,** 384-388.

Sakata-Yanagimoto, M., T. Enami, K. Yoshida, Y. Shiraishi, R. Ishii, Y. Miyake, H. Muto, N. Tsuyama, A. Sato-Otsubo, Y. Okuno, S. Sakata, Y. Kamada, R. Nakamoto-Matsubara, N. B. Tran, K. Izutsu, Y. Sato, Y. Ohta, J. Furuta, S. Shimizu, T. Komeno, T. Ito, M. Noguchi, E. Noguchi, M. Sanada, K. Chiba, H. Tanaka, K. Suzukawa, T. Nanmoku, Y. Hasegawa, O. Nureki, S. Miyano, N. Nakamura, K. Takeuchi, S. Ogawa & S. Chiba (2014) Somatic RHOA mutation in angioimmunoblastic T cell lymphoma. *Nat Genet,* 46**,** 171-5.

Thienpont, B., J. Steinbacher, H. Zhao, F. D'Anna, A. Kuchnio, A. Ploumakis, B. Ghesquière, L. Van Dyck, B. Boeckx, L. Schoonjans, E. Hermans, F. Amant, V. N. Kristensen, K. Peng Koh, M. Mazzone, M. Coleman, T. Carell, P. Carmeliet & D. Lambrechts (2016) Tumour hypoxia causes DNA hypermethylation by reducing TET activity. *Nature,* 537**,** 63-68.

Tsagaratou, A., T. Äijö, C. W. Lio, X. Yue, Y. Huang, S. E. Jacobsen, H. Lähdesmäki & A. Rao (2014) Dissecting the dynamic changes of 5-hydroxymethylcytosine in T-cell development and differentiation. *Proc Natl Acad Sci U S A,* 111**,** E3306-15.

Walsh, C. P., J. R. Chaillet & T. H. Bestor (1998) Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet,* 20**,** 116-7.

Wu, D., D. Hu, H. Chen, G. Shi, I. S. Fetahu, F. Wu, K. Rabidou, R. Fang, L. Tan, S. Xu, H. Liu, C. Argueta, L. Zhang, F. Mao, G. Yan, J. Chen, Z. Dong, R. Lv, Y. Xu, M. Wang, Y. Ye, S. Zhang, D. Duquette, S. Geng, C. Yin, C. G. Lian, G. F. Murphy, G. K. Adler, R. Garg, L. Lynch, P. Yang, Y. Li, F. Lan, J. Fan, Y. Shi & Y. G. Shi (2018) Glucose-regulated phosphorylation of TET2 by AMPK reveals a pathway linking diabetes to cancer. *Nature,* 559**,** 637-641.

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

Xu, G. L., T. H. Bestor, D. Bourc'his, C. L. Hsieh, N. Tommerup, M. Bugge, M. Hulten, X. Qu, J. J. Russo & E. Viegas-Péquignot (1999) Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature,* 402**,** 187-91.

Zeller, P., J. Padeken, R. van Schendel, V. Kalck, M. Tijsterman & S. M. Gasser (2016) Histone H3K9 methylation is dispensable for Caenorhabditis elegans development but suppresses RNA:DNA hybrid-associated repeat instability. *Nat Genet,* 48**,** 1385-1395.

Zhang, X., J. Su, M. Jeong, M. Ko, Y. Huang, H. J. Park, A. Guzman, Y. Lei, Y. H. Huang, A. Rao, W. Li & M. A. Goodell (2016) DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells. *Nat Genet,* 48**,** 1014-23.

Zhou, W., H. Q. Dinh, Z. Ramjan, D. J. Weisenberger, C. M. Nicolet, H. Shen, P. W. Laird & B. P. Berman (2018) DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat Genet,* 50**,** 591-602.

Zhu, Q., N. Hoong, A. Aslanian, T. Hara, C. Benner, S. Heinz, K. H. Miga, E. Ke, S. Verma, J. Soroczynski, J. R. Yates, T. Hunter & I. M. Verma (2018) Heterochromatin-Encoded Satellite RNAs Induce Breast Cancer. *Mol Cell,* 70**,** 842-853.e7.

Chapter 5

Materials and methods

External data. The external data was downloaded from Gene Expression Omnibus (GEO) and the European Nucleotide Archive (ENA). See *Table 1-2* for details on datasets.

WGBS mapping and analysis. We employed BSMAP (v2.9) (Xi and Li 2009) to align reads from bisulfite-treated samples to the mm10 mouse reference genome allowing 4 mismatches. Reads mapping to multiple locations in the reference genome with the same mapping score were removed as well as the 5' ends bearing quality lower than 20 (mapping parameters: -n 1 -v 4 -w 2 -r 0 -q 20 -R -p 8). Single and paired-end reads were mapped as appropriately.

Duplicate reads caused by PCR amplification were removed by BSeQC (v1.0.3) (Lin et al. 2013) using default parameters. An effective genome size of 1.87e9 (as suggested in BSeQC for *Mus musculus* genome) was employed to calculate maximum coverage at the same genomic location. In addition, BSeQC was employed for removing DNA methylation artefacts introduced by end repair during adaptor ligation. For paired-end sequencing, overlapping segments of two mates of a pair were reduced to only one copy to avoid considering the same region twice during the DNA methylation quantification.

To estimate CpG DNA methylation, we employed the methratio.py tool included in BSMAP (v2.9) (Xi and Li 2009), merging DNA methylation at each CpG di-nucleotide (combining CpG methylation ratios on both DNA strands). We required each CpG to be covered by at least 5 reads (merging biological replicates) in order to be considered in the downstream analysis. Only CpGs within the autosomes were considered for the analysis (no sex chromosomes included). For the window analysis and the integration with Hi-C data, we only considered for the analysis1 kb windows with at least 3 CpGs, and 10kb windows with at least 10 CpGs.

Hi-C mapping and analysis. Reads corresponding to each extreme of a fragment were trimmed after the corresponding restriction site (e.g. MboI in the case of the NKT datasets) using HOMER (Heinz et al. 2010) *homerTools trim* and independently mapped employing BWA-aln

(v0.7.13) (single-end mode) (Li and Durbin 2010). Reads were filtered out if they had a MAPQ score of less than 30, and only reads that were at least 25 bp were considered for the rest of the analysis. Only reads falling within the autosomes were considered for the analysis (no sex chromosomes included).

Hi-C analysis was performed using HOMER (Heinz et al. 2010) and its Hi-C data analysis suite. Independently-mapped reads were paired using the makeTagDirectory command, allowing only 1 tag per bp (*-tbp 1*). Reads were filtered to remove uninformative reads (contiguous genomic fragments, self-ligation, re-ligation, and reads originating from regions of unusually high tag density) and also filtered based on the distance tor restriction sites (*-genome mm10 -removePEbg -restrictionSite GATC -both -removeSelfLigation -removeSpikes 10000 5*).

To perform the principal component analysis (PCA) of Hi-C data (A/B compartment identification), we used the tool runHiCpca.pl on the normalized interaction matrix, with the options *-res 50000 -superRes 100000 -genome mm10.* For analysis involving the Hi-C A/B compartments (e.g. integration with WGBS data), only the bins associated to the same Hi-C compartment in all biological replicates (of a given sample) were considered in the analysis.

CMS-IP and TAB-seq mapping and analysis. CMS-IP data were mapped in a similar way to WGBS. Signal per 1kb window (log_2 enrichment over input) was computed using MEDIPS (Bioconductor package) (Lienhard et al. 2014), using the functions MEDIPS.createSet (with the options *extend=300, shift=0, window_size=1000 , BSgenome="BSgenome.Mmusculus.UCSC.mm10", uniq=1e-5, paired=F* for single-end data; *extend=0, shift=0, window_size=1000 , BSgenome="BSgenome.Mmusculus.UCSC.mm10", uniq=1e-5, paired=T* for paired-end data) and MEDIPS.meth (with the options *p.adj = "bonferroni", diff.method = "edgeR", minRowSum = 10, diffnorm = "tmm"*) for statistical comparisons. TAB-seq data were processed in a similar way to WGBS data.

ChIP-seq and ATAC-seq mapping and analysis. ChIP-seq and ATAC-seq data were mapped employing BWA v0.7.13 (Li and Durbin 2010). Depending on the read length and sequencing type, BWA-aln was used in single or paired-end mode to map reads that were shorter than 70 bp, and reads with length >=70 bp were mapped using BWA-mem. In both cases, *Mus musculus* genome (mm10 downloaded from UCSC website) was used as reference. Reads were filtered out if they had a MAPQ score of less than 30, and only reads that were at least 25 bp were considered for the rest of the analysis. Only reads falling within the autosomes were considered for the analysis (no sex chromosomes included). For differential enrichment or occupancy analysis, the signal per 1kb genomic window was computed using MEDIPS (Bioconductor package) (Lienhard et al. 2014), using the functions MEDIPS.createSet (with the options *extend=300, shift=0, window_size=1000 , BSgenome="BSgenome.Mmusculus.UCSC.mm10", uniq=1e-5, paired=F* for single-end data; *extend=0, shift=0, window_size=1000 , BSgenome="BSgenome.Mmusculus.UCSC.mm10", uniq=1e-5, paired=T* for paired-end data) and MEDIPS.meth (with the options *p.adj = "bonferroni", diff.method = "edgeR", minRowSum = 10, diffnorm = "tmm"*) for statistical comparisons.

Replication timing and Lamina B data. Processed data for replication timing (Hiratani et al. 2008, Peric-Hupkes et al. 2010) was downloaded from https://www2.replicationdomain.com/

RNA-seq mapping and transposable element (TE) analysis. Quality and adapter trimming was performed on raw RNA-seq reads using TrimGalore! 431 v0.4.5 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with default parameters, retaining reads with minimal length of 25 bp. Ribosomal RNA reads were filtered out using Tagdust2. Resulting reads were aligned to mouse genome mm10 using STAR v2.5.3a (Dobin et al. 2013)(Dobin et al., 2013) with alignment parameter *--outFilterMismatchNmax 4 --outFilterMultimapNmax 100 --winAnchorMultimapNmax 200*.

We employed TETranscripts (Jin et al. 2015) to quantify gene and transposable element transcript abundances. This program proportionally assigns read counts to the corresponding gene or transposable element. We used this package on mode *-multi* to be able to use ambiguously mapped reads to perform the differential expression analysis. We used the transcript annotations of the mouse genome mm10, and the repeat element annotation from UCSC RepeatMasker track of mouse genome mm10.

The DESeq2 package v1.14.1 (Love, Huber and Anders 2014) was used to normalize the raw counts and identify differentially expressed genes or transposable elements (FDR cutoff of $p<0.1$). Genes or repeat elements with less than 10 reads total were pre-filtered in all comparisons as an initial step. For total (ribodepleted) RNA-seq sample analysis, the highest expressed genes were used as control genes for size factor estimation in DESeq2. For polyA+ RNA-seq sample analysis, *p*-values of two independent experiments (same biological conditions, different library preparation methods, TruSeq and SMARTseq) were combined using the Fisher method, as implemented in the R package metaRNASeq (https://cran.r-project.org/web/packages/metaRNASeq).

Whole-genome sequencing (WGS) mapping. WGS libraries were sequenced on the Illumina Hiseq 2500 using paired-end reads at a >20X coverage per sample. Adapters and low-quality bases were trimmed before mapping, and reads with length >=70 bp were mapped to the *Mus musculus* genome (mm10 downloaded from UCSC website) using BWA-mem (Li and Durbin 2010) with default options. Optical duplicate reads were removed Picard MarkDuplicates tool.

Tumor-specific variant calling. Following GATK (McKenna et al. 2010) best practices for variant detection, additional pre-processing steps including recalibration of base quality scores were performed prior to variant detection. MuTect2 (Cibulskis et al. 2013) somatic variant caller was employed to identify single-nucleotide variants (SNVs), using matched (samples Mouse A, B and C) tail information as normal (non-tumor tissue), as well as a panel of mutations observed in

the recipient B6.SJL-PtprcaPep3bBoyJ mice (recipient mouse strain). In order to avoid false detection of tumor-specific SNVs (false positives), the variant calling process was repeated in a pairwise manner using the unmatched tails as normal (*e.g. Tail B and C for Mouse A*), and only SNVs detected in all three comparisons were included in the analysis. SNV filtering was performed using MuTect2 (Cibulskis et al. 2013) default parameters. Mutational signature analysis was performed with Bioconductor's package MutationalPatterns (Blokzijl et al. 2018). ANNOVAR (Wang, Li and Hakonarson 2010) was used to perform functional annotation of mutations (synonymous, nonsynonymous, frameshift and nonsense mutations).

Copy number analysis. HMMcopy (Ha et al. 2012) was employed to detect copy number alterations, correcting for CG content and mappability bias, employing 500kb bins (or as indicated). For WGS data, this was performed in the aligned and duplicate-removed reads; similarly for Hi-C data. For the copy number analysis on WGBS data, all cytosines were computationally converted to thymine in the sequencing reads and in the reference genome (mm10), the reads were mapped employing BWA-mem to the C-to-T converted reference genome (only one BS-read per fragment was considered for the analysis), duplicate reads were removed Picard MarkDuplicates tool.

TCR repertoire analysis. The overlapping paired-end reads (250x250) were merged into a single longer read, and ClonotypeR (18) was employed to detect clonotypes in the sequence reads, extract the CDR3 sequences and quantify TCR repertoire abundances.

Mice. Mice were housed in a pathogen-free animal facility at the La Jolla Institute. They were used according to protocols approved by the Institutional Animal Care and Use Committee (IACUC). *Tet2-/-* mice were generated by crossing *CMVCre* mice to *Tet2fl/fl* mice, in which exons 8, 9 and 10 that code for the catalytic HxD domain, were floxed (flanked by LoxP sites) (Ko et al. 2011). *Tet3fl/fl* mice were generated by targeting exon 2 (Ko et al. 2015) *Tet2-/-* and *Tet3fl/fl* mice were crossed with *CD4Cre* (Lee et al. 2001) mice to generate *Tet2-/- Tet3fl/fl CD4Cre* mice (DKO

mice). The *Tet2/3* DKO mice are in the C57BL/6 background. B6.SJL-PtprcaPep3bBoyJ (CD45.1^{+}) mice, C57BL/6 (CD45.2^{+}) mice mice were purchased from Jackson laboratory (B6(C)-*Cd1d1tm1.2Aben*/J). Both male and female mice were used in this study with similar findings. Invariant NKT cells were isolated from young mice (3-4 weeks old). The recipients were of the same sex as the donors. Both male and female recipients were used and similar results were obtained.

B6.Cg-Tg(UBC-cre/ERT2)1Ejb/J mice (Jackson Laboratory, #008085) were crossed with *Tet1*$^{Flox/Flox}$*; Tet2*$^{Flox/Flox}$*; Tet3*$^{Flox/Flox}$ and Gt(ROSA)26Sor<tm1Ytchn>/J (Jackson Laboratory, #021847) mice in order to generate Cre/ERT2$^{Wt/Ki}$; *Tet1*$^{Flox/Flox}$*; Tet2*$^{Flox/Flox}$*; Tet3*$^{Flox/Flox}$; Gt(ROSA)26Sor<tm1Ytchn>/$^{Wt/Ki}$ mice, which were time-mated with *Tet1*$^{Flox/Flox}$*; Tet2*$^{Flox/Flox}$*; Tet3*$^{Flox/Flox}$ mice so as to produce E3.75 blastocysts for mESCs derivation.

Flow Cytometry associated with NKT cell experiments. Cells were isolated from thymus, spleen, lymph nodes and bone marrow. Surface staining was performed using antibodies from Biolegend and eBioscience: CD4 (RM4-5), CD8 (53-6.7), TCRb (H57-597), B220 (RA3-6B2), CD45.1 (A20), CD45.2 (104). TCRVb2 (B20.6), TCRVb 5.1, 5.2 (MR9-4), TCRVb7 (TR310), TCRVb8.1, 8.2 (MR5-2) were purchased from BD Pharmingen. aGalCer-CD1d tetramer was obtained from the NIH Tetramer Core. Vα14*i* NKT cells were routinely defined as TCRb intermediate, B220-negative and positive for aGalCer-CD1d tetramer binding. For the pH2Ax staining the Alexa Fluor 647 anti-H2Ax-Phosphorylated (Ser139) (clone 2F3)(Biolegend) was used. Acquisition was performed in a BD LSR Fortessa (BD Biosciences) using the BD FACSDiva Software. Data analysis was performed with FlowJo (Treestar).

Flow cytometry associated with mESC experiments. After dissociation, cells were resuspended in sterile ice-cold PBS 1X containing 0.5% BSA. Cells were labeled with a Fixable Viability Dye and stained for surface markers: SSEA-1 (BD Biosciences, cat. no. 562705), Thy1.2 (Biolegend, cat. no. 140319) and CD326 (Biolegend, cat. no. 118227). Subsequently, cells were

fixed in PBS 1X + PFA 4%, rinsed, and were permeabilized in PBS 1X + Triton X-100 0.1% for 5-10 min. After rinsing the cells once with PBS 1X., the DNA was denatured by treating the fixed ESCs with HCl 3.5 N for 12 min. Following the removal of HCl, the pH was neutralized with Tris-HCl 100 mM (pH 8.5) for 10 min. at room temperature. mESCs were rinsed in PBS 1X with BSA 0.5%, blocked in PBS 1X, 2% BSA, 2% Normal Donkey Serum, and 0.02% Tween-20, prior to the addition of the primary antibody. mESCs were washed thrice with PBS 1X, 0.5% BSA, and 0.02% Tween-20, before adding the secondary antibody. Acquisition was performed using a BD LSR Fortessa (BD Biosciences) and the BD FACSDiva software. Data analysis was performed with FlowJo software.

Isolation of Vα14*i* NKT cells. Vα14*i* NKT-cell preparations for adoptive transfer and DNA isolation experiments were performed using in case of control mice a pool of cells (isolated from thymus or spleen as indicated on each case) from C57BL/6 mice and from age- and sex-matched DKO mice. For fluorescence-activated cell sorting (FACS), cells from wild type mice were depleted of $CD19^+$ (6D5), TER-119^+ (TER119), $CD8^+$ (53-6.7), $CD11c^+$ (N418), $F4/80^+$ (BM8) and $CD11b^+$ (M1/70) cells using biotinylated antibodies (Biolegend) and subsequent binding to magnetic streptavidin beads (Life Technologies). The unbound cells were incubated with 1 mg/ml Streptavidin A (Sigma Aldrich) and subsequently stained with αGalCer-loaded CD1d tetramers and anti-TCRβ, after which tetramer-binding, $TCR\beta^+$ cells were isolated using a FACSAria cell sorter (BD Biosciences). To obtain DKO cells, no depletion was performed since NKTs are massively expanded. Rather, B220-, tetramer-binding, TCRβ+ cells were isolated using a FACSAria cell sorter (BD Biosciences).

Adoptive transfer experiments. NKT-sorted cells were transferred retro-orbitrally to non-irradiated, fully immune-competent congenic (B6.SJL-PtprcaPep3bBoyJ) ($CD45.1^+$) mice.

TCR repertoire sequencing. Va14*i* NKT cells were isolated by FACS from wild type and *Tet2/3* DKO young mice or were magnetically purified by recipients of *Tet2/3 DKO* NKT cells.

RNA was isolated with the E.Z.N.A. HP Total RNA kit (Omega) according to the manufacturer's instructions. cDNA was prepared using Superscript III (Invitrogen). Subsequently, PCR was performed for amplification of the gene segments with specific forward primers (*sequences shown below*) for Vb8.1 (primer MuBV8.1N), Vb8.2 (primer MuBV8.2N) and Vb8.7 (primer MuBV7) regions and a reverse primer for the b chain constant region (primer MuTCB3C). Amplicons were quantified and pooled using HS Qubit (Life Technologies). Adaptors (NEB) were ligated and libraries were amplified using Kapa HiFi (Kapa Biosystems). Amplified libraries were quantified using HS Qubit, their size was evaluated using Bioanalyzer and sequenced in an Illumina Miseq.

MuBV8.1N	GGC TGA TCC ATT ACT CAT ATG TC
MuBV8.2N	TCA TAT GGT GCT GGC AGC ACT G
MuBV7	TAC AGG GTC TCA CGG AAG AAG C
MuTCB3C	AAG CAC ACG AGG GTA GCC T

Whole-genome bisulfite sequencing (WGBS) Library preparation. $V_{\alpha}14i$ NKT cells were isolated by flow cytometry and DNA was isolated using the PureLink genomic DNA mini kit (Life technologies). DNA was fragmented. 1.5 µg of the fragmented DNA was used for the library preparation and bisulfite treatment was done as described in *ref. 26*. After the bisulfite conversion the purified DNA was amplified for 4 cycles (low amplification) using Kapa HiFi Uracil^{+} (Kapa Biosystems). 2 independent WGBS samples per genotype were evaluated.

Whole Genome Sequencing (WGS) Library preparation. Genomic DNA was isolated from purified NKT cells using the PureLink genomic DNA mini kit (Life technologies). DNA was fragmented to an average size of 400 bp using the Adaptive Focused Acoustics Covaris S2 instrument. Libraries were prepared using the TruSeq DNA PCR-Free Sample Preparation kit (Illumina) according to the manufacturer's guidelines. Libraries were purified, pooled according to the instructions of the manufacturer and sequenced in an Illumina HiSeq 2500 instrument.

Hi-C Library preparation. Between 0.6 and 1.5x10^6 NKT cells were fixed in complete medium containing 1% Formaldehyde, then quenched with 125 mM glycine and washed twice with an excess of PBS. Cells were then resuspended in lysis buffer containing 0.5% SDS and lysed at 62°C for 7 minutes. This step also allows to remove proteins that were not fixed to the chromatin. SDS was further quenched with 1% Triton-X-100 at 37°C for 15 minutes. Next, permeabilized nuclei were reacted with 100 units of MboI overnight at 37°C. After subsequent washing and inactivation of MboI, the restriction sites were further filled in with Biotin-14-dATP and Klenow polymerase at room temperature for 40 minutes. Samples were transferred into a ligation solution containing 600 units of T4 DNA ligase. Proximity ligation was stopped by addition of 2-fold molar excess of EDTA, and samples were decrosslinked at 65°C for 16h00. DNA was further purified by proteinase K digestion and phenol/chloroform extraction. For library preparation, 800 ng of DNA was sonicated to an average of 300bp fragments length, and was used for subsequent library preparation that includes blunting of DNA, A-tailing, ligation of sequencing adapters, and amplification of library.

Total (ribodepleted) RNA-seq Library preparation. 10 million cells were sorted and then whole RNA was isolated using the RNeasy Plus Mini Kit (Qiagen). Ribo-zero RNA-seq libraries were prepared using the TruSeq Stranded Total RNA Library Prep Gold kit (Illumina) with minor modifications. The starting RNA was 800 ng. Ribosomal RNAs were depleted using magnetic beads. Next, RNA was fragmented, and cDNA was synthesized using Superscript II (Invitrogen). After A-tailing and adaptor ligation, libraries were generated by amplifying the cDNA for 12 cycles.

Cell Culture and Isolation of EGFP+ mESCs. ESCs were grown in Serum Replacement ESC media (SRES media) composed of KnockOut™ DMEM/F-12 (Gibco, cat. no. 12660012) supplemented with 15% KnockOut™ Serum Replacement (KSR) (Gibco, cat. no. 10828028), GlutaMAX™-I (Gibco, cat. no. 35050061), 1x MEM non-essential amino acids (Gibco, cat. no.

11140-035), 0.1 mM b-mercaptoethanol, LIF, and with 2i (3 mM CHIR99021 and 0.4 mM PD0325901) over a layer of mitomycin C-treated mouse embryonic fibroblasts (MEFs).

To inducibly delete TET enzymes, ESCs plated on mitotically inactivated MEFs were treated with 1.4 uM 4-hydroxymethyltamoxifen (4OH-TAM) for 2.5 days. After 1.5 additional days of growth in SRES without 4OH-TAM, H2B-EGFP+ mESCs, resulting from a successful Cre-mediated recombination that occurred during 4OH-TAM treatment, were sorted by FACS (Fluorescence-Activated Cell Sorting) using a FACSAria cell sorter (BD Biosciences). ESCs were plated on mitotically inactivated MEFs and allowed to attach for 6 hours prior to initiating the imaging of the cells for 48 hours.

ESCs Derivation. Mouse ESCs were derived from E3.75 blastocysts flushed out from the uterine horn of time-mated females. Blastocysts were collected and hatched on mitomycin C mitotically inactivated MEFs plated 1 day earlier. After attachment, SRES media was replaced every second day until day 6 after attachment of the blastocyst to the feeder MEFs layer. On day 6, the expanded blastocysts were dissociated using TrypLE™ Select Enzyme (10X) (Gibco, cat. no. A1217702) and dissociated cells were plated on mitotically inactivated MEFs. Cells were grown and SRES media was changed every second day until compact cell colonies with typical ESC colony morphology formed. mESCs were passaged 2 to 3 times before being cryopreserved.

Live-Cell Microscopy. To study chromosome segregation by live-cell imaging, mESCs were seeded on mitotically inactivated MEFs into an 96-well high optical quality plastic plates (Greiner Bio-One) and imaged using a Confocal Quantitative Image Cytometer CQ1 benchtop high-content analysis system (Yokogawa) with a 40x 0.95 NA U-PlanApo objective and 2,560 x 2,160-pixel sCMOS camera at 2 x 2 binning. 20-30 fields/well with 6 x 2 mm z-sections per field in GFP channel were captured at 6 minutes intervals for 48 hours. Chromosome segregation was manually analyzed using ImageJ.

Metaphase Spread and G-banding. FACS- or MACS-sorted iNKT cells (Wt and Tet2/3 DKO) were cultured in IMDM, GlutaMAX, FBS, 1% penicillin/streptomycin, b-mercaptoethanol, and IL-17 at 37°C in 10% CO_2 for 5 hours. Karyomax at a final concentration of 0.1 µg/mL was then added to the flask for 1 hour at 37°C in 10% CO_2, before harvesting and rinsing iNKT cells twice in PBS 1X. Cells were swelled in pre-warmed buffered hypotonic solution (Hepes 5 mM + KCl 65 mM) for 12 minutes at 37°C. iNKTs were then pelleted and fixed in fixative solution (methanol:glacial acetic acid, 3:1). The fixative solution was changed twice before storing the cells at -20°C overnight. On the next day, cells were resuspended in a small volume of fixative solution following two rinses. 30 µL of cell suspension were dropped onto clean dry glass slides, that were immediately exposed, face up, into the stream of hot steam (90°C) for 30 seconds, which caused the cells to blow up as the fixative solution evaporated. Metaphase spreads were aged at room temperature for ~10-15 days before performing G-banding. Suitably aged slides were incubated in 2X SSC at 65°C for 1H30. Slides were transferred into 0.85% (w/v) NaCl at room temperature for 5 min., then in 0.85% NaCl + 0.025% trypsin for 15-20 sec. Tryptic activity was stopped by placing the slides back into 0.85% NaCl, followed by 2 rinses in phosphate buffer 20 mM (pH 6.8). Slides were stained for with fresh Giemsa stain in 5 mM phosphate buffer (pH 6.8) for 10 min. Slides were quickly rinsed in phosphate buffer (pH 6.8) and blow-dried. Slides were mounted with Permount and photos were taken with a 100X oil-immersion lens.

Statistical Analysis. For mouse experiments, Mantel-Cox test and Gehan-Brenslow-Wilcoxon test were applied as indicated and the p values are shown for each figure. Statistical evaluations were performed using the unpaired t test. Data are mean $\pm$ SEM. Asterisks indicate statistically significant differences: ****P< 0.0001, ***P< 0.001, **P< 0.01, *P<0.05. If not otherwise indicated the p value was not statistically significant (p>0.05). In the graphs each dot represents a mouse. For all the experiments we used sufficient number of mice to ensure adequate power for our conclusions. Mice from different litters and of different sex were evaluated.

In addition, we ensured that a minimum of 2 independent experiments was performed in each case. No blinding was applied in this study. For the two-sample Kolmogorov-Smirnov test related to methylation analysis, the *D* statistic and pvalues were calculated using the *ks.test* function as implemented in R. In all tests, the alternative hypothesis is that CDF of WT lies below that of *TET*.

5.2 References

Blokzijl, F., R. Janssen, R. van Boxtel & E. Cuppen (2018) MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med,* 10**,** 33.

Cibulskis, K., M. S. Lawrence, S. L. Carter, A. Sivachenko, D. Jaffe, C. Sougnez, S. Gabriel, M. Meyerson, E. S. Lander & G. Getz (2013) Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol,* 31**,** 213-9.

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

Ha, G., A. Roth, D. Lai, A. Bashashati, J. Ding, R. Goya, R. Giuliany, J. Rosner, A. Oloumi, K. Shumansky, S. F. Chin, G. Turashvili, M. Hirst, C. Caldas, M. A. Marra, S. Aparicio & S. P. Shah (2012) Integrative analysis of genome-wide loss of heterozygosity and monoallelic expression at nucleotide resolution reveals disrupted pathways in triple-negative breast cancer. *Genome Res,* 22**,** 1995-2007.

Heinz, S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh & C. K. Glass (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell,* 38**,** 576-89.

Hiratani, I., T. Ryba, M. Itoh, T. Yokochi, M. Schwaiger, C. W. Chang, Y. Lyou, T. M. Townes, D. Schübeler & D. M. Gilbert (2008) Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol,* 6**,** e245.

Jin, Y., O. H. Tam, E. Paniagua & M. Hammell (2015) TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics,* 31**,** 3593-9.

Ko, M., J. An, W. A. Pastor, S. B. Koralov, K. Rajewsky & A. Rao (2015) TET proteins and 5-methylcytosine oxidation in hematological cancers. *Immunol Rev,* 263**,** 6-21.

Ko, M., H. S. Bandukwala, J. An, E. D. Lamperti, E. C. Thompson, R. Hastie, A. Tsangaratou, K. Rajewsky, S. B. Koralov & A. Rao (2011) Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proc Natl Acad Sci U S A,* 108**,** 14566-71.

Lee, P. P., D. R. Fitzpatrick, C. Beard, H. K. Jessup, S. Lehar, K. W. Makar, M. Pérez-Melgosa, M. T. Sweetser, M. S. Schlissel, S. Nguyen, S. R. Cherry, J. H. Tsai, S. M. Tucker, W. M. Weaver, A. Kelso, R. Jaenisch & C. B. Wilson (2001) A critical role for Dnmt1 and DNA methylation in T cell development, function, and survival. *Immunity,* 15**,** 763-74.

Li, H. & R. Durbin (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics,* 26**,** 589-95.

Lienhard, M., C. Grimm, M. Morkel, R. Herwig & L. Chavez (2014) MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. *Bioinformatics,* 30**,** 284-6.

Lin, X., D. Sun, B. Rodriguez, Q. Zhao, H. Sun, Y. Zhang & W. Li (2013) BSeQC: quality control of bisulfite sequencing experiments. *Bioinformatics,* 29**,** 3227-9.

Love, M. I., W. Huber & S. Anders (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol,* 15**,** 550.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly & M. A. DePristo (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res,* 20**,** 1297-303.

Peric-Hupkes, D., W. Meuleman, L. Pagie, S. W. Bruggeman, I. Solovei, W. Brugman, S. Gräf, P. Flicek, R. M. Kerkhoven, M. van Lohuizen, M. Reinders, L. Wessels & B. van Steensel (2010) Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol Cell,* 38**,** 603-13.

Wang, K., M. Li & H. Hakonarson (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res,* 38**,** e164.

Xi, Y. & W. Li (2009) BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics,* 10**,** 232.