

Clonal copy-number mosaicism in autoreactive T lymphocytes in diabetic NOD mice

Maha Alriyami,^{1,2} Luc Marchand,¹ Quan Li,^{1,3} Xiaoyu Du,¹ Martin Olivier,⁴ and Constantin Polychronakos¹

¹The Endocrine Genetics Laboratory, Child Health and Human Development Program and Department of Pediatrics, McGill University Health Centre Research Institute, Montreal, Quebec H3H 1P3, Canada; ²Department of Biochemistry, College of Medicine and Health Sciences, Sultan Qaboos University, 123, Muscat, Oman; ³Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, ON M5G 2C1, Canada; ⁴Departments of Medicine, Microbiology, and Immunology, McGill University Health Centre Research Institute, Montreal, Quebec H3H 1P3, Canada

Concordance for type 1 diabetes (T1D) is far from 100% in monozygotic twins and in inbred nonobese diabetic (NOD) mice, despite genetic identity and shared environment during incidence peak years. This points to stochastic determinants, such as postzygotic mutations (PZMs) in the expanding antigen-specific autoreactive T cell lineages, by analogy to their role in the expanding tumor lineage in cancer. Using comparative genomic hybridization of DNA from pancreatic lymph-node memory CD4⁺ T cells of 25 diabetic NOD mice, we found lymphocyte-exclusive mosaic somatic copy-number aberrations (CNAs) with highly nonrandom independent involvement of the same gene(s) across different mice, some with an autoimmunity association (e.g., *Ilf3* and *Dgka*). We confirmed genes of interest using the gold standard approach for CNA quantification, multiplex ligation-dependent probe amplification (MLPA), as an independent method. As controls, we examined lymphocytes expanded during normal host defense (17 NOD and BALB/c mice infected with *Leishmania major* parasite). Here, CNAs found were fewer and significantly smaller compared to those in autoreactive cells ($P = 0.0019$). We determined a low T cell clonality for our samples suggesting a prethymic formation of these CNAs. In this study, we describe a novel, unexplored phenomenon of a potential causal contribution of PZMs in autoreactive T cells in T1D pathogenesis. We expect that exploration of point mutations and studies in human T cells will enable the further delineation of driver genes to target for functional studies. Our findings challenge the classical notions of autoimmunity and open conceptual avenues toward individualized prevention and therapeutics.

[Supplemental material is available for this article.]

Type 1 diabetes (T1D) is an autoimmune disease caused by targeted destruction of the insulin-producing beta cells through infiltration of autoreactive T lymphocytes (Polychronakos and Li 2011). The disease is antigen-specific, in which this autoimmune process of infiltration destroys only the insulin-producing beta cells. Although T1D is known to depend on both inherited susceptibility and environmental factors, these alone may not explain all of the disease. Concordance in monozygotic twins is only 65% and age of onset can differ by several decades (Redondo et al. 2008). A shared environment in early life, at the onset in the first twin, also raises some doubt about whether environment accounts for this difference (Knip et al. 2005). Similarly, in the inbred nonobese diabetic (NOD) mouse model, not all females develop the disease and males have an incidence of <50% despite being genetically identical and kept in a standardized environment (Makino et al. 1980). These observations suggest stochastic events. One plausible such event could consist of postzygotic genetic changes in the expanding antigen-specific autoreactive T cell lineages.

The hypothesis suggesting the contribution of postzygotic mutations (PZM) in the pathogenesis of autoimmune diseases was first put forward in 1972 by Burnet (Burnet 1972), who proposed that the stochastic nature of autoimmune diseases might

be caused by a combination of germline and somatic mutations that interrupt normal mechanisms for eliminating self-reactive lymphocytes and causing the development of “forbidden clones.” The hypothesis was proposed again in 2007 by Goodnow (Goodnow 2007), who hypothesized a major contribution of PZMs in the pathogenesis of autoimmune diseases, in a paradigm similar to the pathogenesis of cancer. In 2004, Holzelova (Holzelova et al. 2004) detected heterozygous dominant *Fas* mutations in a fraction of T cells of sporadic cases of the autoimmune lymphoproliferative syndrome (ALPS) without the development of lymphoma. This condition follows the conventional two-hit cancer model, with the somatic mutation compounding one inherited on the opposite allele (Dowdell et al. 2010; Magerus-Chatinet et al. 2011). Here, we hypothesized that the phenomenon applies more generally in autoimmunity and involves modulation (not necessarily complete loss of function) of multiple genes.

In blood cells, PZMs (copy-number or point mutations) result in a mosaic state that can occasionally be detected in the peripheral whole blood of healthy individuals (Forsberg et al. 2012; Jacobs et al. 2012; Laurie et al. 2012). These findings almost certainly underestimate the frequency of these events in the general population, as peripheral whole-blood is a heterogeneous mixture, within which the PZM mosaicism is too low to cause a clinical phenotype or to be detectable by conventional methods (Jacobs et al.

Corresponding author: constantin.polychronakos@mcgill.ca

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.247882.118>. Freely available online through the *Genome Research* Open Access option.

© 2019 Alriyami et al. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

2012). PZM frequency increases with age, indicating that their rise to detectable levels is due to some proliferation/survival advantage. The contribution of copy-number somatic mutations in the pathogenesis of cancer has been established and has enabled therapeutic advances.

In this study, we investigated the PZM hypothesis as part of the cause of diabetes in NOD mice, a model of spontaneous insulinitis that closely recapitulates the destruction of the beta cells by autoreactive CD4⁺ and CD8⁺ T cells in T1D (Polychronakos and Li 2011; Pearson et al. 2016). Similar to human T1D, diabetes in NOD mice is caused by a combination of polygenic inheritance and environmental factors (Polychronakos and Li 2011; Pearson et al. 2016). Female mice are predominantly affected (90%–100%), while males develop it at an older age with lower frequency. We hypothesize that PZMs cause T cells to escape self-tolerance checkpoints, with expansion of autoreactive lineages. As a first step, we focused on copy-number mutations, since they are known to occur in blood cells and have more obvious functional effects.

Results

To test the hypothesis, we compared the genomes of the autoreactive lymphocytes (memory cells from pancreatic lymph nodes [PLN]) to the germline genome (tail DNA) of the diabetic mice by comparative genomic hybridization (CGH) to identify somatic copy-number mutations in the proliferating immune cells.

In our vivarium, 90%–100% of the female NOD mice and about 50% of males become diabetic (Supplemental Fig. S1). We euthanized NOD mice, within 2–4 d of hyperglycemia onset and obtained the PLNs and a tail-clip sample. Memory CD4⁺ cells from PLNs were isolated on a FACSaria sorter based on the SELL^{neg}/CD44^{hi} phenotype (Fig. 1A,B). Average yield was 124,672 CD4⁺ cells per mouse. In six out of the 25 mice examined, all T cells were negative for selectin, lymphocyte (SELL, also known as CD62L), indicating a strong preponderance of memory cells in the PLN (Fig. 1C,D). DNA from these cells was analyzed by CGH, with tail DNA as reference.

Positive and negative controls

CGH data of all samples revealed the expected deletions at the T cell receptor loci (B, G, and A–D subunits) due to genomic rearrangements to create TCR diversity in lymphocytes—a naturally occurring positive control (Fig. 2).

Technical controls evaluating the specificity of the method demonstrated an absence of amplification artifacts in CNA calls. As expected, no CNAs were called when the same sample was used as test and reference due to the identity of the two samples (Sample M26; DLRSpread value: 0.17). Moreover, no CNAs were

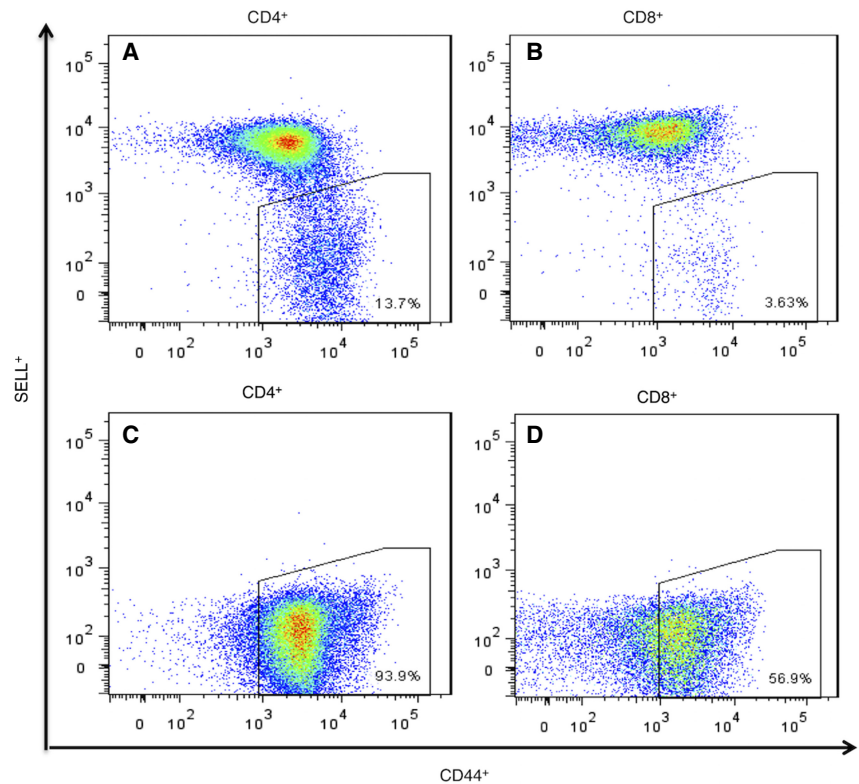


Figure 1. Flow-sorting of memory CD4⁺ and CD8⁺ T cells from NOD mouse PLNs using the established surface-marker phenotype SELL^{neg}/CD44^{hi}. (A,B) Normal profile observed in most samples of CD4⁺ and CD8⁺ T cells ($n=19$). (C,D) Profile of sorted memory cells in six animals ($n=6$) in which all cells were SELL^{neg} with CD44⁺ in the lower part of the range for memory cells.

called for the duplicate samples M9 and M9B (DLRSpread value: 0.18 and 0.21, respectively).

To evaluate the sensitivity of CGH for mosaic CNAs, we evaluated mixtures of NOD and C57BL6/J tail DNA against C57BL6/J as a reference. Known CNVs between the two strains were detectable and statistically significant even at 25% mosaicism (Fig. 3A) and correlated linearly with the Cy3/Cy5 ratio (Fig. 3B; Supplemental Data S1), allowing a rough estimation of mosaicism.

Recurrent postzygotic copy number changes in pancreatic lymph nodes of T cells of different mice

The 25 tested mice (22 females and three males) (Supplemental Table S1) had a total of 403 somatic CNAs (Supplemental Data S2, S3). Four of the mice were also investigated for CNAs in CD4⁺ from the peripheral LN (M16L, M17L, M20L, and M21L). No CNAs were called in samples obtained from the peripheral LN of these mice. The number of CNAs per mouse varied widely. M23 and M25 (two of the mice with memory-only cells in PLNs) had an unusually large number of CNAs.

We found a total of 83 overlaps between mice, most of them involving the two outliers, M23 and M25. Not counting CNAs exclusively between M23 and M25, we identified eight recurrent CNAs that occurred independently in two or more mice (Table 1; Fig. 4; Supplemental Fig. S3). This is very unlikely to happen by chance alone (none of these loci overlap with the CNVs that we observed between NOD and C57BL6/J). After 100,000 iterations of randomizing the genomic coordinates of the probes defining

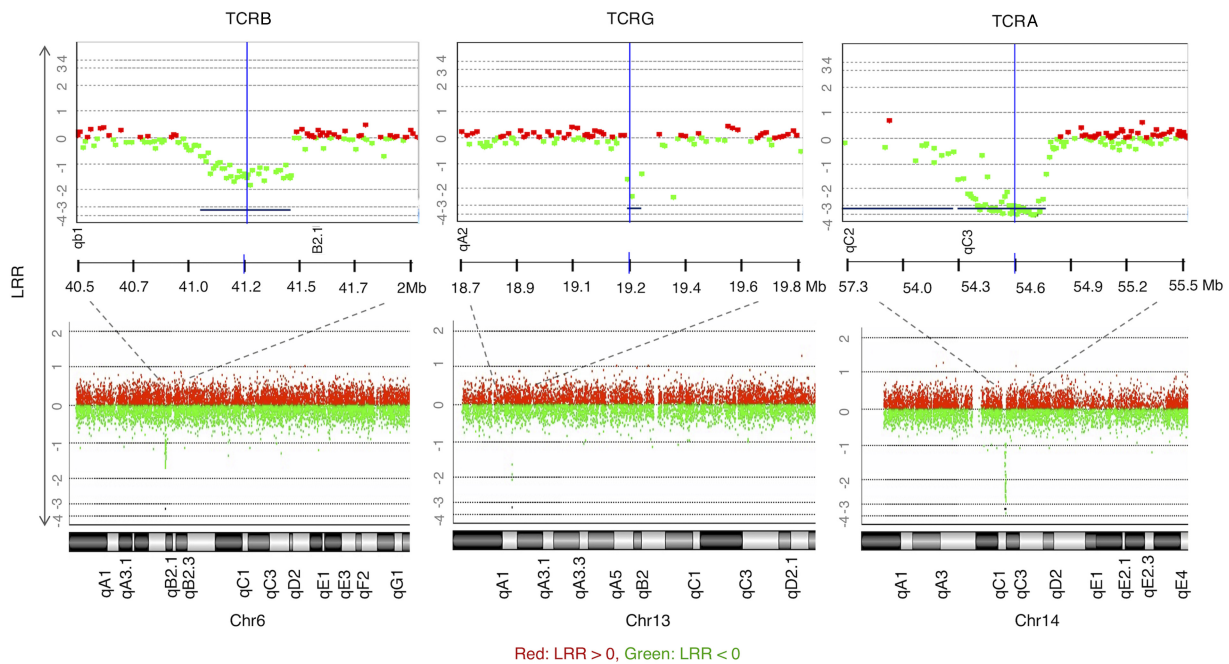


Figure 2. CGH array results demonstrating positive controls for T cell CNAs observed in all samples (to illustrate, M1 is shown here), copy loss (regions represented by lines below the graph) at regions of TCR genomic rearrangements of the β -chain on Chr 6, γ -chain on Chr 13, and α -chain on Chr 14 (DLRspread value: M1: 0.21). LRR (log relative ratio) is the base 2 log of the Cy3/Cy5 ratio.

these CNAs, not a single one gave the number of recurrent CNAs observed or more ($P < 10^{-6}$). The same result was found after removing M23 and M25, responsible for many of the overlaps. This nonrandom recurrence supports causal-functional importance and, in the cancer paradigm, it is considered a strong indicator of pathogenicity. Many of these overlaps involve genes related to immune function.

Two of the recurrent CNAs span Homeobox clusters (*Hoxd* and *Hoxa*). Both are copy losses at Chr 2: 74,474,401–74,598,078, spanning the *Hoxd* gene cluster and at Chr 6: 52,116,749–52,207,847, spanning the *Hoxa* gene cluster, both found independently in CD4⁺ T cells of two separate mice M4 and M12 (Table 1; Supplemental Fig. S3A,C). *Hox* genes control proliferation and differentiation of hematopoietic cells and have a crucial role in T cell development, being the most active cluster during early hematopoiesis (Lawrence et al. 1996; Magli et al. 1997; Tachon et al. 2002, 2003). They are implicated in leukemogenesis (Kawagoe et al. 1999) and act as tumor-suppressors (Shah and Sukumar 2010).

Another recurrent CNA is a copy gain at Chr 3: 94,767,165–94,821,010 (Table 1; Supplemental Fig. S3B) seen independently in three mice M23, M25, and M22. This CNA spans *Pi4kb* encoding phosphatidylinositol 4-kinase beta, highly expressed in hematopoietic cells (Wu et al. 2016). It phosphorylates phosphatidylinositol (PI) as the first step in the production of inositol-1, 4, 5, -triphosphate (Ins(1,4,5)P₃), an essential signaling molecule (Heilmeyer et al. 2003). Overexpression of *Pi4kb* due to the amplification observed might be contributing to autoimmunity.

Another recurrent CNA at Chr 8: 72,706,464–72,777,429 is a copy gain in two diabetic mice (M3 and M10) (Table 1; Supplemental Fig. S3D). A common affected gene is *Armc6*, with coding sequence within the confidence interval of the breakpoint locations, indicating a high likelihood of disruption of the gene by

the duplications. *Armc6* is highly conserved with Gene Ontology (GO) that suggests involvement in hematopoietic progenitor cell differentiation (GO:0002244) and in pancreatic cancer (Gress et al. 1996).

Copy gain at Chr 11: 68,880,665–68,909,840 (Table 1; Fig. 4A) was found in three mice (M22, M23, and M25). This CNA spans two genes, *Tmem107* (transmembrane protein 107) and *Vamp2* (vesicle-associated membrane protein 2). *Tmem107* encodes a transmembrane protein 107, an uncharacterized gene that is highly expressed in T cells (<http://ist.medisapiens.com>). Another transmembrane encoding gene (*Igflr1*) was also included in a recurrent multicopy amplification at Chr 7: 31,316,157–31,433,964 in M25 (Supplemental Data S3) spanning several genes, including some with high expression in hematopoietic cells, such as *Igflr1*, *Lin37*, *Cox6b1*, and *Rbm42* (Wu et al. 2016). *Igflr1* encodes the transmembrane protein IGF-like family receptor 1 that is exclusively expressed in hematopoietic cells (Wu et al. 2016) and, in mice, primarily on T cells (Lobito et al. 2011). This gene is also uncharacterized, but it is structurally similar to the tumor necrosis factor receptor family (Lobito et al. 2011), and a potential inflammatory role was suggested for its encoded protein (Lobito et al. 2011).

A recurrent CNA at Chr 14: 52,061,318–52,145,237 in mice M5 and M22 is a copy loss that spans two unstudied genes *Gm7247* and *Vmn2r89* (Table 1; Fig. 4B). These two genes are also spanned by copy losses in mice M6 and M25 that were called by ADM-2 but narrowly missed the CBS cut-off (Supplemental Fig. S4A).

Two CNAs in three of the mice with the unstable genome (M22, M23, and M25) are copy gains at two histone loci at Chr 13: 21,813,250–21,921,804 and Chr 13: 23,648,864–23,661,397 (Table 1; Fig. 4C), spanning two clusters of highly homologous genes that encode nuclear proteins collectively referred to as

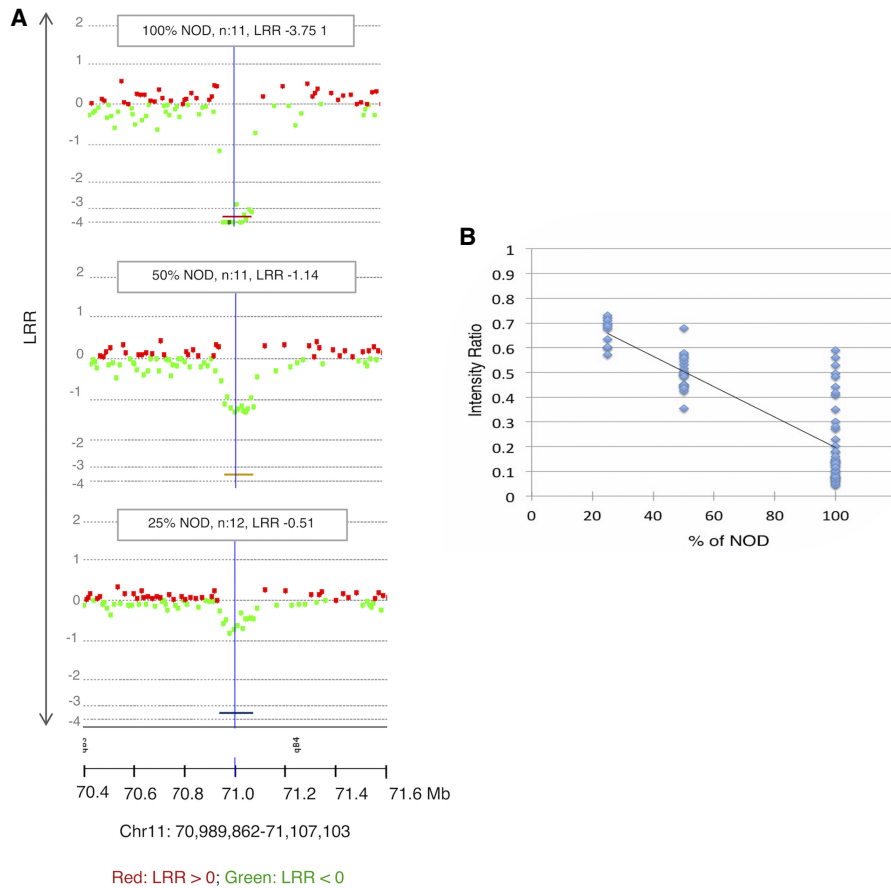


Figure 3. Detection of mosaicism. (A) Example of results of the three mixtures of germline NOD and C57BL/6J at the indicated percentages against 100% C57BL/6J as a reference. A known germline CNV between the two strains at Chr 11: 70,989,862–71,107,103 (region represented by colored lines below the charts) is significant even at 25%. (B) The intensity ratio of NOD copy losses is significantly lower than the baseline represented by 1, allowing estimates of mosaicism even <25% (DLRspread values: 100% NOD: 0.23; 50% NOD: 0.20; 25% NOD: 0.15); *n* = number of probes.

Histone H1, H2, and H3. The histone loci are separated by a normal copy-number stretch encoding other genes. In human CD4⁺ T cells, TCR activation induces a distinctive promoter histone profile that directs the transcriptional activation of *FasI* leading to apoptosis (Ghare et al. 2014).

Finally, the last recurrent CNA is a copy loss at Chr 5: 23,846,753–24,135,248 found independently in M16 and M12 (Supplemental Fig. S4B). It is mentioned because it includes an important candidate gene, despite the fact that it missed (*P* = 0.0018) the strict significance threshold in one of the two algorithms (CBS) in one of the two mice (M16). This CNA spans several genes that are highly expressed in hematopoietic cells, including *Kcnh2*, *Fastk*, *Tmub1*, and *Chpf2* (Wu et al. 2016). *Fastk* encodes a protein that belongs to the serine/threonine protein kinase family, and it is a potent activator of lymphocytes apoptosis (Tian et al. 1995). Moreover, it has been shown that this protein is highly activated during FAS-mediated apoptosis (Tian et al. 1995; Izquierdo and Valcárcel 2007). The observed copy loss might have facilitated escape of self-reactive lymphocytes from proliferation and apoptosis checkpoints.

Four of the six mice with the memory-only sorting profile (M20, M21, M22, and M24) have a comparable number of CNAs to the other 19 samples (M1–M19), but M23 and M25 had a very large

Table 1. List of recurrent CNAs observed among tested diabetic mice excluding overlaps between M23 and M25 (due to their large number) unless shared by at least one other mouse

ID	Genomic coordinates	Locus	# Probes	LRR	<i>P</i> -value	Recurrent genes	DLRS	Figure
M4	Chr 2: 74,475,401–74,598,078	qC3	15	-0.34	8.55 × 10 ⁻¹⁰	<i>Hoxd</i> gene cluster	0.21	Supplemental Fig. S3A
M12	Chr 2: 74,501,193–74,598,078	qC3	13	-0.42	2.33 × 10 ⁻¹²		0.19	
M23	Chr 3: 94,759,757–94,857,185	qF2.1	12	1.47	3.10 × 10 ⁻²⁶	<i>Pi4kb</i> , <i>Zfp687</i>	0.28	Supplemental Fig. S3B
M25	Chr 3: 94,759,757–94,844,573	qF2.1	11	0.73	1.87 × 10 ⁻¹¹		0.23	
M5	Chr 3: 94,767,165–94,821,010	qF2.1	7	0.49	6.82 × 10 ⁻¹⁰		0.19	
M4	Chr 6: 52,104,495–52,281,548	qB3	18	-0.35	7.06 × 10 ⁻¹²	<i>Hoxa</i> gene cluster	0.21	Supplemental Fig. S3C
M12	Chr 6: 52,116,749–52,207,847	qB3	9	-0.48	1.65 × 10 ⁻¹¹		0.19	
M10	Chr 8: 72,706,464–72,777,429	qB3.3	8	0.51	3.83 × 10 ⁻¹⁰	<i>Slc25a42</i> , <i>Armc6</i> , <i>Sugp2</i>	0.18	Supplemental Fig. S3D
M3	Chr 8: 72,718,314–72,795,346	qB3.3	9	0.47	1.18 × 10 ⁻¹⁰		0.20	
M25	Chr 11: 69,752,914–68,919,294	qB3	17	0.71	7.10 × 10 ⁻³⁷	<i>Tmem107</i> , <i>Vamp2</i>	0.23	Fig. 4A
M23	Chr 11: 68,766,792–68,909,840	qB3	15	1.65	5.46 × 10 ⁻¹¹⁸		0.28	
M22	Chr 11: 68,880,665–68,909,840	qB3	4	0.79	2.01 × 10 ⁻¹⁴		0.19	
M22	Chr 13: 21,798,640–21,921,804	qA3.1	13	0.40	7.92 × 10 ⁻¹²	Histones family	0.19	Fig. 4C
M23	Chr 13: 21,813,250–21,921,804	qA3.1	11	2.05	4.45 × 10 ⁻¹²⁷		0.28	
M25	Chr 13: 21,813,250–21,921,804	qA3.1	11	1.01	4.62 × 10 ⁻⁴⁷		0.23	
M22	Chr 13: 23,648,864–23,661,297	qA3.1	3	0.86	1.11 × 10 ⁻¹¹	Histones family	0.19	Fig. 4C
M23	Chr 13: 23,648,864–23,725,304	qA3.1	10	2.12	7.57 × 10 ⁻¹¹		0.28	
M25	Chr 13: 23,847,542–23,847,542	qA3.1	22	0.88	6.56 × 10 ⁻⁶⁹		0.23	
M5	Chr 14: 52,061,318–52,152,579	qC1–qC2	8	-0.47	1.47 × 10 ⁻¹⁰	<i>Gm7247</i> , <i>Vmn2r89</i>	0.19	Fig. 4B
M22	Chr 14: 52,061,318–52,145,237	qC1–qC2	7	-0.57	6.92 × 10 ⁻¹³		0.19	

A complete list of CNAs can be found in Supplemental Data S2 and S3.

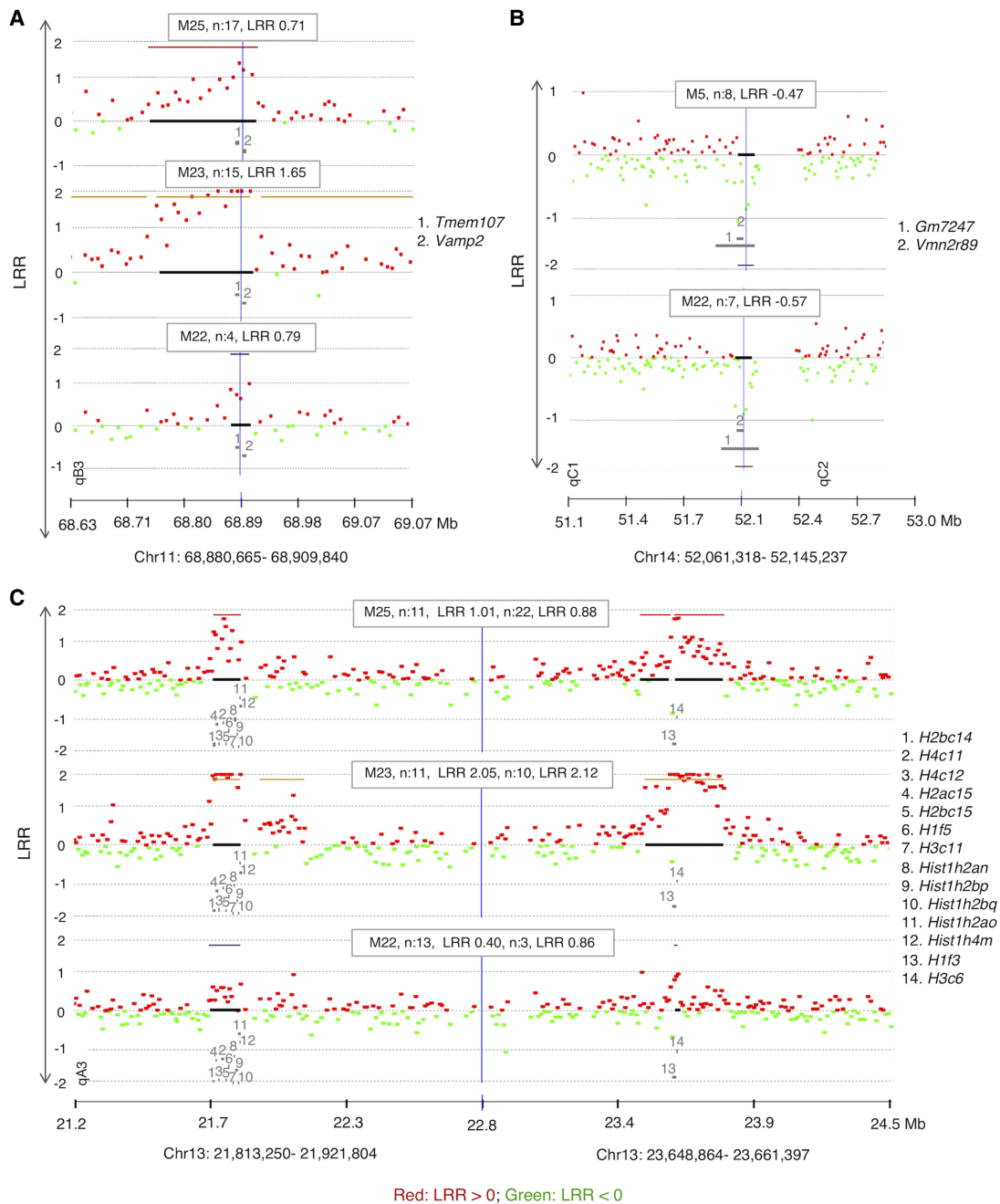


Figure 4. CGH array results showing recurrent CNAs in memory cells from PLNs of diabetic mice (n : number of affected probes, location of genes is indicated by numbers). (A) A recurrent copy gain in cells of three mice spanning several genes including *Tmem107*. (B) Recurrent copy loss spanning two undocumented genes (*Gm7247* and *Vmn2r89*). (C) An independent recurrence of copy gain in three different mice that are spanning two Histone families' loci. The formula to calculate the percentage of mosaicism of deletions is $100 \times (1 - 2^{\text{LRR}}) \times 2$ (example: the percentage of mosaicism for -0.25 is 31.8% mosaicism and for -0.5 is 58.6%). DLRSpread values of samples are in Table 1 and Supplemental Fig. S2.

number of CNAs, most of which were recurrent loci between the two mice. Chr X and Y probes clearly showed a male and female genome, ruling out possible sample duplication through a labeling error (Supplemental Table S2). The high numbers of CNAs in these two mice is reminiscent of CNAs observed in a malignant genome of a leukemic cell and involve many genes, some of them of known immune function. Three examples of these CNAs include a recurrent CNA in M23 and M25 that is a copy gain at Chr 9: 20,856,727–

21,236,076 spanning several genes, including *Keap1*, *S1pr5*, and *Ilf3* with hematopoietic cells expression and an immune-related function (Table 2; Supplemental Fig. S5A; Wu et al. 2016). Another large CNA is at Chr 10: 79,084,598–81,041,680. This recurrent CNA spans a large number of genes of which several also have a high or exclusive expression in hematopoietic cells, including *Madcam1*, *Gzmm*, *Rps15*, *Mknk2*, *Dapk3*, *Matk*, *S1pr4*, and *Abca7* (Table 2; Supplemental Fig. S5B; Wu et al. 2016). Another

Table 2. List of genes of interest within the three CNAs confirmed by MLPA in the original unamplified DNA as an independent method

	Genes	Lymphocyte expression	Immune/regulatory related function
Chr 9: 20,856,727–21,236,076	<i>Keap1</i>	Yes	Inflammatory regulatory function (Awuh et al. 2015). Suppresses the nuclear factor NFE2L2 causing its degradation (Rockwell et al. 2012). <i>Nfe2l2</i> activation has an anti-inflammatory effect, while <i>Nfe2l2</i> deletion has pro-inflammatory effects (Innamorato et al. 2008; Rockwell et al. 2012).
	<i>S1pr5</i>	High and exclusive in NK cells	SIPD1 controls cell proliferation and apoptosis and regulates T and B cells migration (Cyster 2005; Rosen and Goetzl 2005; Rosen et al. 2007). S1PR5 is necessary for the circulation of NK cells to inflamed tissue (Walzer et al. 2007).
	<i>Ilf3</i>	Yes	Encodes a double-stranded RNA-binding protein that is required for T cell interleukin 2 expression (Jarbouli et al. 2011). Highly expressed in human lymphocytes (Wu et al. 2016). Associated with psoriasis (Tsoi et al. 2012) and rheumatoid arthritis (Okada et al. 2014).
Chr 10: 79,084,598–81,041,680	<i>Madcam1</i>	High in spleen and lymph nodes	Encodes an endothelial cells adhesion molecule, a member of the immunoglobulins. Ligand to lymphocytes homing receptor (Grayson et al. 2003). Strongly up-regulated in active celiac disease (Di Sabatino et al. 2009).
	<i>Gzmm</i>	High in lymphocytes	Encoding granzyme M proteins. Group of enzymes is expressed by NK cells and activated lymphocytes and stored in cytoplasmic granules (Lu et al. 2006).
	<i>Ptbp1</i>	High in lymphocytes	Encodes a nuclear ribonucleoprotein. Increases CD5 protein levels upon T cell activation (Domingues et al. 2016).
	<i>Rps15</i>	Yes	Encodes the RS15P protein belonging to a family of ribosomal proteins. Recurrently mutated in aggressive chronic lymphocytic leukemia (CLL) (Ljungström et al. 2016). Highly expressed in human lymphocytes (Wu et al. 2016).
	<i>Mknk2</i>	Yes	Encodes a member of the protein kinases superfamily. Crucial role of this protein in T cells function was suggested (Zhong et al. 2008; Gorenli et al. 2013). Highly expressed in human lymphocytes (Wu et al. 2016).
	<i>Dapk3</i>	Yes	Encodes death-associated protein kinase 3. Inducing of apoptosis and regulation of cell cycle progression and proliferation. Tumor suppressor function was suggested (Brognard et al. 2011).
	<i>Matk</i>	Yes	Crucial in the signal transduction of hematopoietic cells. Inhibitory function in the regulation of T cell proliferation (Tan et al. 2011). Exclusive expressed in lymphocytes (Wu et al. 2016).
	<i>S1pr4</i>	High in lymphocytes	Encodes sphingosine-1-phosphate receptor 4. Cell signaling of several cell types (Wang et al. 2014).
	<i>Abca7</i>	High in lymphocytes	Encodes a transporter protein detected principally in myelo-lymphatic tissues. Expression signature implies a function in lipid homeostasis in immune cells (Tomioka et al. 2017).
Chr 10: 127,684,108–128,334,229	<i>Il23a</i>	Yes	Encodes a subunit of IL-23 that stimulates the production of interferon-gamma and acts on memory CD4 ⁺ T cells (Revu et al. 2018). Higher levels of IL-23 have been associated with autoimmune diseases such as multiple sclerosis and inflammatory bowel disease (Duerr et al. 2006; Abraham and Cho 2009).
	<i>Pa2g4</i>	Yes	High in hematopoietic cells, leukemic and lymphoma cells (Wu et al. 2016). Encodes a protein that functions in cell proliferation, transducing growth regulatory signals and inducing differentiation of human cancer cells (Nguyen et al. 2015, 2016). Highly expressed in human lymphocytes (Wu et al. 2016).
	<i>Dgka</i>	High in lymphocytes	Involved in cell cycle progression and the function of immune cells (Prinz et al. 2012; Poli et al. 2017). Associated with T1D (Evangelou et al. 2014).
	<i>Coq10a</i>	High in CD4 ⁺ and CD8 ⁺	Unstudied gene

is a recurrent copy gain CNA in these two mice also at Chr 10: 127,684,108–128,334,229, also spanning several genes with high or exclusive hematopoietic cell expression and/or an immune function, including *Il23a*, *Pa2g4*, *Dgka*, and *Coq10a* that are expressed in lymphocytes (*Dgka* is exclusively expressed in CD4⁺ and CD8⁺ cells) (Table 2; Supplemental Fig. S5C).

Since the above recurrent CNAs span several genes with a high lymphocyte expression and have an immune and proliferation function in immune cells (Table 2), we aimed to confirm these CNAs in the original unamplified DNA by multiplex ligation-dependent probe amplification (MLPA), the gold standard method for CNV diagnostics (Feuk et al. 2006) as independent validation. We selected three genes within the above three recurrent CNAs, that are expressed in lymphocytes and have an immune- and/or proliferation-related function and in which we have enough of the original unamplified DNA to be examined by MLPA: *Ilf3*, *Abca7*, and *Dgka*. *Ilf3* on Chr 9 is highly expressed in

lymphocytes and encodes a double-stranded RNA-binding protein that is required for T cell interleukin 2 expression (Jarbouli et al. 2011). Moreover, it is associated with two immune-related diseases: psoriasis (Tsoi et al. 2012) and rheumatoid arthritis (Okada et al. 2014). *Abca7* on Chr 10 encodes a transporter protein detected principally in myelo-lymphatic tissues with an expression signature that implies a function in lipid homeostasis in immune cells (Tomioka et al. 2017). *Dgka* is exclusively expressed in CD4⁺ and CD8⁺ cells, and its encoded protein is involved in cell cycle progression and the function of immune cells (Prinz et al. 2012; Poli et al. 2017) and has been reported to be associated with T1D (Evangelou et al. 2014). We successfully confirmed by MLPA the above three CNAs (Fig. 5A,B). Further detailed annotation of the CNAs seen exclusively in M23 and M25 was not attempted because of their large number and the likelihood that they are a consequence rather than a cause of the rapid proliferation.

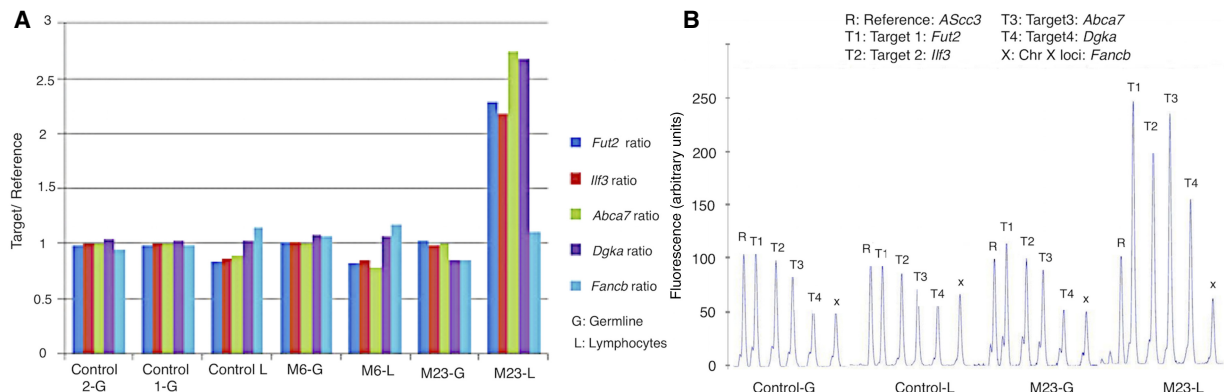


Figure 5. Independent confirmation of CNAs in original unamplified DNA by MLPA. Four recurrent loci with copy gain in M23 and M25 spanning genes with immune function (Supplemental Fig. S5) were tested, along with DNA from three mice that did not have the CNA (M6-L and -G, control L and 1-G, and control 2-G). (A) The height of the MLPA peak at each locus for each mouse is shown in proportion to the reference probe. Only the four loci in the T cell (but not the tail representing the germline) DNA from mouse M23 show an increase over 1. (B) Representative peaks shown as a percentage of the reference probe. Only mouse M23 shows doubling of the peak in T cells.

T cells involved in host defense have smaller copy number changes

As a normal host-defense control, we tested DNA from memory CD4⁺ T cells from the popliteal lymph nodes of 10 NOD mice (eight females and two males; 9–14 wk) infected with *Leishmania major* (*L. major*). Out of 10, two had CNAs (one had one and the other one had two CNAs) (Supplemental Table S3; Supplemental Data S4). The same experiment was done on seven non-autoimmunity susceptible BALB/c mice (19 wk) to determine if the occurrence of CNAs in lymphocytes is promoted by an autoimmune genetic background. Out of the seven mice, three had CNAs (1–2 CNAs) (Supplemental Table S3; Supplemental Data S4). The mean size of CNAs in autoreactive cells of diabetic mice (26.02 ± 1.69 , mean \pm SEM, $n = 403$) was significantly twofold larger than that in samples from *L. major*-infected mice (7.71 ± 2.03 , $n = 7$; Fisher-Pitman permutation test, $P = 0.0019$).

The CNAs occur in prethymic precursors

Sequencing the TCR of the CD4⁺ memory cells from PLN of 10 diabetic mice showed low clonality in both the α -chain and β -chain (Supplemental Data S5), too low to explain the level of mosaicism found in the CNAs, suggesting prethymic occurrence in a marrow progenitor. We found that the most common TCR is (CDR3-V-J: CVVGDGRGALGRLHF-TRAV11*01-TRAJ18*01) sequence for the α -chain and (CDR3-V-J-D: CAASSNTDKVVF-TRBV2*01-TRBJ1-1*01-TRBD1*01) sequence for the β -chain are present within all samples with relatively high frequency.

Discussion

This study demonstrates a novel discovery of somatic copy-number mutations in immune-related genes in memory cells of diabetic NOD mice, an excellent model for human T1D. Some of these genes are highly expressed in T cells and may well represent an important causal contribution in the autoimmunity of the NOD model and, quite plausibly, of human T1D. Their causal role is supported by the nonrandom recurrence of eight of these PZMs affecting the same genes in different mice. In the paradigm of cancer, independent recurrence of somatic mutations, either point or copy-number, in different patients is considered strong evidence that the mutation is causal. In fact, this is the criterion by which

most tumor-suppressor genes have been identified. Somatic mutations are a common occurrence, but, being confined to one single cell and its progeny, they are unlikely to cause disease, unless the cell progeny expands exponentially, as is the case in cancer (Jacobs et al. 2012; Li et al. 2014) and, to a lesser extent, in autoreactive cells.

We considered alternative explanations for our observations and found them unlikely. Hyperglycemia could accelerate the normal accumulation of somatic mutations in blood cells, as described for type 2 diabetes (Bonfond et al. 2013; Loftfield et al. 2018), but the alterations we observed were much more drastic, over a much shorter timeframe (euthanasia within days of hyperglycemia onset). It is possible that some of our findings are DNA amplification artifacts. Against it is the control experiments described under Methods and the MLPA confirmation using unamplified DNA, especially in mice M23 and M25, who had the highest number of CNAs, i.e., the most likely to be due to technical artifacts. Also, in our mixture experiments, we found the same loci in all three mixes, or in the two highest ones, always at the same locus. This is very unlikely to happen at random, if these were amplification artifacts. Overlaps also are against artifacts, which should have happened at random.

The contribution of somatic mutations in the pathogenesis of cancer has been well established. A PZM in a tumor suppressor gene in a single cell causes a monoclonal expansion during which additional PZMs occur, causing more pathological phenotypes and further release from proliferation controls (Watson et al. 2013). The identification of such driver mutations in cancer has enabled therapeutic advances by identifying drug targets. One example is Gleevec, a selective inhibitor of the BCR-ABL tyrosine kinase, the pathogenic fusion transcript in chronic myeloid leukemia (CML) (Hernández-Boluda and Cervantes 2002).

In the case of autoimmune diseases, similar to cancer, disease is caused by a proliferation of clonal lymphocytes from a small number of autoreactive cells that have escaped natural tolerance to self-antigens. Although autoimmunity involves more functional steps than proliferation and resistance to apoptosis, it does share these two features with cancer. Natural tolerance involves many checkpoints. Its failure is likely a multistep process, causing autoimmune diseases to develop over many years. It has been estimated that as many as 50% of all T cell receptor V(D)J rearrangements

will recognize some autoantigen (Ignatowicz et al. 1996). Most of these will be deleted by negative selection in the thymus, where tissue-specific antigens like insulin are expressed (Vafiadis et al. 1997; Liston et al. 2004). However, some autoreactive T cells escape to the periphery and must be dealt with by peripheral tolerance. A large number of molecular players are involved in each of these checkpoints (for review, see Goodnow et al. 2005), and many others may be important but still unstudied.

Epidemiological evidence has connected autoimmunity to hematologic malignancy in several scenarios. The risk of lymphoid malignancy is higher in individuals affected by many autoimmune conditions, such as celiac disease, lupus, rheumatoid arthritis, and Sjorgen's syndrome (Goodnow 2007). Somatic *FAS* mutations in non-Hodgkin's lymphoma are often seen in patients with a prior history of autoimmunity (Grønbaek et al. 1998). This suggests overlaps between immune tolerance checkpoints, important in autoimmunity, and proliferation checkpoints, important in lymphoid malignancy.

The plausibility of this model depends on how often such PZMs happen in autoimmunity-relevant hematopoietic lineages. PZMs originated in hematopoietic stem cells and progenitor cells have been identified in peripheral whole blood (Busque et al. 2012; Genovese et al. 2014; Xie et al. 2014; Young et al. 2016). SNP-array data show that whole-blood DNA has mosaic, Mb-sized CNAs in about 5% of healthy individuals (Forsberg et al. 2012; Jacobs et al. 2012; Laurie et al. 2012). This detectable mosaicism is due to a proliferative or survival advantage of the copy-number mutation (versus early occurrence in embryogenesis), as it is age-dependent and variable over time (Forsberg et al. 2012). Only a very small fraction of all people carrying them will develop leukemia, leaving a broad gray zone for a role in nonmalignant disease. In a related study of SNP-array discrepancies between monozygotic twins, we calculated an average, ~130 point-PZMs per individual, at high enough mosaicism, in whole blood, to flip a SNP genotype call (Li et al. 2014). Single-cell sequencing has more recently revealed that somatic mutations are common and accumulate in normal cells (Abyzov et al. 2012; McConnell et al. 2013; Lodato et al. 2015, 2018; Blokzijl et al. 2016; Bae et al. 2018).

In logarithmically expanded autoantigen-specific lineages, mosaicism might be much higher and still not detectable without isolating these lineages from the rest of the blood cells. A recent publication identified an exclusive somatic mutation in the expanded CD8⁺ memory subset in newly diagnosed rheumatoid arthritis (Savola et al. 2017).

In six of the 25 mice, all cells in PLNs were memory, and two of these mice showed a very unstable genome with many recurrent CNAs. This instability might be explained by the copy gain in two histone loci (Fig. 4C). Histone overexpression due to copy gain is very likely to have increased the normally tightly regulated expression of these nucleosomal proteins, whose correct expression upon DNA replication is crucial to genomic integrity (Mejlvang et al. 2014). The multiple nearly identical genes encoding these two histones also make the locus susceptible to rearrangements. The remaining CNAs cover a large number of RefSeq genes, most of which are not functionally annotated and have no published literature.

Activated T cells in the context of host defense with *L. major* also have CNAs, but they are fewer (seven in 17 mice versus 403 in 25 mice) and significantly smaller than those observed in T cells involved in NOD autoimmunity. Moreover, their occurrence in BALB/c indicates that they are not due to the germline determinants of autoimmunity.

Identification of the genes affected by the CNAs in the autoreactive lymphocytes could serve as a potential therapeutic target as their CNAs exist only in disease-causing cells. One example of these is recurrent CNAs that span *Igflr1* (Supplemental Data S3) and *Tmem107* (Fig. 4A) encoding transmembrane proteins highly expressed in T cells. Their amplification or deletion should be detectable by flow cytometry in blood cells, providing a potential biomarker and therapeutic target.

Sequencing the TCR of 10 diabetic mice revealed a percentage of clonotypes too low to explain the observed levels of mosaicism, suggesting that the CNAs are occurring prior to the thymic selection in a bone marrow progenitor and not in the antigen-specific lineage. Additional PZMs, occurring during the autoantigen-activated logarithmic expansion of T cells likely to happen but their level of mosaicism would have been too low to detect by our approach, as we did not isolate cells by antigen specificity.

It is possible that the overlap could be due to a mechanistic origin due to the fact that some parts of the genome are more fragile than others. However, there is a very large number of such loci, making their recurrence unlikely. Moreover, none of our findings correspond to known CNVs between mouse strains.

Since these postzygotic mutations do not have to pass selection for fitness of the whole organism, they are likely to have more drastic functional effects than germline variants. If so, they could be ideal therapeutic targets, as they exist only in disease-causing cells. These data must be confirmed and expanded using CNA detection in single cells by sequencing of multiple cell clones that will enable a more accurate detection and estimate of the level of genetic mosaicism (Abyzov et al. 2012; McConnell et al. 2013; Lodato et al. 2015; Blokzijl et al. 2016; Bae et al. 2018; Lodato et al. 2018). Future studies should address the causal role by gene editing and extend the studies to human T1D autoimmunity. Other autoimmune diseases should be likewise examined and detection of point mutations by exome sequencing is planned, to explore the full PZM spectrum.

Methods

Study approval

In this study, all experiments on animals were approved by the Facility Animal Care Committee (FACC) according to the guidelines established by the Canadian Council on Animal Care and the McGill University Animal Care Committee (AUP#7517).

Experimental model

Animals were obtained from standard sources and the diagnosis of diabetes, euthanasia, and dissection were carried out by standard procedures (Supplemental Methods).

Isolation of memory T lymphocytes

Nodes were placed in 1 mL cold Dulbecco's saline DPBS (1×) (cat#: 14190-144, Life Technologies) and lymphocytes isolated by sheer-force slide dissociation. Next, fluorescence activated cell sorting (FACS) was used to isolate memory T lymphocytes (CD4⁺/SELL⁻/CD44⁺ by a FACSria cell sorter (Becton-Dickinson). In these experiments, cells were stained using: anti-CD4⁺-FITC (cat#: 11-0042-81, eBioscience), anti-CD8α⁺-PE (cat#: 12-0081-82, eBioscience), anti-CD62L⁺-APC (naive-cell marker; cat#: 130-091-805, Miltenyi Biotec), and anti-Human/Mouse CD44⁺-PreCP-cy5.5 (memory-cell marker; cat#: 45-0441-82,

eBioscience). All samples were kept on ice throughout isolation and staining.

Whole-genome amplification

DNA was amplified using a GenomePlex complete whole-genome amplification (WGA-2) (cat#: WGA2-50RXN) kit, specifically approved by Agilent for aCGH to detect copy-number differences. DNA was next purified using a PCR Cleanup kit from Sigma-Aldrich (cat#: NA1020).

Detection of PZMs

We sent 1.5 µg of amplified DNA to Oxford Gene Technologies (OGT) for comparative genomic hybridization analysis on the mouse 4 × 180 K Agilent array. DNA from memory cells was the test sample, with germline DNA from the tail as reference. To model the sensitivity to detect levels of mosaicism, germline NOD DNA was mixed with C57BL6/J DNA prior to amplification. The NOD proportions were 100%, 50%, and 25%, to imitate levels of mosaicism at inter-strain copy-number variants (CNV). Reference was 100% C57BL6/J.

Copy-number alterations (CNAs) were detected by the Agilent dual color, 4 × 180 k mouse CGH array (Supplemental Methods). CGH quality control (QC)- derivative log ratio spread (DLRS) values were within Agilent Technologies specifications (Supplemental Fig. S2).

To evaluate the specificity of the method, two technical controls were tested in which DNA from lymphocytes from C57BL6/J mouse was amplified twice by the same kit (WGA-2), and then one of the two amplified samples was used as test and the other as a reference in CGH (4 × 180 K) analysis. Moreover, DNA was amplified twice from memory CD4⁺ T cells obtained from PLN of a diabetic mouse (M9) and twice from germline DNA from a tail-clip sample of the same mouse. These duplicate samples were analyzed separately (M9 and M9B) by CGH using in both samples the amplified memory CD4⁺ DNA as test and the amplified germline DNA as reference.

Bioinformatics analysis

Test and reference signals were calculated for each probe and internal normalization equalized the systematic difference between the two fluorochromes. The base-2 log of the Cy3/Cy5 ratio (LRR) was plotted against chromosomal coordinates. Probes covering NOD or BALB/c germline deletions were eliminated. Regions bearing copy-number aberrations were identified using two independent algorithms, Agilent Genomic Workbench 7.0 (Agilent Technologies) (Roy and Motsinger Reif 2013) implementing ADM-2 algorithm and DNACopy software implementing the circular binary segmentation (CBS) algorithm (Venkatraman and Olshen 2007). In ADM-2 (threshold 6.0), we applied Agilent Technologies recommended filtering standards of a minimum of three aberrant consecutive probes. The mouse genome UCSC mm9 NCBI build 37/July 2007 was used to match the indexing of the Agilent probes (converting coordinates did not affect the findings).

CNAs were kept as true if called by both algorithms ADM-2 ($LRR \geq |0.25|$, $P\text{-value} \leq 2.90 \times 10^{-07}$, Bonferroni threshold) and CBS ($LRR|0.25|$, $P\text{-value} \leq 1 \times 10^{-04}$). The lower statistical significance threshold for CBS is justified by the vastly smaller number of hypotheses tested (number of ADM-2-called CNAs).

Tissue-specificity of the expression levels of the genes involved was obtained from mouse and human BioGPS data, 2018 (Wu et al. 2016).

Statistics

The statistical significance of the recurrence of overlapping CNAs in different mice was estimated by permutation testing (Supplemental Methods). Either the actual length in kilobases of each CNA or the number of probes covering it were used in this simulation. Both gave similar results, but we retained the results of the probe permutation, a more conservative approach that avoids inflating the statistical significance because of nonhomogenous distribution of the probes (favoring coding genes).

The difference in the number of probes covering CNAs in T cells from diabetic pancreatic nodes versus *L. major*-infected popliteal nodes was assessed by the Fisher-Pitman permutation test, which does not require an assumption about normality of distribution. We compared the number of probes (number of consecutive probes covered); because of the gene-centered positioning of the probes, they better reflect the genes covered by them than the actual CNA length.

Multiplex ligation-dependent probe amplification

Some CNAs were selected for confirmation by MLPA on unamplified DNA by a standard protocol (Supplemental Methods). Primer and probes used for the MLPA experiment can be found in Supplemental Table S4.

TCR sequencing

RNA samples were extracted from the CD4⁺ memory cells isolated from the PLN of diabetic mice and amplified by a 5' RACE approach using the SMARTer RACE cDNA amplification kit by Clontech and sequenced on an Illumina MiSeq apparatus (Supplemental Methods).

Data access

All raw and processed CGH data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE114660. All raw and processed sequencing data generated in this study have been submitted to the NCBI Sequence Read Archive (SRA; <https://www.ncbi.nlm.nih.gov/sra/>) under accession number SRP148447.

Acknowledgments

We thank Michael Saraga for his help in analyzing four of the mice CGH results. This work was supported by grants from the Canadian Institutes of Health Research (MOP-137050) and from the Canadian Diabetes Association (OG-3-11-3429CP). M.A. holds a doctoral-studies scholarship from Sultan Qaboos University, Oman.

Author contributions: C.P. designed and supervised the study. M.A., L.M., and X.D. performed the experiments. M.O. injected mice with *L. major*. M.A., L.M., and Q.L. analyzed data. All other authors provided samples or analysis tools. M.A. and C.P. wrote the manuscript. All authors read and approved the final manuscript.

References

- Abraham C, Cho JH. 2009. IL-23 and autoimmunity: new insights into the pathogenesis of inflammatory bowel disease. *Annu Rev Med* **60**: 97–110. doi:10.1146/annurev.med.60.051407.123757
- Abyzov A, Mariani J, Palejev D, Zhang Y, Haney MS, Tomasini L, Ferrandino AF, Rosenberg Belmaker LA, Szekely A, et al. 2012. Somatic copy number

- mosaicism in human skin revealed by induced pluripotent stem cells. *Nature* **492**: 438–442. doi:10.1038/nature11629
- Awuh JA, Haug M, Mildenerberger J, Marstad A, Do CP, Louet C, Stenvik J, Steigedal M, Damás JK, Halaas O, et al. 2015. Keap1 regulates inflammatory signaling in *Mycobacterium avium*-infected human macrophages. *Proc Natl Acad Sci* **112**: E4272–E4280. doi:10.1073/pnas.1423449112
- Bae T, Tomasini L, Mariani J, Zhou B, Roychowdhury T, Franjic D, Pletikos M, Pattini R, Chen BJ, Venturini E, et al. 2018. Different mutational rates and mechanisms in human cells at pregastrulation and neurogenesis. *Science* **359**: 550–555. doi:10.1126/science.aan8690
- Blokzijl F, de Ligt J, Jager M, Sasselli V, Roerink S, Sasaki N, Huch M, Boymans S, Kuijk E, Prins P, et al. 2016. Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* **538**: 260–264. doi:10.1038/nature19768
- Bonnefond A, Skrobek B, Lobbens S, Eury E, Thuillier D, Cauchi S, Lantieri O, Balkau B, Riboli E, Marre M, et al. 2013. Association between large detectable clonal mosaicism and type 2 diabetes with vascular complications. *Nat Genet* **45**: 1040–1043. doi:10.1038/ng.2700
- Brognaard J, Zhang YW, Puto LA, Hunter T. 2011. Cancer-associated loss-of-function mutations implicate DAPK3 as a tumor-suppressing kinase. *Cancer Res* **71**: 3152–3161. doi:10.1158/0008-5472.CAN-10-3543
- Burnet FM. 1972. *Autoimmunity and autoimmune disease*. Medical and Technical Publishing Co Ltd, Lancaster, UK.
- Busque L, Patel JP, Figueroa ME, Vasanthakumar A, Provost S, Hamilou Z, Mollica L, Li J, Viale A, Heguy A, et al. 2012. Recurrent somatic *TET2* mutations in normal elderly individuals with clonal hematopoiesis. *Nat Genet* **44**: 1179–1181. doi:10.1038/ng.2413
- Cyster JG. 2005. Chemokines, sphingosine-1-phosphate, and cell migration in secondary lymphoid organs. *Annu Rev Immunol* **23**: 127–159. doi:10.1146/annurev.immunol.23.021704.115628
- Di Sabatino A, Rovedatti L, Rosado MM, Carsetti R, Corazza GR, MacDonald TT. 2009. Increased expression of mucosal addressin cell adhesion molecule 1 in the duodenum of patients with active celiac disease is associated with depletion of integrin $\alpha_4\beta_7$ -positive T cells in blood. *Hum Pathol* **40**: 699–704. doi:10.1016/j.humpath.2008.10.014
- Domingues RG, Lago-Baldaia I, Pereira-Castro I, Fachini JM, Oliveira L, Drpic D, Lopes N, Henriques T, Neilson JR, Carmo AM, et al. 2016. CD5 expression is regulated during human T-cell activation by alternative polyadenylation, PTBP1, and miR-204. *Eur J Immunol* **46**: 1490–1503. doi:10.1002/eji.201545663
- Dowdell KC, Niemela JE, Price S, Davis J, Hornung RL, Oliveira JB, Puck JM, Jaffe ES, Pittaluga S, Cohen JI, et al. 2010. Somatic *FAS* mutations are common in patients with genetically undefined autoimmune lymphoproliferative syndrome. *Blood* **115**: 5164–5169. doi:10.1182/blood-2010-01-263145
- Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, Steinhardt AH, Abraham C, Regueiro M, Griffiths A, et al. 2006. A genome-wide association study identifies *IL23R* as an inflammatory bowel disease gene. *Science* **314**: 1461–1463. doi:10.1126/science.1135245
- Evangelou M, Smyth DJ, Fortune MD, Burren OS, Walker NM, Guo H, Onengut-Gumuscu S, Chen WM, Concannon P, Rich SS, et al. 2014. A method for gene-based pathway analysis using genomewide association study summary statistics reveals nine new type 1 diabetes associations. *Genet Epidemiol* **38**: 661–670. doi:10.1002/gepi.21853
- Feuk L, Carson AR, Scherer SW. 2006. Structural variation in the human genome. *Nat Rev Genet* **7**: 85–97. doi:10.1038/nrg1767
- Forsberg LA, Rasi C, Razzaghi HR, Pakalapati G, Waite L, Thilbeault KS, Ronowicz A, Wineinger NE, Tiwari HK, Boomsma D, et al. 2012. Age-related somatic structural changes in the nuclear genome of human blood cells. *Am J Hum Genet* **90**: 217–228. doi:10.1016/j.ajhg.2011.12.009
- Genovese G, Kähler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, Chambert K, Mick E, Neale BM, Fromer M, et al. 2014. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med* **371**: 2477–2487. doi:10.1056/NEJMoa1409405
- Ghare SS, Joshi-Barve S, Moghe A, Patil M, Barker DF, Gobejishvili L, Brock GN, Cave M, McClain CJ, Barve SS. 2014. Coordinated histone H3 methylation and acetylation regulate physiologic and pathologic Fas ligand gene expression in human CD4⁺ T cells. *J Immunol* **193**: 412–421. doi:10.4049/jimmunol.1400055
- Goodnow CC. 2007. Multistep pathogenesis of autoimmune disease. *Cell* **130**: 25–35. doi:10.1016/j.cell.2007.06.033
- Goodnow CC, Sprent J, Fazekas de St Groth B, Vinuesa CG. 2005. Cellular and genetic mechanisms of self tolerance and autoimmunity. *Nature* **435**: 590–597. doi:10.1038/nature03724
- Gorentla BK, Krishna S, Shin J, Inoue M, Shinohara ML, Grayson JM, Fukunaga R, Zhong XP. 2013. Mnk1 and 2 are dispensable for T cell development and activation but important for the pathogenesis of experimental autoimmune encephalomyelitis. *J Immunol* **190**: 1026–1037. doi:10.4049/jimmunol.1200026
- Grayson MH, Hotchkiss RS, Karl IE, Holtzman MJ, Chaplin DD. 2003. Intravital microscopy comparing T lymphocyte trafficking to the spleen and the mesenteric lymph node. *Am J Physiol Heart Circ Physiol* **284**: H2213–H2226. doi:10.1152/ajpheart.00999.2002
- Gress TM, Muller-Pillasch F, Geng M, Zimmerhackl F, Zehetner G, Friess H, Buchler M, Adler G, Lehrach H. 1996. A pancreatic cancer-specific expression profile. *Oncogene* **13**: 1819–1830.
- Grønbaek K, Straten PT, Ralfkiaer E, Ahrenkiel V, Andersen MK, Hansen NE, Zeuthen J, Hou-Jensen K, Guldberg P. 1998. Somatic Fas mutations in non-Hodgkin's lymphoma: association with extranodal disease and autoimmunity. *Blood* **92**: 3018–3024. doi:10.1182/blood.V92.9.3018
- Heilmeyer LM Jr., Vereb G Jr., Vereb G, Kakuk A, Szivák I. 2003. Mammalian phosphatidylinositol 4-kinases. *IUBMB Life* **55**: 59–65. doi:10.1080/1521654031000090896
- Hernández-Boluda JC, Cervantes F. 2002. Imatinib mesylate (Gleevec, Glivec): a new therapy for chronic myeloid leukemia and other malignancies. *Drugs Today (Barc)* **38**: 601–613. doi:10.1358/dot.2002.38.9.696536
- Holzelova E, Vonarbourg C, Stolzenberg MC, Arkwright PD, Selz F, Prieur AM, Blanche S, Bartunkova J, Vilmer E, Fischer A, et al. 2004. Autoimmune lymphoproliferative syndrome with somatic *Fas* mutations. *N Engl J Med* **351**: 1409–1418. doi:10.1056/NEJMoa040036
- Ignatowicz L, Kappler J, Marrack P. 1996. The repertoire of T cells shaped by a single MHC/peptide ligand. *Cell* **84**: 521–529. doi:10.1016/S0092-8674(00)81028-4
- Innamorato NG, Rojo AI, García-Yagüe AJ, Yamamoto M, de Ceballos ML, Cuadrado A. 2008. The transcription factor Nr2f2 is a therapeutic target against brain inflammation. *J Immunol* **181**: 680–689. doi:10.4049/jimmunol.181.1.680
- Izquierdo JM, Valcárcel J. 2007. Fas-activated serine/threonine kinase (FAST K) synergizes with TIA-1/TIAR proteins to regulate Fas alternative splicing. *J Biol Chem* **282**: 1539–1543. doi:10.1074/jbc.C600198200
- Jacobs KB, Yeager M, Zhou W, Wacholder S, Wang Z, Rodriguez-Santiago B, Hutchinson A, Deng X, Liu C, Horner MJ, et al. 2012. Detectable clonal mosaicism and its relationship to aging and cancer. *Nat Genet* **44**: 651–658. doi:10.1038/ng.2270
- Jarbouli MA, Wynne K, Elia G, Hall WW, Gautier VW. 2011. Proteomic profiling of the human T-cell nucleolus. *Mol Immunol* **49**: 441–452. doi:10.1016/j.molimm.2011.09.005
- Kawagoe H, Humphries RK, Blair A, Sutherland HJ, Hogge DE. 1999. Expression of HOX genes, HOX cofactors, and MLL in phenotypically and functionally defined subpopulations of leukemic and normal human hematopoietic cells. *Leukemia* **13**: 687–698. doi:10.1038/sj.leu.2401410
- Knip M, Veijola R, Virtanen SM, Hyoty H, Vaarala O, Akerblom HK. 2005. Environmental triggers and determinants of type 1 diabetes. *Diabetes* **54**: S125–S136. doi:10.2337/diabetes.54.suppl_2.S125
- Laurie CC, Laurie CA, Rice K, Doherty KF, Zelnick LR, McHugh CP, Ling H, Hetrick KN, Pugh EW, Amos C, et al. 2012. Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat Genet* **44**: 642–650. doi:10.1038/ng.2271
- Lawrence HJ, Sauvageau G, Humphries RK, Largman C. 1996. The role of *HOX* homeobox genes in normal and leukemic hematopoiesis. *Stem Cells* **14**: 281–291. doi:10.1002/stem.140281
- Li R, Montpetit A, Rousseau M, Wu SY, Greenwood CM, Spector TD, Pollak M, Polychronakos C, Richards JB. 2014. Somatic point mutations occurring early in development: a monozygotic twin study. *J Med Genet* **51**: 28–34. doi:10.1136/jmedgenet-2013-101712
- Liston A, Lesage S, Gray DH, O'Reilly LA, Strasser A, Fahrner AM, Boyd RL, Wilson J, Baxter AG, Gallo EM, et al. 2004. Generalized resistance to thymic deletion in the NOD mouse; a polygenic trait characterized by defective induction of *Bim*. *Immunity* **21**: 817–830. doi:10.1016/j.immuni.2004.10.014
- Ljungström V, Cortese D, Young E, Pandzic T, Mansouri L, Plevova K, Ntoufa S, Baliakas P, Clifford R, Sutton LA, et al. 2016. Whole-exome sequencing in relapsing chronic lymphocytic leukemia: clinical impact of recurrent *RPS15* mutations. *Blood* **127**: 1007–1016. doi:10.1182/blood-2015-10-674572
- Lobito AA, Ramani SR, Tom I, Bazan JF, Luis E, Fairbrother WJ, Ouyang W, Gonzalez LC. 2011. Murine insulin growth factor-like (IGFL) and human IGFL1 proteins are induced in inflammatory skin conditions and bind to a novel tumor necrosis factor receptor family member, IGFLR1. *J Biol Chem* **286**: 18969–18981. doi:10.1074/jbc.M111.224626
- Lodato MA, Woodworth MB, Lee S, Evrony GD, Mehta BK, Karger A, Lee S, Chittenden TW, D'Gama AM, Cai X, et al. 2015. Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**: 94–98. doi:10.1126/science.aab1785
- Lodato MA, Rodin RE, Bohrsen CL, Coulter ME, Barton AR, Kwon M, Sherman MA, Vitzthum CM, Luquette LJ, Yandava CN, et al. 2018. Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science* **359**: 555–559. doi:10.1126/science.aao4426

- Lofffield E, Zhou W, Graubard BI, Yeager M, Chanock SJ, Freedman ND, Machiela MJ. 2018. Predictors of mosaic chromosome Y loss and associations with mortality in the UK Biobank. *Sci Rep* **8**: 12316. doi:10.1038/s41598-018-30759-1
- Lu H, Hou Q, Zhao T, Zhang H, Zhang Q, Wu L, Fan Z. 2006. Granzyme M directly cleaves inhibitor of caspase-activated DNase (CAD) to unleash CAD leading to DNA fragmentation. *J Immunol* **177**: 1171–1178. doi:10.4049/jimmunol.177.2.1171
- Magerus-Chatinet A, Neven B, Stolzenberg MC, Daussy C, Arkwright PD, Lanzarotti N, Schaffner C, Cluet-Dennetiere S, Haerynck F, Michel G, et al. 2011. Onset of autoimmune lymphoproliferative syndrome (ALPS) in humans as a consequence of genetic defect accumulation. *J Clin Invest* **121**: 106–112. doi:10.1172/JCI43752
- Magli MC, Largman C, Lawrence HJ. 1997. Effects of HOX homeobox genes in blood cell differentiation. *J Cell Physiol* **173**: 168–177. doi:10.1002/(SICI)1097-4652(199711)173:2<168::AID-JCP16>3.0.CO;2-C
- Makino S, Kunimoto K, Muraoka Y, Mizushima Y, Katagiri K, Tachino Y. 1980. Breeding of a non-obese, diabetic strain of mice. *Jikken Dobutsu* **29**: 1–13. doi:10.1538/expanim1978.29.1_1
- McConnell MJ, Lindberg MR, Brennand KJ, Piper JC, Voet T, Cowing-Zitron C, Shumilina S, Lasken RS, Vermeesch JR, Hall IM, et al. 2013. Mosaic copy number variation in human neurons. *Science* **342**: 632–637. doi:10.1126/science.1243472
- Mejlvang J, Feng Y, Alabert C, Neelsen KJ, Jasencakova Z, Zhao X, Lees M, Sandelin A, Pasero P, Lopes M, et al. 2014. New histone supply regulates replication fork speed and PCNA unloading. *J Cell Biol* **204**: 29–43. doi:10.1083/jcb.201305017
- Nguyen LXT, Lee Y, Urbani L, Utz PJ, Hamburger AW, Sunwoo JB, Mitchell BS. 2015. Regulation of ribosomal RNA synthesis in T cells: requirement for GTP and Ebp1. *Blood* **125**: 2519–2529. doi:10.1182/blood-2014-12-616433
- Nguyen LXT, Zhu L, Lee Y, Ta L, Mitchell BS. 2016. Expression and role of the ErbB3-binding protein 1 in acute myelogenous leukemic cells. *Clin Cancer Res* **22**: 3320–3327. doi:10.1158/1078-0432.CCR-15-2282
- Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, Kochi Y, Ohmura K, Suzuki A, Yoshida S, et al. 2014. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**: 376–381. doi:10.1038/nature12873
- Pearson JA, Wong FS, Wen L. 2016. The importance of the Non Obese Diabetic (NOD) mouse model in autoimmune diabetes. *J Autoimmun* **66**: 76–88. doi:10.1016/j.jaut.2015.08.019
- Poli A, Fiume R, Baldanzi G, Capello D, Ratti S, Gesi M, Manzoli L, Graziani A, Suh PG, Cocco L, et al. 2017. Nuclear localization of diacylglycerol kinase α in K562 cells is involved in cell cycle progression. *J Cell Physiol* **232**: 2550–2557. doi:10.1002/jcp.25642
- Polychronakos C, Li Q. 2011. Understanding type 1 diabetes through genetics: advances and prospects. *Nat Rev Genet* **12**: 781–792. doi:10.1038/nrg3069
- Prinz PU, Mandler AN, Masouris I, Durner L, Oberneder R, Noessner E. 2012. High DGK- α and disabled MAPK pathways cause dysfunction of human tumor-infiltrating CD8⁺ T cells that is reversible by pharmacologic intervention. *J Immunol* **188**: 5990–6000. doi:10.4049/jimmunol.1103028
- Redondo MJ, Jeffrey J, Fain PR, Eisenbarth GS, Orban T. 2008. Concordance for islet autoimmunity among monozygotic twins. *N Engl J Med* **359**: 2849–2850. doi:10.1056/NEJMc0805398
- Revu S, Wu J, Henkel M, Rittenhouse N, Menk A, Delgoffe GM, Poholek AC, McGeachy MJ. 2018. IL-23 and IL-1 β drive human Th17 cell differentiation and metabolic reprogramming in absence of CD28 costimulation. *Cell Rep* **22**: 2642–2653. doi:10.1016/j.celrep.2018.02.044
- Rockwell CE, Zhang M, Fields PE, Klaassen CD. 2012. Th2 skewing by activation of Nrf2 in CD4⁺ T cells. *J Immunol* **188**: 1630–1637. doi:10.4049/jimmunol.1101712
- Rosen H, Goetzl EJ. 2005. Sphingosine 1-phosphate and its receptors: an autocrine and paracrine network. *Nat Rev Immunol* **5**: 560–570. doi:10.1038/nri1650
- Rosen H, Sanna MG, Cahalan SM, Gonzalez-Cabrera PJ. 2007. Tipping the gatekeeper: S1P regulation of endothelial barrier function. *Trends Immunol* **28**: 102–107. doi:10.1016/j.it.2007.01.007
- Roy S, Motsinger Reif A. 2013. Evaluation of calling algorithms for array-CGH. *Front Genet* **4**: 217. doi:10.3389/fgene.2013.00217
- Savola P, Kelkka T, Rajala HL, Kuuliala A, Kuuliala K, Eldfors S, Ellonen P, Lagström S, Lepistö M, Hannunen T, et al. 2017. Somatic mutations in clonally expanded cytotoxic T lymphocytes in patients with newly diagnosed rheumatoid arthritis. *Nat Commun* **8**: 15869. doi:10.1038/ncomms15869
- Shah N, Sukumar S. 2010. The Hox genes and their roles in oncogenesis. *Nat Rev Cancer* **10**: 361–371. doi:10.1038/nrc2826
- Taghon T, Stolz F, De Smedt M, Cnockaert M, Verhasselt B, Plum J, Leclercq G. 2002. HOX-A10 regulates hematopoietic lineage commitment: evidence for a monocyte-specific transcription factor. *Blood* **99**: 1197–1204. doi:10.1182/blood.V99.4.1197
- Taghon T, Thys K, De Smedt M, Weerkamp F, Staal FJ, Plum J, Leclercq G. 2003. Homeobox gene expression profile in human hematopoietic multipotent stem cells and T-cell progenitors: implications for human T-cell development. *Leukemia* **17**: 1157–1163. doi:10.1038/sj.leu.2402947
- Tan SY, Ooi AS, Ang MK, Koh M, Wong JC, Dykema K, Ngeow J, Loong S, Gatter K, Tan L, et al. 2011. Nuclear expression of MATK is a novel marker of type II enteropathy-associated T-cell lymphoma. *Leukemia* **25**: 555–557. doi:10.1038/leu.2010.295
- Tian Q, Taupin J, Elledge S, Robertson M, Anderson P. 1995. Fas-activated serine/threonine kinase (FAST) phosphorylates TIA-1 during Fas-mediated apoptosis. *J Exp Med* **182**: 865–874. doi:10.1084/jem.182.3.865
- Tomioka M, Toda Y, Mañucac NB, Akatsu H, Fukumoto M, Kono N, Arai H, Kioka N, Ueda K. 2017. Lysophosphatidylcholine export by human ABCA7. *Biochim Biophys Acta* **1862**: 658–665. doi:10.1016/j.bbalip.2017.03.012
- Tsoi LC, Spain SL, Knight J, Ellinghaus E, Stuart PE, Capon F, Ding J, Li Y, Tejasvi T, Gudjonsson JE, et al. 2012. Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. *Nat Genet* **44**: 1341–1348. doi:10.1038/ng.2467
- Vafiadis P, Bennett ST, Todd JA, Nadeau J, Grabs R, Goodyer CG, Wickramasinghe S, Colle E, Polychronakos C. 1997. Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus. *Nat Genet* **15**: 289–292. doi:10.1038/ng0397-289
- Venkatraman ES, Olshen AB. 2007. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23**: 657–663. doi:10.1093/bioinformatics/btl646
- Walzer T, Chiossone L, Chaix J, Calver A, Carozzo C, Garrigue-Antar L, Jacques Y, Baratin M, Tomasello E, Vivier E. 2007. Natural killer cell trafficking *in vivo* requires a dedicated sphingosine 1-phosphate receptor. *Nat Immunol* **8**: 1337–1344. doi:10.1038/ni1523
- Wang C, Mao J, Redfield S, Mo Y, Lage JM, Zhou X. 2014. Systemic distribution, subcellular localization and differential expression of sphingosine-1-phosphate receptors in benign and malignant human tissues. *Exp Mol Pathol* **97**: 259–265. doi:10.1016/j.yemp.2014.07.013
- Watson IR, Takahashi K, Futreal PA, Chin L. 2013. Emerging patterns of somatic mutations in cancer. *Nat Rev Genet* **14**: 703–718. doi:10.1038/nrg3539
- Wu C, Jin X, Tsueng G, Afrasiabi C, Su AI. 2016. BioGPS: building your own mash-up of gene annotations and expression profiles. *Nucleic Acids Res* **44**: D313–D316. doi:10.1093/nar/gkv1104
- Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC, McMichael JF, Schmidt HK, Yellapantula V, Miller CA, et al. 2014. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat Med* **20**: 1472–1478. doi:10.1038/nm.3733
- Young AL, Challen GA, Birmann BM, Druley TE. 2016. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. *Nat Commun* **7**: 12484. doi:10.1038/ncomms12484
- Zhong XP, Guo R, Zhou H, Liu C, Wan CK. 2008. Diacylglycerol kinases in immune cell function and self-tolerance. *Immunol Rev* **224**: 249–264. doi:10.1111/j.1600-065X.2008.00647.x

Received January 4, 2019; accepted in revised form November 2, 2019.