

ORIGINAL ARTICLE

Functional analyses of polymorphic variants of human terminal deoxynucleotidyl transferase

A Troshchynsky¹, I Dzneladze^{1,4}, L Chen², Y Sheng¹, V Saridakis¹ and GE Wu^{1,2,3}

Human terminal deoxynucleotidyl transferase (hTdT) is a DNA polymerase that functions to generate diversity in the adaptive immune system. Here, we focus on the function of naturally occurring single-nucleotide polymorphisms (SNPs) of hTdT to evaluate their role in genetic-generated immune variation. The data demonstrate that the genetic variations generated by the hTdT SNPs will vary the human immune repertoire and thus its responses. Human TdT catalyzes template-independent addition of nucleotides (N-additions) during coding joint formation in V(D)J recombination. Its activity is crucial to the diversity of the antigen receptors of B and T lymphocytes. We used *in vitro* polymerase assays and *in vivo* human cell V(D)J recombination assays to evaluate the activity and the N-addition levels of six natural (SNP) variants of hTdT. *In vitro*, the variants differed from wild-type hTdT in polymerization ability with four having significantly lower activity. *In vivo*, the presence of TdT varied both the efficiency of recombination and N-addition, with two variants generating coding joints with significantly fewer N-additions. Although likely heterozygous, individuals possessing these genetic changes may have less diverse B- and T-cell receptors that would particularly effect individuals prone to adaptive immune disorders, including autoimmunity.

Genes and Immunity (2015) 16, 388–398; doi:10.1038/gene.2015.19; published online 4 June 2015

INTRODUCTION

Polymorphisms are responsible for generating much of the genetic diversity of humans.^{1,2} It is of great interest and importance to understand how polymorphisms translate into differences in activity and function of our genome. In the immune system, inherited polymorphisms have been instrumental in diversifying our major histocompatibility loci that in turn is reflected in our almost limitless array of cell identity markers.^{3,4} The human adaptive immune system is a particularly useful system to study the effects of genetic changes on function in part because of the immune system's redundancies and its inherent diversity. Single-nucleotide polymorphism (SNPs) resulting in single-amino-acid changes can be found in immune-related enzymes, but the effect of these genetic changes on the immune system is largely unknown. Here, we investigate one such diversifying enzyme, human terminal deoxynucleotidyl transferase (hTdT), to determine whether and how these genetic changes may affect the T- and B-cell receptor diversity of the immune system in humans harboring the variants.

The genes encoding the B- and T-cell receptors are generated by the process of V(D)J recombination.^{5–8} This process is responsible for the primary B- and T-cell repertoire generation by choosing and joining, in an inaccurate manner, V, D and J gene segments, that when joined, form the genes encoding the receptors. V(D)J recombination is a complex process involving a cascade of enzymes specific to V(D)J recombination and a number of cellular DNA repair-associated activities.^{9–13} TdT participates at the coding end processing stage of the gene segments (V, D or J) being joined. TdT catalyzes the addition of nucleotides, known as N-addition, in a template-independent manner to the coding

ends.^{14–18} Through N-addition, TdT promotes junctional diversity among V and J, and V, D and J gene segments genetically diversifying the available antigen receptor pool of B and T cells.

TdT is an X-family DNA polymerase expressed in the primary lymphoid tissues, thymus and bone marrow. It catalyzes a nucleotidyl transfer reaction between an incoming deoxynucleotide and the 3' hydroxyl end of DNA primer strand.^{19,20} Two functionally independent TdT regions have been identified: the breast cancer susceptibility protein BRCA1 C-terminal (BRCT) domain at the N-terminus and the polymerase β -like domain at the C-terminus (catalytic domain) (Figure 1a).²¹ The crystal structure of the catalytic domain of murine TdT revealed a largely α helical structure containing a central antiparallel β sheet where the active site residues reside along with two regions (Loop1 and 2) (Figure 1b).²⁰ The TdT catalytic domain is further subdivided into four subdomains named: 8 kDa, finger, palm and thumb domains. The active sites of X-family DNA polymerases are highly conserved. Three invariant catalytic aspartate residues co-ordinate two metal ions involved in the nucleotidyl transfer reaction in the palm subdomain.^{22,23} The finger subdomain aligns the incoming nucleotide and the DNA primer strand in proper orientation for the phosphoryl transfer reaction. The role of the thumb subdomain is to position the DNA primer strand to position after catalysis. The 8 kDa subdomain contacts the thumb subdomain to create TdT's ring-like shape that allows the DNA primer strand and incoming nucleotide to reach the active site.^{14,19,20,23} Three human mRNA TdT splice variants have been identified. These translate into three mature protein isoforms: short (TdTS) and two long (TdTL1 and TdTL2). The length of nucleotide

¹Department of Biology, York University, Toronto, Ontario, Canada; ²School of Kinesiology and Health Science, York University, Toronto, Ontario, Canada and ³School of Biological Sciences, University of Canterbury, Christchurch, New Zealand. Correspondence: Dr V Saridakis or Dr GE Wu, School of Kinesiology and Health Sciences, York University, 4700 Keele Street, Room 124 Farquharson Building, Toronto, M3J 1P3 Ontario, Canada.

E-mail: vsaridak@yorku.ca or gillwu@yorku.ca

⁴Current address: Department of Medical Biophysics, University of Toronto and the Princess Margaret Cancer Center, University Health Network, Toronto, Canada.

Received 28 February 2015; revised 19 April 2015; accepted 23 April 2015; published online 4 June 2015

addition during V(D)J recombination is thought to be regulated through the combined activity of these isoforms.^{19,20,23–26}

TdT has been implicated in human diseases. The evidence is strongest in B- and T-cell leukemias (specifically acute lymphocytic leukemia and acute myeloid leukemia). Patients with leukemias that are TdT positive have a poorer prognosis than those whose leukemias are TdT negative.^{27–29} A role for TdT in autoimmune disease has been documented in mouse models but not yet in humans. TdT-knockout mice lacking N-additions have a significant loss of B-cell receptor specificity toward peptide antigens compared with TdT-positive mice. TdT-deficient lupus-prone MRL-*Fas*^{Lpr} mice demonstrated prolonged lifespan and decreased severity of disease symptoms as compared with the TdT-positive mice of this strain, suggesting that N-additions resulted in specificities in their BCRs that rendered them more susceptible to (early onset) disease.^{30–34} Similarly, TdT-deficient diabetes-prone non-obese diabetic mice showed a decrease in type I diabetes incidence when compared with littermates with active TdT.³⁵ These mouse model data led us to consider that the functional activity of TdT gene products could affect the onset and severity of autoimmune diseases in humans.

In this study, we explore the effects of six human TdT genes harboring SNPs encoding variant amino acids, by evaluating their polymerase and V(D)J recombination activities. We hypothesize that the variability in enzymatic activity by the genetic changes will affect the diversity of antigen receptors, thus potentially impacting the adaptive immune response in healthy and autoimmune-prone individuals possessing these genetic variations.

RESULTS

Selection of hTdT genetic variants

The National Center for Biotechnology Information (NCBI) SNP database was used to identify potential human TdT short isoform candidate genes with SNPs within the coding region of the protein (NCBI reference NP_004079.3). The BRCT domain of the hTdT protein was not included in the SNP search as it is not associated with TdT polymerase activity. Human TdT gene SNPs were classified based on the chemical nature of the amino-acid change and on the results of the modeling of the relative proximity of the SNP to the murine TdT active site (PDB ID 1KEJ) using *PyMol* software (Schrodinger, NY, USA) (Figures 1b and c). TdT multi-sequence alignments with other species' genes were completed to identify highly conserved residues (Supplementary Figure 1a, represented as percent similarity in Table 1), as we considered these residues were likely to be significant for TdT structure and function. Based on the above analyses, six hTdT variants with the SNPs D280H, L397S, R431C, A445T, T450S and R460Q were selected for the study. Among them, D280H was chosen to be a control because of its location at the periphery of the protein, away from the active site. The SNP analyses are detailed in Table 1.

SNPs affect TdT polymerase activity

Polymerase activity. The wild-type hTdT form (Supplementary Figure 1b) was cloned from the Jurkat cell line and the six selected variant SNPs generated using standard mutagenesis. The expressed hTdT proteins were isolated from bacteria as described

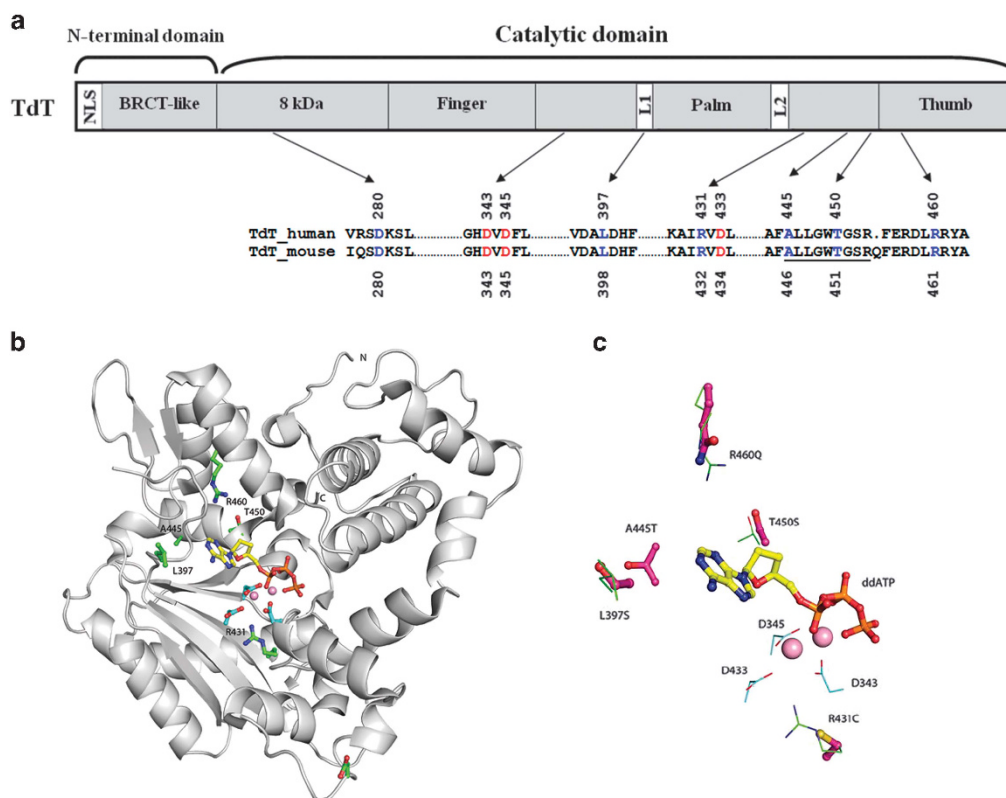


Figure 1. Positions of hTdT SNPs. **(a)** Linear arrangement of TdT domains with the positions of the six hTdT SNP variants (blue). L1 and L2 refer to loop 1 and loop 2, respectively. The three conserved aspartic acid residues (D343, D345 and D433) of the active site are shown in red. The incoming dNTP binding site motif is underlined. **(b)** The crystal structure of murine TdT (PDB ID 1KEJ) used to model the positions of hTdT SNPs. The three conserved aspartic acid residues are cyan. The incoming nucleotide substrate, ddATP, is yellow. Two cobalt ions are pink. The relative positions of the six hTdT SNPs are green. **(c)** Active site of murine TdT showing the relative positions of L397S, R431C, A445T, T450S and R460Q SNPs within 8 Å of incoming nucleotide substrate, ddATP. The SNP amino acids substituted (magenta) are superimposed on the wild-type amino-acid residues (green). Residue positions numbered according to human TdT sequence. Figure generated using *PyMol* software.

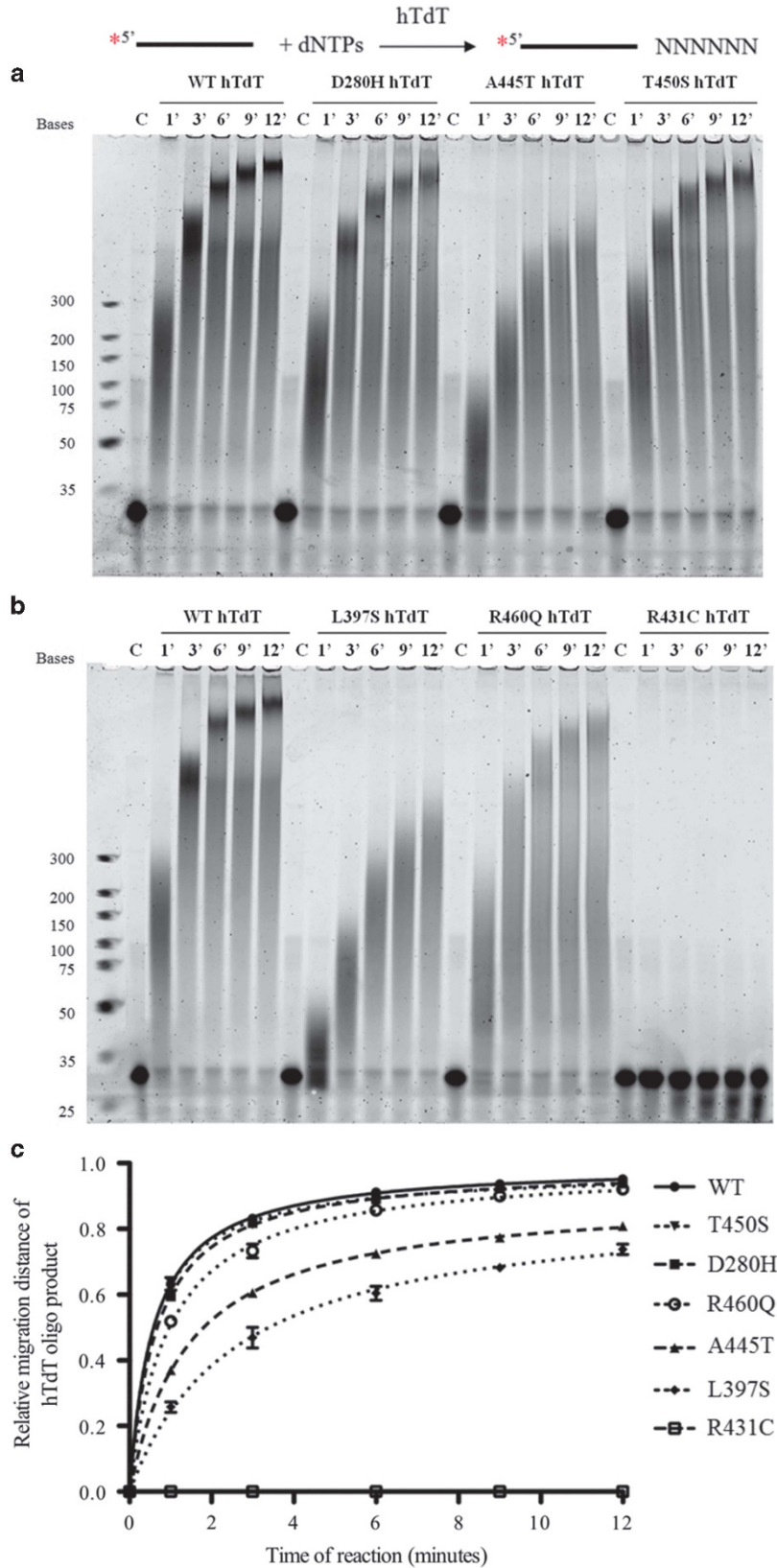


Figure 2. *In vitro* polymerase time course of purified wild-type and SNP hTdT variants. The samples were resolved on a 10% denaturing TBE-urea gel (89 mM Tris, 89 mM boric acid, 2 mM EDTA pH 8.0, 7M urea) as described in the Materials and methods section. **(a)** Polymerase time course of wild-type, D280H, A445T and T450S hTdT. **(b)** Polymerase time course of wild-type, L397S, R460Q and R431C hTdT. **(c)** Relative migration distance of hTdT products at its maximum (top of the product smear) with respect to time of the *in vitro* reaction. The distance has been normalized to the total resolving distance of the 10% TBE-urea gel (89 mM Tris, 89 mM boric acid, 2 mM EDTA pH 8, 7M urea). The assay was independently repeated three times and the standard error is shown.

in the Materials and methods section and Supplementary Figure 2 and optimal conditions for polymerization determined. Supplementary Figure 3a shows the *in vitro* oligomer extension activity assay being used to determine the optimal amount of TdT to be used in the polymerase assay (3.0 μ M), Supplementary Figure 3b, the activity of the hTdT's in the absence and presence of enzyme and dNTPs, and Supplementary Figure 3c, the WT and variant R431C in the absence and presence of the Cobalt cation.

Figure 2 is an example of polymerase activity of the wild type and the six hTdT variants under the determined conditions. Four of the hTdT variants demonstrated marked different polymerase activities compared with wild-type hTdT. Five out of the six hTdT genetic variants were functional polymerases (Figures 2a and b). R431C hTdT was non-functional as a polymerase and unable to extend the DNA oligomer substrate and thus was termed a 'dead' *in vitro* mutant. Figure 2c shows the maximum migration distance of hTdT products with respect to time of reaction. D280H and T450S have profiles of product extension similar to wild-type TdT. R460Q, A445T and L397S variants have lower polymerase activity than wild-type. These three variants extend the oligomer substrate at 83%, 60% and 41%, respectively, compared with the wild-type hTdT at the 1-min time point (Figure 2). We interpret these results to indicate that the amino-acid changed in three of the genetic variants resulted in a change in their polymerase activity.

DNA substrate preference. In V(D)J recombination, the substrate of hTdT is the opened hairpin of the coding end of the gene segments being recombined. Thus, we evaluated whether the single-stranded oligomers influenced the polymerization activity of hTdT by testing the ability of the hTdT variants to utilize different types of Cy3-labeled DNA substrates. We compared single-stranded with double-stranded oligomers with a blunt end or with a 3' end overhang (Figure 3). The oligomer substrates produced polymerized products with the enzymes tested (excluding the 'dead' R431C variant). The variation in polymerization ability paralleled that demonstrated with the single-stranded oligomer (Figure 2). The double-stranded 3' overhang oligomer substrate yielded the longest products in all reactions (Figure 3a). The double-stranded blunt end DNA substrate was the least preferred in all reactions by all the variant hTdT enzymes (Figure 3b). Ten percent of initial double-stranded 3' overhang substrate remained unused after 3 min of reaction time with each hTdT variant, compared with approximately 50% of the initial double-stranded blunt ended substrate (Figure 3c). R431C hTdT, the 'dead' variant, demonstrated no polymerase activity with any of the DNA substrate oligomers (Figures 3a–c). We conclude from these results that the genetic variants differ from the wild-type expressed gene in ways that are unlikely to be related to type of substrate in the reaction mixture.

Extrachromosomal recombination assay

A445T, L397S and R431C hTdT were chosen from the six hTdT SNP variants for the *in vivo* investigation of hTdT activity^{36,37} because they demonstrated marked differences in the *in vitro* functional assays compared with the wild-type enzyme.

As described in the Materials and methods section, a human embryonic kidney cell line, HEK293T, was used in the extrachromosomal V(D)J recombination assays. It was examined for endogenous hTdT using reverse transcriptase PCR and western blotting. As expected, as it is not a lymphocyte, HEK293T did not express endogenous TdT (Supplementary Figure 4) therefore any detected hTdT expression is dependent on ectopic expression of transfected hTdT variants.

V(D)J recombination frequencies (R-values). The *in vivo* functional activity of the hTdT genetic variants was evaluated using the extrachromosomal recombination assay as described in the Materials and methods section. HEK293T cells were co-transfected with plasmids that express RAG1, RAG2 and hTdT genes, required for V(D)J recombination, and the extrachromosomal substrate plasmid pGG51 (Figure 4a). The deletion of DNA in pGG51 between its two RSS regions via V(D)J recombination generates a recombined plasmid that confers chloramphenicol resistance on bacteria harboring it (Figures 4b and c). The frequency with which pGG51 was recombined, the R-value, was determined as the percentage of ampicillin-resistant bacterial colonies that were doubly resistant to ampicillin and chloramphenicol (Supplementary Table 1). Figure 5a plots individual R-values and the mean R-value of seven independent HEK293T cell transfections. As expected, owing to the nature of the cell line (a human embryonic kidney cell line and not an endogenously recombination competent RAG+ lymphocyte), the R-values are low; lower than reported in the literature using RAG-positive lymphocyte lines.^{38,39}

V(D)J mediated recombination occurred in every experimental condition that included RAG1 and RAG2. As has been observed by others,²⁶ the highest R-values occurred in the absence of any hTdT with the mean R-value being twofold higher than that for the wild-type hTdT gene transfection (0.0066% versus 0.0032%). We interpret this result to indicate that either the transfection efficiency was affected by four (rather than three) plasmids or hTdT itself may add a complexity to the coding end joining process that can lower its efficiency.²⁶ The mean R-value for L397S hTdT is 1.5 fold higher when compared with wild-type hTdT (0.0050% versus 0.0032%). No statistically significant differences were observed among A445T, R431C and wild-type hTdT R-values. We conclude that the absence of hTdT and the presence of L397S hTdT in the transfection mix have a positive effect on recombination competence and no variant has a negative effect.

Analysis of V(D)J recombined substrate pGG51 sequences. Extrachromosomal substrate plasmid pGG51 DNA was purified from

Table 1. Human TdT SNP information

SNP ID number	mRNA position	Base change	Amino-acid position	Amino-acid change	Chemical nature of amino-acid change	Amino-acid homology across species ^a
rs41291616	1008	G to C	280	Asp to His	Negatively to positively charged	100%
rs142078021	1360	T to C	397	Leu to Ser	Non-polar to polar	100%
rs139202533	1461	C to T	431	Arg to Cys	Positively charged to polar	100%
rs149947558	1503	G to A	445	Ala to Thr	Non-polar to polar	100%
rs145011041	1518	A to T	450	Thr to Ser	No change; polar to polar	95%
rs142141382	1549	G to A	460	Arg to Gln	Positively charged to polar	100%

Abbreviations: SNP, single-nucleotide polymorphism; TdT, terminal deoxynucleotidyl transferase. ^aPercent amino-acid identity based on amino-acid sequence multi-alignment across all available TdT protein sequences (NCBI database, Supplementary Figure 1).

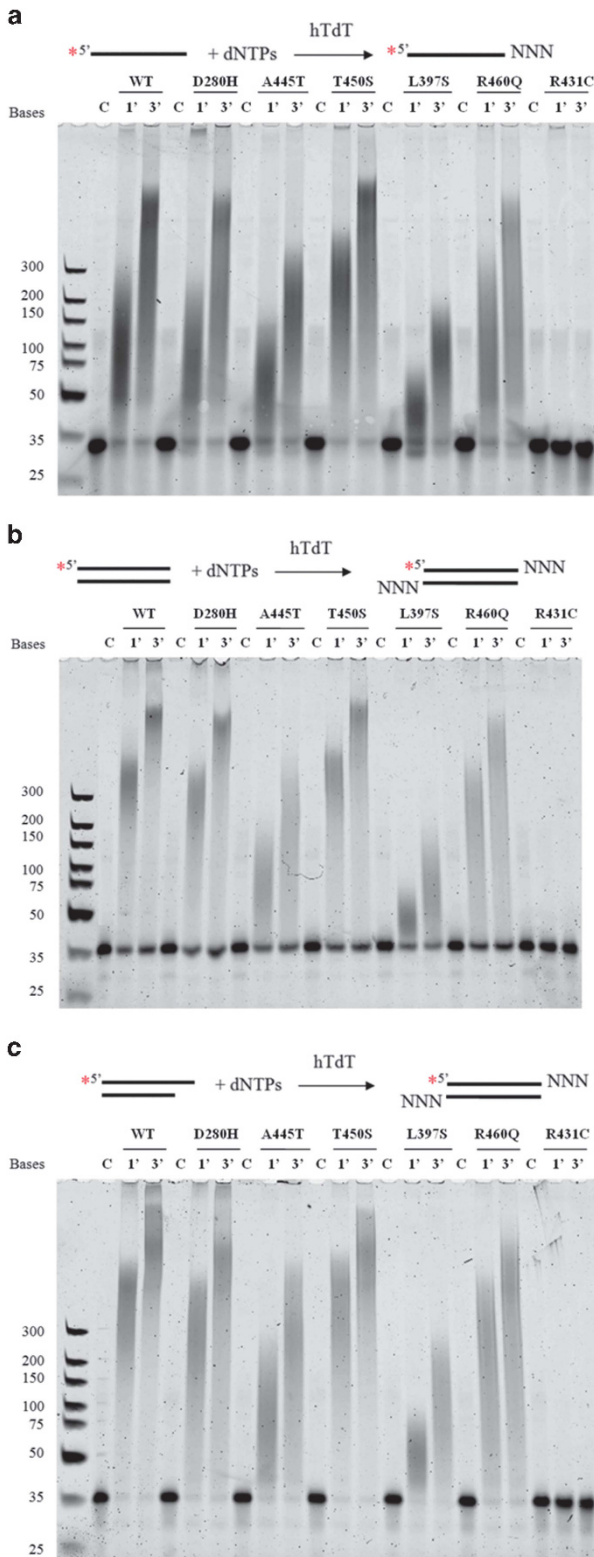


Figure 3. *In vitro* DNA substrate preference polymerase activity of purified hTdT wild-type and SNP variants. The polymerase reaction contained Cy3 oligo: (a) single-stranded Cy3-labeled oligonucleotide; (b) double-stranded Cy3-labeled blunt end oligonucleotide; (c) double-stranded Cy3-labeled 3' overhang oligonucleotide.

bacteria that were both chloramphenicol and ampicillin resistant and sequenced to reveal the V(D)J recombination generated coding junctions. Two hundred seventy-six unique recombined substrate plasmid sequences were obtained from the tested experimental conditions (Figure 4c and Supplementary Table 2 details all unique sequences). Repeated recombined sequences from within the same transfection were excluded from the analyses as they could be the result of replication of pGG51 in HEK293T cells after V(D)J recombination. The coding end junctions of the sequences were analyzed with respect to nucleotide additions, the frequency of N-additions, the distribution of N-nucleotide additions per recombined joint and the G/C base content of the additions.

N-additions. The majority of random non-templated nucleotides (N-addition) at the recombined joints of the antigen receptor genes are dependent on TdT.⁹ Therefore, the number and type of N-additions observed at the joints provide a direct measure of the activity of the polymerases (Table 2). In the absence of any hTdT (negative control), two single N-additions were present in the 48 recovered sequences. In the presence of wild-type hTdT, the highest average number of N-additions of all hTdT variants tested was observed, namely, an average of 2.2 nucleotides per joint when all sequences were considered and 3.0 nucleotides when only sequences with N's were considered. The length of the N-additions are lower than those reported in the literature with recombination competent lines^{40,41} and, similarly to the R-values, likely reflect the non-endogenous V(D)J recombination ability of the human embryonic kidney cell line. The average length of the N-additions generated by the variant hTdT A445T was not significantly different from wild-type values (1.9 nucleotides per joint when all sequences were considered and 2.4 nucleotides when only sequences with N's were considered). Coding joints from L397S and R431C hTdT transfections contained significantly fewer N-additions per sequence when all sequences were considered (0.47 and 0.32 nucleotides per joint, respectively) and R431C generated sequences had significantly fewer additions per joint than wild-type when only sequences with N were considered (1.4 nucleotides for R431C versus 3.0 for WT).

The proportion of recombined sequences containing N-additions was analyzed (Table 2). In the negative control transfection, as expected, only 4.2% of all recombined sequences contained N-additions, whereas, also as expected, 74% of sequences from wild-type hTdT transfection contained N-additions (P -value < 0.0001). A445T hTdT transfections produced similar proportions of N-containing joints when compared with wild-type hTdT transfections (78% versus 74%). However, both L397S and R431C hTdT transfections generated a significantly lower proportion of N-containing joints as compared with wild-type hTdT gene transfections (both at 24% versus 74%, P -values < 0.0001). The hTdT variants, L397S and R431C, have over a threefold difference in the frequency of N-addition in their recombined sequences compared with wild-type hTdT.

It is generally accepted that N-additions are most commonly G's or C's.¹⁹ In the joints from wild-type and A445T hTdT transfections, 61% and 72% of the N-nucleotides were G or C (G/C) bases, respectively. However, in the R431C hTdT transfections, only 27% were G/C bases (P -value 0.0164 when compared with wild-type gene transfections). This value is curious and will be discussed below.

The distribution of N-nucleotide additions per single recombined joint is shown in Figure 5b. The distribution of A445T and L397S hTdT was similar to wild-type hTdT. A striking difference was observed in the R431C hTdT transfection sequences in which 75% of all N-containing joints were 1N-nucleotide additions.

Deletions and P-additions. Given what is known about the mechanism of joining, the number and frequency of deletions

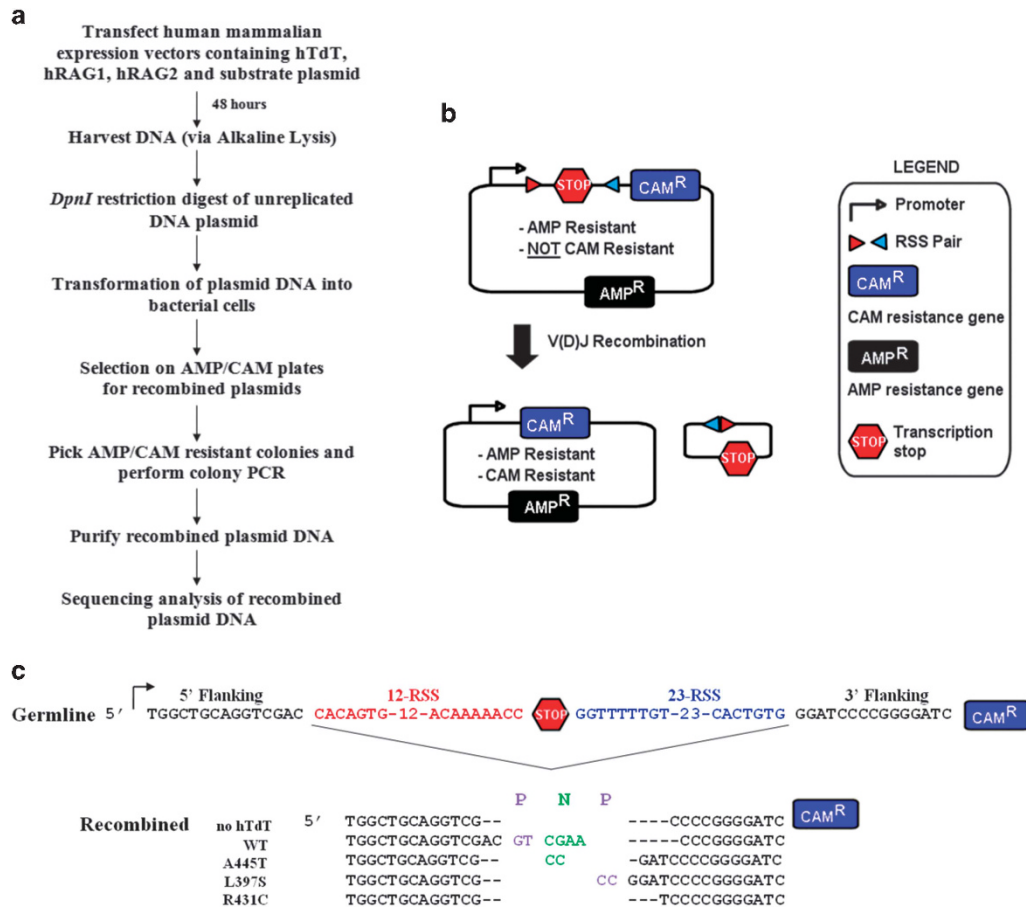


Figure 4. The *in vivo* extrachromosomal recombination assay. (a) Overview. (b) Design of the recombination substrate plasmid, pGG51 in the unrecombined configuration and the recombined configurations of the coding joint and of the signal joint products. (c) Examples of the recombined substrate plasmid sequences.

and P-additions at the coding ends would not be expected to be affected by the short form of hTdT (hTdT_S). When nucleotide deletions at the recombined joints were enumerated, the presence of hTdT in the transfection mixture did not affect deletions (Table 3). There were no statistically significant differences between the negative control transfection, lacking the hTdT gene, and those with the hTdT genes. Similarly, as detailed in Table 3, the presence of hTdT in the transfections did not significantly influence P-nucleotide addition.

DISCUSSION

Single-amino-acid changes in genes have long been known to cause major, as well as subtle changes in the resulting protein's structure and function.^{42,43} One would, *a priori*, expect subtle changes in function of some of the gene variants of human TdT assessed here. Nonetheless, it was surprising that the hTdT SNP gene variants tested had observable, and in two cases major effects on hTdT function, phenotypes that could modify the functioning of the immune system. Albeit humans are diploid, a person harboring these variant hTdT SNPs, particularly the *in vitro* 'dead' variant R431C, will likely have fewer N-additions and hence a lower level of coding joint diversity than individuals who do not have these amino-acid changes. The data herein give reason to speculate that human's varying responses to pathogens, non-self and self, so exquisitely encoded in the B- and T-cell receptors of the adaptive arm of the immune system, are genetically determined by the SNPs one harbors in genes affecting immune diversity. TdT is one such gene.

Six non-synonymous hTdT SNPs in proximity to the active site of hTdT were chosen from the NCBI SNP database as the best candidates to have altered polymerase activity (Table 1). The active site of hTdT contains three highly conserved aspartic acid residues (D343, D345 and D433) and a cation-binding site that aids in stabilizing the negative charge associated with a triphosphate group of an incoming dNTP (Figures 1b and c). Two gene variants had SNPs positioned within one of the conserved-sequence motifs forming the dNTP binding site (consensus sequence ALLGWTGSR (residues 445–453) (A and T are two of the variants studied, namely A445T (small non-polar residue to polar residue) and T450S (polar residue to polar residue)).²⁰ *In vitro*, T450S's change from wild-type was a decrease in the amount of product, but it was not a significant change leading us to conclude that this conservative (polar to polar) change did not strongly influence polymerase activity (Figure 2). A445T had a noted decrease in polymerase activity *in vitro* (Figure 2) but *in vivo* A445T hTdT demonstrated N-addition activity comparable to the wild-type enzyme in terms of the number of N-containing joints, frequency of N-additions, the N's per joint and the proportion of G/C bases (Table 2, Supplementary Table 2). We conclude that a person carrying the A445T hTdT allele may have a subtle change but would not be expected to have a strong reduction in B- and T-cell receptor repertoire diversity.

The L397S SNP is located within Loop1, which is associated with template-independent polymerase activity (ESTFEQPSRVKDALDH, residues 382–401) (Figure 1b). Romain *et al.*⁴⁴ demonstrated that a 13 residue deletion of Loop1, including L397, managed to switch the template-independent activity of TdT to template-dependent

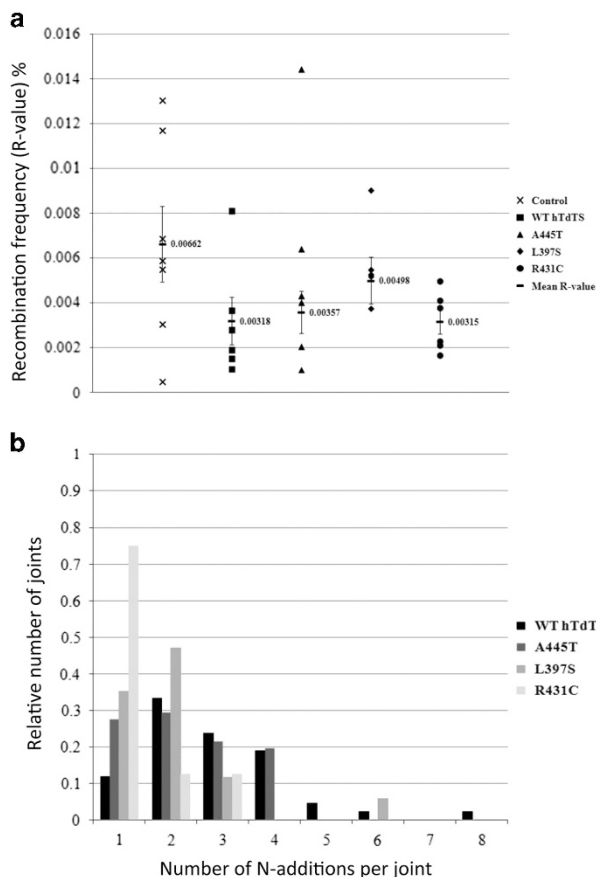


Figure 5. Analyses of extrachromosomal recombination assays. **(a)** Recombination frequencies of the *in vivo* V(D)J recombination assay. The individual R-values computed based on seven independent HEK293T cell transfections per single experimental condition. The mean R-values are shown along with the standard error, represented by error bars. The mean R-values: Control = $0.00662 \pm 0.00169\%$; Wild-type hTdT = $0.00318 \pm 0.00106\%$; A445T hTdT = $0.00357 \pm 0.00094\%$; L397S hTdT = $0.00498 \pm 0.00104\%$; R431C hTdT = $0.00315 \pm 0.00054\%$. **(b)** Distribution of N-nucleotide additions per single recombined joint. The frequencies of N-nucleotide occurrence were computed as a ratio of the number of joints having a certain number of N-additions to the number of joints containing N-additions.

activity comparable to polymerase λ that lacks a Loop1 region. Structural studies led to the conclusion that the Loop1 region likely sterically excludes double-stranded templates. Site-directed mutagenesis studies of L397 by Gouge *et al.*²², revealed its role in base stacking interruption of the incoming DNA strand, thus participating in the closing of Loop1 onto the substrate DNA strand. The data presented here also reveal the importance of this position for TdT activity as the L397S variant has impaired polymerase activity *in vitro* and impaired N-additions in the *in vivo* V(D)J assay. The SNP mutation may disrupt an essential contact of Loop1 with substrate DNA, leading to a reduction in catalytic efficiency of TdT resulting in a lower frequency of N-containing joints and a lower average number of N-additions compared with wild-type enzyme (Table 2).

Variant R431C had the most uncharacteristic and impaired TdT activity. We considered whether its production in bacteria and in the cell line could hold the reason for its impaired activity. Variant R431C had the same apparent size and the same purity as the other recombinants and wild-type hTdT (Supplementary Figure 2). Moreover, each of the hTdT proteins were isolated from bacteria at least twice and assayed completely more than five times each.

In no case did any of the purification procedures change the characteristics of the wild type or variants in the assays of their function. However, it remains a possibility that R431C had a post-translational modification in the bacteria that the other isolates did not have. The residue R431 forms a salt bridge with the catalytic D343, which contributes to TdT catalysis and stability of the active site. The presence of a cysteine instead of an arginine likely disrupts the essential salt bridge, thus destabilizing the active site of TdT. The importance of this position for TdT function was demonstrated by variant R431C's loss of polymerase activity (a 'dead' mutant) in the *in vitro* assay, and having the lowest and uncharacteristic activity in the *in vivo* V(D)J assay. We also considered whether variant R431C's *in vitro* lack of activity actually was revealing a gained exonuclease ability because of the disruption of the salt bridge. Its profile in the *in vitro* polymerase assays, especially in the absence of the cobalt cation (Supplementary Figure 3c) revealed what we interpreted as probable exonuclease activity. Although an appealing hypothesis, which would explain R431C's *in vitro* profile, the *in vivo* V(D)J recombination assay, revealed no evidence of exonuclease activity (Table 2). The number of deletions at the ends was not statistically different from any of the hTdT's assayed in the analyses of all unique R431C sequences (4.4 nt deleted on average) nor in the analyses of sequences without any additions (4.8 nt deleted on average). (We reasoned that perhaps an exonuclease activity would be revealed in joints without additions). Nevertheless, R431C's activity is markedly different from WT. Although WT hTdT had 23% (13/57) of its joints without N or P-additions, R431C had 62% (21/34). In addition, it proved more difficult to recover recombinants in the transfections with R431C (34 recombinants recovered compared with 57, 65, 72 recombinants for the other hTdT's, and more transformations were needed to obtain the 34) (Tables 2 and 3, Supplementary Table 2). However, based on R431C's inability to function *in vitro*, we were surprised that there were any joints with N-additions (24%, Table 2). It has been well established that polymerases other than TdT can be responsible for template-independent additions at DNA breaks. Indeed, TdT knockout mice have template-independent additions at nearly 5% of their V(D)J junctions generated during normal development in B and T lineage cells.^{15,16,31,35} Polymerase μ possesses both template dependent and template-independent polymerase activity and therefore may be responsible for adding these non-templated bases.^{23,41} It remains unknown what polymerases are present in HEK293T or, what repair polymerases are activated or induced by the transfection stress, and it is unclear whether the presence of the TdT enzyme affects other polymerase's activity on non-homologous end joining.^{41,45,46}

It is generally accepted that random additions are beneficial for annealing of the coding ends during non-homologous end joining because of 1 or 2 nucleotides of terminal microhomology.^{15,16,41} In this study, the 2 out of 48 (4%) joints that had 'N'-additions under non-TdT conditions may be due to the microhomologies (both had 1 N-addition per joint). Seventy-five percent of the joints recovered from the R431C-containing transfection contained a single N-addition (6/8) (versus 14% (6/42) for joints from wild-type transfections). The G/C base frequency was 27% (four A's, one C, two G's, four T's, Supplementary Table 2), compared with WT's hTdT G/C frequency of 61%. Even so, the finding that 24% of the joints were N-containing joints in the presence of R431C gene variant indicates that the R431C variant has some influencing activity *in vivo*, but R431C's activity is not that of a highly diversifying TdT polymerase. *In vivo*, N-additions are usually in the range of one to five nucleotides.¹⁸ A length limit is required in order to properly maintain the shape and size of the encoded antigen receptor. Excess nucleotide additions between V, D and J gene segments would result in a misfolded and non-functional receptor. On the other hand, TdT's *in vitro* polymerase ability is extensive (a trait used commercially for end labeling, for

Table 2. Analyses of N-additions in the V(D)J recombination generated coding joints from pGG51

	Control (no TdT)	W-T hTdT	A445T hTdT	L397S hTdT	R431C hTdT
Number of unique sequences	48	57	65	72	34
Number containing N-additions	2	42	51	17	8
Proportion containing N-additions (%)	4.2	74	78	24	24
<i>P</i> -value		< 0.0001	0.6704	< 0.0001	< 0.0001
<i>N</i> -additions					
Number of N-additions	2	126	122	34	11
Average N-additions in sequences with N's	1.0	3.0	2.4	2.0	1.4
<i>P</i> -value		0.0336	0.0335	0.0218	0.0062
Average N-additions in all unique sequences	0.042	2.2	1.9	0.47	0.32
<i>P</i> -value		< 0.0001	0.27	< 0.0001	< 0.0001
<i>Frequency of G/C bases</i>					
Number of G/C bases in N-additions (% of total N-additions)		77	88	16	3
% Of N-additions that are G or C		61	72	47	27
<i>P</i> -value			0.0814	0.1691	0.0164

Abbreviation: hTdT, human terminal deoxynucleotidyl transferase. *P*-values for WT hTdT experimental conditions computed using a comparison with the control condition. All other *P*-values computed through a comparison with WT hTdT experimental conditions. Average N-additions in all sequences = total number of N-additions/total number of unique sequences. Average N-additions in sequences with N = total number of N-additions/number of sequences containing N-additions. Proportion of sequences is the %. Percentage of AT or GC bases in N-additions (%) = (total number of AT or GC bases/total number of N-additions) × 100%.

Table 3. Deletions and potential P-additions in V(D)J recombination generated coding joints

	Control (no TdT)	WT hTdT	A445T hTdT	L397S hTdT	R431C hTdT
Number of unique sequences	48	57	65	72	34
<i>Deletions</i>					
Number of nucleotides deleted	230	258	298	372	148
Average deletion	4.8	4.5	4.6	5.2	4.4
<i>P</i> -value		0.50	0.86	0.072	0.66
<i>P</i> -additions					
Number of P-additions	19	18	25	18	12
Number of sequences with P-additions	12	10	16	13	8
Proportion containing P-additions (%)	25	18	25	18	24
<i>P</i> -value		0.30	0.23	0.049	0.20

Abbreviation: hTdT, human terminal deoxynucleotidyl transferase. *P*-values for WT hTdT experimental conditions computed using a comparison with the control condition. All other *P*-values computed through a comparison with WT hTdT experimental conditions. Average number of deletions = total number of nucleotide deletions/total number of unique sequences. Proportion of sequences containing P-additions is the % of the number with P-additions/number of unique sequences.

example, BD Biosciences TUNEL assays, Mississauga, ON, Canada). The associated DNA repair factors, *in vivo*, may limit TdT's polymerase activity, and thus the *in vitro* assays (Figure 2) are what demonstrate their inherent variant activity. When the variant enzymes are in the conditions that have the associated V(D)J *in vivo* constraints and DNA-binding components, gene variant L397S, as well as the *in vitro* 'dead' variant, R431C, were unable to effectively add N-nucleotides (nucleotides added per unique joint: 0.47, 0.32 and 0.042 for L397S, R431C and the reaction without TdT, respectively), compared with 2.2 and 1.9 nucleotides per unique joint for wild-type and A445T, respectively. It seems likely that the *in vitro* assay reveals the polymerase ability of hTdT and the *in vivo* assay reveals a modified enzyme function in the presence of the other components of coding end joining.

The substrate for TdT during the V(D)J recombination is believed to be variable, as the hairpinned coding end may be opened in a way that generates a bunt ended double-stranded DNA or an overhang. All three of the DNA substrates tested in the *in vitro* assays were extended by the wild-type and by the variant hTdT's (Figure 3) with the blunt ended double-stranded DNA the least preferred and the double-stranded 3' overhang DNA

substrate the most utilized substrate (Figure 3c). Sterically, a hairpin opening that produces a 3' overhang would be expected to be the most effective opening given the DNA polymerization reaction. The fact none of the variants had a substrate preference different from the wild-type hTdT indicates that the regions where the changed amino acids reside are not in the substrate-entering region but rather in the polymerization region as predicted by the murine model.^{22,35,44}

The specific role of TdT in autoimmune models remains controversial. Absence of TdT in inbred mice has been shown to correlate with decreased severity of autoimmune diseases, such as lupus and diabetes.^{30,31,35} TdT deficiency has been linked to a decreased number of autoantibody-producing cells. This deficiency may provide a protective role against autoimmunity because of lower polyreactivity toward self-antigens and reduction in autoantibody affinities. It has been speculated that this autoimmune protection is the reason for the delayed onset of TdT activity during fetal development.^{30,31,35,47} Although the effect of these alleles in the heterozygous state is undetermined, autoimmune-prone individuals carrying a low activity TdT allele (L397S and R431C in this study) in the heterozygous state may

experience a lesser extent of autoreactivity that may then promote disease control.

TdT overexpression in B- and T-cell acute lymphocytic leukemia and acute myeloid leukemia generally has a poorer prognosis. More than 90% of acute lymphocytic leukemia leukemic cells show high TdT activity, with remission rates of TdT-positive patients twofold lower as compared with TdT-negative patients.^{19,28,29} Individuals carrying the low polymerase activity alleles may be expected to be protected from TdT generated mutations because of the effect of those alleles although the influence of the other (probably wild type) allele is not known. As genetic testing has become commonplace and personalized medicine a reality, the presence of some TdT alleles could be a positive marker for people with TdT+ leukemias.

The decreased activity of the tested variants will affect the junctional diversity at V(D)J junctions of antigen receptors of B and T cells. As TdT is critically responsible for antigen receptor diversity allowing recognition of broad pathogens by the adaptive immune system, diminished activity would be expected to impair the development of a diverse antigen receptor repertoire needed for the immune system to promote an adequate response to pathogens. A hTdT polymorphism on one allele may not have a profound effect on antibody diversity, although someone with two such polymorphisms would have such defects. Thus, although the effect of these alleles in the heterozygous state is undetermined, heterozygous, healthy individuals carrying hTdT polymorphisms, particularly, L397S and R431C, may have variations in their adaptive immune response, adding to the diversity generated by genetic variation. Humans carry many polymorphisms with the obvious resulting diversity of the population.

MATERIALS AND METHODS

Human TdT protein purification for *in vitro* polymerase activity assays

The wild-type hTdT short form was isolated from Jurkat cell RNA (Applied Biosystems, Foster City, CA, USA) using normal complementary DNA (cDNA) cloning methods (Supplementary Figure 1b). The hTdT cDNA was cloned into p15TV-L, a bacterial expression vector (GenBank accession EF456736) and six single-nucleotide mutations of hTdT gene, corresponding to the six hTdT SNP candidates, were generated using site-directed mutagenesis (QuickChange Agilent Technologies, Mississauga, ON, Canada). Correct clones were identified using standard procedures and the inserts sequenced for verification. The hTdT variants were over-expressed and purified from *E. coli* BL21(DE3) cells. Isopropyl β -D-1-thiogalactopyranoside-induced BL21(DE3) cultures were centrifuged at 1500xg at 4 °C for 30 min. The bacterial pellets were resuspended in lysis buffer (50 mM Tris pH 7.5, 500 mM NaCl, 5 mM imidazole, 1 mM benzamidine, 0.5 mM phenylmethylsulfonyl fluoride and 1 mM β -mercaptoethanol) and sonicated for 8 min on ice. Cell debris was removed by centrifugation at 40 000xg at 4 °C for 30 min. The supernatant was incubated with nickel-nitrilotriacetic acid (Ni-NTA, Qiagen, Toronto, ON, Canada) metal-affinity chromatography resin for 1 h at 4 °C with rotation. The beads were washed three times with wash buffer (50 mM Tris pH 7.5, 500 mM NaCl, 20 mM imidazole, 1 mM benzamidine, 0.5 mM phenylmethylsulfonyl fluoride and 1 mM β -mercaptoethanol). His-tagged hTdT wild-type and variants were eluted with elution buffer (50 mM Tris pH 7.5, 500 mM NaCl, 500 mM imidazole, 1 mM benzamidine, 0.5 mM phenylmethylsulfonyl fluoride and 1 mM β -mercaptoethanol) and dialysed overnight at 4 °C in TdT storage buffer (50 mM K_2PO_4 , 100 mM NaCl, 1.43 mM β -mercaptoethanol, 50% glycerol, 0.1% Triton X-100 pH 7.3 at 25 °C), as per storage conditions of recombinant calf TdT (New England Biolabs, Whitby, ON, Canada). Protein samples were analyzed by 10% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (0.375 M Tris pH 8.8, 0.1% sodium dodecyl sulfate) (Supplementary Figure 2) and used for the *in vitro* polymerase activity assays. Protein samples from each plasmid were purified separately at least twice and their characteristics in the assays did not vary.

In vitro polymerase activity assay

An *in vitro* polymerase activity assay was used to determine the optimal amount of TdT by varying its concentration up to 4 μ M (Supplementary Figure 2a). Its extension ability was also evaluated in the absence and presence of the enzyme, dNTPs and 0.25 mM cobalt chloride ($CoCl_2$) (Supplementary Figures 2b and c). The reaction conditions of the *in vitro* oligomer extension assays were set up according to a DNA tailing reaction recommended by New England Biolabs. The standard components of are: 0.3 μ M single-stranded fluorescently labeled Cy3 oligomer (5'-Cy3-CTACTGGTACTTCGATCTCTGGGGCCGTGACGC-3') (Sigma-Aldrich, The Woodlands, TX, USA), 0.15 mM dNTP mix, 0.25 mM $CoCl_2$, 1xTdT reaction buffer (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate at pH 7.9 at 25 °C; New England Biolabs), 3.0 μ M of each purified hTdT protein for a total reaction volume of 30 μ l. The reaction was incubated at 37 °C for various time points and stopped by the addition of an equal volume of TdT stopping buffer (95% formamide, 50 mM EDTA, 0.05% bromophenol blue). The completed hTdT reactions were then boiled for 5 min to denature any duplexed DNA and immediately placed on ice. The samples were resolved on a 10% denaturing TBE gel (89 mM Tris, 89 mM Boric Acid, 2 mM EDTA pH 8.0, 7M urea) and the Cy3-labeled DNA was visualized by the Typhoon Trio Biomolecular Imager (GE Healthcare, Baie-d'Urfé, QC, Canada) (scanning done at an excitation wavelength of 532 nm, and an emission wavelength of 580nm) using ImageQuant software (Molecular Dynamics, GE Healthcare).

In vitro DNA substrate preference assay

The DNA substrate was varied under the same reaction conditions as described above for the *in vitro* polymerase activity assay. Purified wild-type and mutant hTdT variants were incubated with either 0.03 nM of single-stranded Cy3 oligomer (5'-Cy3-CTACTGGTACTTCGATCTCTGGGGCCGTGACGC-3'), or of a double-stranded Cy3 oligomer having a blunt end (5'-Cy3-CTACTGGTACTTCGATCTCTGGGGCCGTGACGC-3', 5'-GCGTCACGGCCCCAGAGATCGAAGTACCAGTAG-3') or of a double-stranded Cy3 oligomer having a 3' end overhang of six nucleotides (5'-Cy3-CTACTGGTACTTCGATCTCTGGGGCCGTGACGC-3', 5' CGGCCCCAGAGATCGAATGACCAGTAG-3'). The double-stranded Cy3 oligomer substrates were obtained by annealing the single-stranded Cy3-labeled oligomer to an excess of unlabeled cDNA oligomer using standard procedures.

Endogenous TdT expression evaluation

Reverse transcriptase PCR procedures. Reverse transcriptase PCR was performed to test for endogenous TdT production in HEK293T cells. The cells were passaged for at least a week before to transfection in Dulbecco's modified Eagle's medium/high glucose media (Sigma-Aldrich's Dulbecco's modified Eagle's medium—high glucose with 4500 mg l⁻¹ glucose, L-glutamine and sodium bicarbonate) supplemented with 10% fetal bovine serum at 37 °C and 5% CO₂. Cells were seeded for transfection in 60 mm tissue culture dishes at ~80–90% confluency in a total volume of 2.5 ml culture media. One hour before transfection, cells were washed with warm 1xphosphate-buffered saline (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 2 mM KH₂PO₄) and fresh culture media was added for a total volume of 2.5 ml per 60 mm dish. HEK293T cells were transfected (using Poly-Jet DNA transfection reagent (FrogggaBio, North York, ON, Canada)) with either substrate recombination plasmid pGG51, empty pcDNATM4/myc-His A vector, wild-type hTdT pcDNATM4/myc-His A expression vector or co-transfected with pGG51 plasmid, wild-type hTdT pcDNA, pEBG-hRAG1 and pEBG-hRAG2 expression vectors.

The reverse transcription reaction used the ThermoScript RT-PCR system (Invitrogen, Burlington, ON, Canada). First cDNA synthesis reaction contained 150ng of RNA sample, 50 μ M oligo (dT)₂₀ primer, 1.5 mM dNTP mix and diethyl dicarbonate-treated water for a total volume of 12 μ l. The reaction was incubated in a thermocycler at 65 °C for 5 min to denature the sample and then placed on ice. Next, cDNA synthesis mix was added to the reaction (4 μ l of 5x cDNA synthesis buffer, 1 μ l of 0.1M DTT, 40 units of RNaseOUT, 15 units of ThermoScript RT, diethyl dicarbonate-treated water up for a total volume of 8 μ l). The 20 μ l cDNA reaction was incubated in a thermocycler for 50 min at 50 °C, followed by reaction termination for 5 min at 85 °C.

PCR amplification reaction of target cDNA was set up using hTdT gene-specific primers that were designed to span an intron–exon boundary (DNIT_F and DNIT_R 5'-CACCAAGCTTGTGTGAGAAGAGAC-3', 5'-CTCTCTCAACTGCCGGGAGCCAGT-3', respectively). Each PCR reaction included 2 μ l cDNA, 1x reaction buffer (10 mM Tris-HCl (pH 9.0), 50 mM KCl, 1% Triton

X-100, 1.5 mM MgCl₂), 50 pmol DNTT_F and DNTT_R primers, 0.2 mM dNTP mix, 2.5 units EconoTaq DNA Polymerase (Lucigen, Mississauga, ON, Canada) and dH₂O to 50 μ l. The negative PCR control reaction did not include cDNA template, the positive PCR control reaction included 0.25 μ g wild-type hTdT pcDNA expression vector as template DNA. Following PCR amplification, the reaction products were electrophoresed on a 0.7% agarose gel stained with ethidium bromide and visualized under ultraviolet light.

Western blot procedures. Endogenous protein hTdT expression was evaluated by western blot using rabbit polyclonal hTdT-specific antibody (Novus Biologicals, Burlington, ON, Canada, #NBP1-58254) raised against N-terminal peptide sequence of hTdT (FQDLVVFLEIKMGTRRAFLME-LARRKGRFVENELSDSVTHIVAENNSG). HEK293T cells were transfected (using Poly-Jet DNA transfection reagent; FroggaBio) with either wild-type hTdT pcDNATM4/myc-His A expression vector or co-transfected with substrate recombination pGG51 plasmid, wild-type hTdT pcDNA, pEBG-hRAG1 and pEBG-hRAG2 expression vectors. Transfected and non-transfected HEK293T cells were harvested 48-h post-transfection. The cells were spun at 1500 r.p.m. for 5 min at room temperature and the supernatant removed. The cell pellet was resuspended in 75 μ l RIPA lysis buffer (150 mM NaCl, 50 mM Tris pH 8.0, 0.5% NP-40, 1x protease inhibitor). The samples were incubated for 20 min at 4 °C on a rotator. The cells were then further lysed by sonication (0.5-s pulse on, 2-s pulse off, at 10% amplitude for a total pulse time of 3 s). The lysed samples were centrifuged at 15 000 r.p.m. for 5 min at 4 °C and loaded onto an sodium dodecyl sulfate–polyacrylamide gel electrophoresis gel. The protein samples were then transferred onto a membrane and blocked in 1xPBST (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 2 mM KH₂PO₄, 0.1% Tween-20 pH 7.4) containing 5% milk (w/v) for 2 h, on a rocker, at room temperature. TdT protein was detected with 1:1000 dilution of rabbit polyclonal TdT antibody (Novus Biologicals #NBP1-58254); GAPDH was detected with 1:5000 dilution of mouse monoclonal GAPDH antibody (Santa Cruz Biotech, Dallas, TX, USA, #SC-47724). The secondary antibodies were anti-rabbit horseradish peroxidase-conjugated antibody (Promega, Madison, WI, USA, #W4011) and anti-mouse horseradish peroxidase-conjugated antibody (Promega #W4021). In the final step, Clarity Western ECL reagent (Bio-Rad, Mississauga, ON, Canada) was added to the membrane and developed in the dark.

Extrachromosomal recombination assay

V(D)J recombination assay plasmids. The V(D)J recombination substrate, pGG51, has been previously described (a generous gift of Dr Michael Lieber, University of Southern California, Los Angeles).^{40,48} It contains both eukaryotic SV40 and prokaryotic ColE1 replication origins. Plasmid pGG51 contains a 12-RSS (recombination signal sequence), and 23-RSS separated by a prokaryotic transcriptional termination sequence (Figures 4b and c). A chloramphenicol resistance gene is downstream of the prokaryotic transcription termination sequence; it is only expressed (and bacteria harboring it are only resistant to chloramphenicol) after deletion of the transcription termination sequence. A V(D)J recombination process acting on the RSS of the plasmid results in a product that has deleted the two RSS's (and the transcription termination sequence) and retains a coding joint. The recombined plasmid confers ampicillin and chloramphenicol resistance onto bacteria harboring it (Figure 4c).

The generated p15TV-L bacterial expression vectors containing wild-type hTdT and hTdT variants were used in cloning hTdT into the pcDNA 4/myc-His A mammalian expression vector. The hTdT genes were excised from the bacterial expression vector, using *Hind*III and *Xho*I restriction sites, and ligated back into the mammalian expression vector using the same restriction sites. Correct clones were identified using standard procedures and the inserts sequenced for verification.

The plasmids expressing human RAG1 and RAG2 proteins, pEBG-hRAG1 and pEBG-hRAG2, were a generous gift of Dr Patricia Q Cortes, Mount Sinai Medical School, New York.⁴⁹

Extrachromosomal recombination assay protocol. Analysis of functional activity of wild-type hTdT and the hTdT variants was carried out in human embryonic kidney cells HEK293T, which lack endogenous TdT expression. The cells were originally obtained from ATCC (Manassas, VA, USA) in the early 2000s, continuously passaged and maintained in liquid nitrogen and tested for mycoplasma contamination routinely. They were passaged for a week before transfection in Dulbecco's modified Eagle's medium/high glucose media (Sigma-Aldrich) supplemented with 10% fetal bovine serum

at 37 °C and 5% CO₂. Cells were co-transfected at 80–90% confluency with substrate plasmid pGG51, one of the hTdT wild-type or variant pcDNA 4/myc-His A expression vectors, and pEBG-hRAG1 and pEBG-hRAG2, the vectors expressing human RAG1 and RAG2 proteins. Cells were transfected via Poly-Jet DNA transfection reagent (FroggaBio). In all, 2.5 μ g DNA per 60 mm culture dish was transfected at 1:3 DNA: Poly-Jet ratio. The transfected cells were incubated at 37 °C and 5% CO₂ for 6 h. Fresh culture media supplemented with caffeine for a final concentration of 1 mM was added 6-h post-transfection.^{50,51}

The assay was carried out as previously described (Figure 4a).^{36,37} Briefly, substrate plasmid DNA was harvested 48-h post-transfection via rapid alkaline lysis. Plasmid DNA was *Dpn*I-digested to select for plasmids that had replicated during the incubation in HEK293T cells. The recovered DNA was transformed into *E. coli* ElectroMAX DH10B cells via electroporation. Bacteria harboring replicated substrate plasmids were recovered from LB agar plates containing 100 μ g ml⁻¹ ampicillin, and bacteria with recombinant substrate plasmids from LB agar plates containing 33 μ g ml⁻¹ chloramphenicol and 100 μ g ml⁻¹ ampicillin. Recombination frequency (R-value) was calculated as the ratio of the recombined substrate plasmids (double-resistant colonies) to the total number of replicated substrate plasmid (ampicillin-resistant colonies).

Colony PCR was performed on double-resistant colonies (pGG51-For 5'-TTGTCGATGAATTCCTGT-3', pGG51-Rev 5'-ATCGAAGGAATCGAGGA CTT-3') and the coding joints of confirmed recombinant plasmids sequenced using pGG51-Forward primer and analyzed (Supplementary Table 2).

Coding joint analysis. The following parameters were analyzed: number of nucleotide deletions and additions (palindromic and N-additions), frequency of nucleotide additions and G/C addition frequency. Repeated recombinant pGG51 substrate sequences within the same mammalian transfection were excluded from the analysis because of the possibility that they were products of plasmid replication in mammalian cells after recombination.

Statistical analysis

Statistical analysis was performed using SPSS Statistics (IBM, Armonk, NY, USA). Chi-square and the Standard T tests were used. A *P*-value < 0.05 was considered statistically significant.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

These studies were supported by the Arthritis Society (grant number 560609) to GEW and by the Canadian Institutes of Health Research (CIHR) (grant number 106583) to VS and YS. AT is the recipient of a CIHR Graduate Student Scholarship. GEW is the recipient of a York University Faculty Association Award held at the University of Canterbury, New Zealand. The substrate plasmid, pGG51, was a generous gift of Dr Michael Lieber, University of Southern California, Los Angeles. The vectors pEBG-hRAG1 and pEBG-hRAG2 were a generous gift of Dr Patricia Q Cortes, Mount Sinai Medical School, New York.

REFERENCES

- Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K *et al*. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 2014; **506**: 376–381.
- Mabert K, Cojoc M, Peitzsch C, Kurth I, Souchelnytskyi S, Dubrovskaya A. Cancer biomarker discovery: current status and future perspectives. *Int J Radiat Biol* 2014; **90**: 659–677.
- Danchin E, Vitiello V, Vienne A, Richard O, Gouret P, McDermott MF *et al*. The major histocompatibility complex origin. *Immunol Rev* 2004; **198**: 216–232.
- Trowsdale J, Knight JC. Major histocompatibility complex genomics and human disease. *Annu Rev Genomics Hum Genet* 2013; **14**: 301–323.
- Gozalbo-Lopez B, Andrade P, Terrados G, de Andres B, Serrano N, Cortegano I *et al*. A role for DNA polymerase μ in the emerging DJH rearrangements of the postgastrulation mouse embryo. *Mol Cell Biol* 2009; **29**: 1266–1275.
- Lewis SM, Wu GE. The origins of V(D)J recombination. *Cell* 1997; **88**: 159–162.
- Lewis SM. The mechanism of V(D)J joining: lessons from molecular, immunological, and comparative analyses. *Adv Immunol* 1994; **56**: 27–150.

- 8 Repasky JA, Corbett E, Boboila C, Schatz DG. Mutational analysis of terminal deoxynucleotidyltransferase-mediated N-nucleotide addition in V(D)J recombination. *J Immunol* 2004; **172**: 5478–5488.
- 9 Gellert M. V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu Rev Biochem* 2002; **71**: 101–132.
- 10 Schatz DG, Spanopoulou E. Biochemistry of V(D)J recombination. *Curr Top Microbiol Immunol* 2005; **290**: 49–85.
- 11 Lewis SM, Wu GE. The old and the restless. *J Exp Med* 2000; **191**: 1631–1636.
- 12 Swanson PC. A RAG-1/RAG-2 tetramer supports 12/23-regulated synapsis, cleavage, and transposition of V(D)J recombination signals. *Mol Cell Biol* 2002; **22**: 7790–7801.
- 13 Grundy GJ, Ramon-Maiques S, Dimitriadis EK, Kotova S, Biertumpfel C, Heymann JB *et al*. Initial stages of V(D)J recombination: the organization of RAG1/2 and RSS DNA in the postcleavage complex. *Mol Cell* 2009; **35**: 217–227.
- 14 Fowler JD, Suo Z. Biochemical, structural, and physiological characterization of terminal deoxynucleotidyl transferase. *Chem Rev* 2006; **106**: 2092–2110.
- 15 Komori T, Okada A, Stewart V, Alt FW. Lack of N regions in antigen receptor variable region genes of TdT-deficient lymphocytes. *Science* 1993; **261**: 1171–1175.
- 16 Gilfillan S, Dierich A, Lemeur M, Benoist C, Mathis D. Mice lacking TdT: mature animals with an immature lymphocyte repertoire. *Science* 1993; **261**: 1175–1178.
- 17 Boubakour-Azzouz I, Bertrand P, Claes A, Lopez BS, Rougeon F. Terminal deoxynucleotidyl transferase requires KU80 and XRCC4 to promote N-addition at non-V(D)J chromosomal breaks in non-lymphoid cells. *Nucleic Acids Res* 2012; **40**: 8381–8391.
- 18 Benedict CL, Gilfillan S, Thai TH, Kearney JF. Terminal deoxynucleotidyl transferase and repertoire development. *Immunol Rev* 2000; **175**: 150–157.
- 19 Motea EA, Berdis AJ. Terminal deoxynucleotidyl transferase: the story of a mis-guided DNA polymerase. *Biochimica et Biophysica Acta* 2010; **1804**: 1151–1166.
- 20 Delarue M, Boule JB, Lescar J, Expert-Bezancon N, Jourdan N, Sukumar N *et al*. Crystal structures of a template-independent DNA polymerase: murine terminal deoxynucleotidyltransferase. *EMBO J* 2002; **21**: 427–439.
- 21 Fetter M, Zee DS, Koenig E, Dichgans J. Head-shaking nystagmus during vestibular compensation in humans and rhesus monkeys. *Acta Oto-laryngologica* 1990; **110**: 175–181.
- 22 Gouge J, Rosario S, Romain F, Beguin P, Delarue M. Structures of intermediates along the catalytic cycle of terminal deoxynucleotidyltransferase: dynamical aspects of the two-metal ion mechanism. *J Mol Biol* 2013; **425**: 4334–4352.
- 23 Yamtich J, Sweasy JB. DNA polymerase family X: function, structure, and cellular roles. *Biochimica et Biophysica Acta* 2010; **1804**: 1136–1150.
- 24 Thai TH, Purugganan MM, Roth DB, Kearney JF. Distinct and opposite diversifying activities of terminal transferase splice variants. *Nat Immunol* 2002; **3**: 457–462.
- 25 Thai TH, Kearney JF. Isoforms of terminal deoxynucleotidyltransferase: developmental aspects and function. *Adv Immunol* 2005; **86**: 113–136.
- 26 Thai TH, Kearney JF. Distinct and opposite activities of human terminal deoxynucleotidyltransferase splice variants. *J Immunol* 2004; **173**: 4009–4019.
- 27 Di Santo R, Maga G. Human terminal deoxynucleotidyl transferases as novel targets for anticancer chemotherapy. *Curr Med Chem* 2006; **13**: 2353–2368.
- 28 Locatelli GA, Di Santo R, Crespan E, Costi R, Roux A, Hubscher U *et al*. Diketo hexenoic acid derivatives are novel selective non-nucleoside inhibitors of mammalian terminal deoxynucleotidyl transferases, with potent cytotoxic effect against leukemic cells. *Mol Pharmacol* 2005; **68**: 538–550.
- 29 Kodama EN, McCaffrey RP, Yusa K, Mitsuya H. Antileukemic activity and mechanism of action of cordycepin against terminal deoxynucleotidyl transferase-positive (TdT+) leukemic cells. *Biochem Pharmacol* 2000; **59**: 273–281.
- 30 Conde C, Weller S, Gilfillan S, Marcellin L, Martin T, Pasquali JL. Terminal deoxynucleotidyl transferase deficiency reduces the incidence of autoimmune nephritis in (New Zealand Black x New Zealand White)F1 mice. *J Immunol* 1998; **161**: 7023–7030.
- 31 Feeney AJ, Lawson BR, Kono DH, Theofilopoulos AN. Terminal deoxynucleotidyl transferase deficiency decreases autoimmune disease in MRL-Fas(lpr) mice. *J Immunol* 2001; **167**: 3486–3493.
- 32 Marshall AJ, Doyen N, Bentolila LA, Paige CJ, Wu GE. Terminal deoxynucleotidyl transferase expression during neonatal life alters D(H) reading frame usage and Ig-receptor-dependent selection of V regions. *J Immunol* 1998; **161**: 6657–6663.
- 33 Gavin MA, Bevan MJ. Increased peptide promiscuity provides a rationale for the lack of N regions in the neonatal T cell repertoire. *Immunity* 1995; **3**: 793–800.
- 34 Schelonka RL, Ivanov II, Vale AM, Szymanska E, Zemlin M, Gartland GL *et al*. The CDR-H3 repertoire from TdT-deficient adult bone marrow is a close, but not exact, homologue of the CDR-H3 repertoire from perinatal liver. *J Immunol* 2010; **185**: 6075–6084.
- 35 Robey IF, Peterson M, Horwitz MS, Kono DH, Stratmann T, Theofilopoulos AN *et al*. Terminal deoxynucleotidyltransferase deficiency decreases autoimmune disease in diabetes-prone nonobese diabetic mice and lupus-prone MRL-Fas(lpr) mice. *J Immunol* 2004; **172**: 4624–4629.
- 36 Hesse JE, Lieber MR, Gellert M, Mizuuchi K. Extrachromosomal DNA substrates in pre-B cells undergo inversion or deletion at immunoglobulin V-(D)-J joining signals. *Cell* 1987; **49**: 775–783.
- 37 Lieber MR, Hesse JE, Mizuuchi K, Gellert M. Studies of V(D)J recombination with extrachromosomal substrates. *Curr Top Microbiol Immunol* 1988; **137**: 94–99.
- 38 Lieber MR, Hesse JE, Mizuuchi K, Gellert M. Developmental stage specificity of the lymphoid V(D)J recombination activity. *Genes Dev* 1987; **1**: 751–761.
- 39 Raghavan SC, Hsieh CL, Lieber MR. Both V(D)J coding ends but neither signal end can recombine at the bcl-2 major breakpoint region, and the rejoining is ligase IV dependent. *Mol Cell Biol* 2005; **25**: 6475–6484.
- 40 Gauss GH, Lieber MR. Mechanistic constraints on diversity in human V(D)J recombination. *Mol Cell Biol* 1996; **16**: 258–269.
- 41 Bertocci B, De Smet A, Weill JC, Reynaud CA. Nonoverlapping functions of DNA polymerases mu, lambda, and terminal deoxynucleotidyltransferase during immunoglobulin V(D)J recombination *in vivo*. *Immunity* 2006; **25**: 31–41.
- 42 Liu Q, Chen H, Ojode T, Gao X, Anaya-O'Brien S, Turner NA *et al*. WHIM syndrome caused by a single amino acid substitution in the carboxy-tail of chemokine receptor CXCR4. *Blood* 2012; **120**: 181–189.
- 43 Allison AC. Genetic control of resistance to human malaria. *Curr Opin Immunol* 2009; **21**: 499–505.
- 44 Romain F, Barbosa I, Gouge J, Rougeon F, Delarue M. Conferring a template-dependent polymerase activity to terminal deoxynucleotidyltransferase by mutations in the Loop1 region. *Nucleic Acids Res* 2009; **37**: 4642–4656.
- 45 Purugganan MM, Shah S, Kearney JF, Roth DB. Ku80 is required for addition of N nucleotides to V(D)J recombination junctions by terminal deoxynucleotidyl transferase. *Nucleic Acids Res* 2001; **29**: 1638–1646.
- 46 Nick McElhinny SA, Havener JM, Garcia-Diaz M, Juarez R, Bebenek K, Kee BL *et al*. A gradient of template dependence defines distinct biological roles for family X polymerases in nonhomologous end joining. *Mol Cell* 2005; **19**: 357–366.
- 47 Molano ID, Redmond S, Sekine H, Zhang XK, Reilly C, Hutchison F *et al*. Effect of genetic deficiency of terminal deoxynucleotidyl transferase on autoantibody production and renal disease in MRL/lpr mice. *Clin Immunol* 2003; **107**: 186–197.
- 48 Gauss GH, Lieber MR. Unequal signal and coding joint formation in human V(D)J recombination. *Mol Cell Biol* 1993; **13**: 3900–3906.
- 49 Spanopoulou E, Cortes P, Shih C, Huang CM, Silver DP, Svec P *et al*. Localization, interaction, and RNA binding properties of the V(D)J recombination-activating proteins RAG1 and RAG2. *Immunity* 1995; **3**: 715–726.
- 50 Yuan SW, Agard EA, Larijani M, Wu GE. Coding joint diversity in mature and immature B-cell lines. *Scand J Immunol* 2005; **62**: 114–118.
- 51 Menetski JP, Gellert M. V(D)J recombination activity in lymphoid cell lines is increased by agents that elevate cAMP. *Proc Natl Acad Sci USA* 1990; **87**: 9324–9328.

Supplementary Information accompanies this paper on Genes and Immunity website (<http://www.nature.com/gene>)