



OPEN

## Spectrum of mutational signatures in T-cell lymphoma reveals a key role for UV radiation in cutaneous T-cell lymphoma

Christine L. Jones<sup>1</sup>, Andrea Degasperi<sup>2,3</sup>, Vieri Grandi<sup>1</sup>, Tauanne D. Amarante<sup>2,3</sup>, Genomics England Research Consortium\*, Tracey J. Mitchell<sup>1</sup>, Serena Nik-Zainal<sup>2,3</sup> & Sean J. Whittaker<sup>1</sup>✉

T-cell non-Hodgkin's lymphomas develop following transformation of tissue resident T-cells. We performed a meta-analysis of whole exome sequencing data from 403 patients with eight subtypes of T-cell non-Hodgkin's lymphoma to identify mutational signatures and associated recurrent gene mutations. Signature 1, indicative of age-related deamination, was prevalent across all T-cell lymphomas, reflecting the derivation of these malignancies from memory T-cells. Adult T-cell leukemia-lymphoma was specifically associated with signature 17, which was found to correlate with the IRF4 K59R mutation that is exclusive to Adult T-cell leukemia-lymphoma. Signature 7, implicating UV exposure was uniquely identified in cutaneous T-cell lymphoma (CTCL), contributing 52% of the mutational burden in mycosis fungoides and 23% in Sezary syndrome. Importantly this UV signature was observed in CD4+ T-cells isolated from the blood of Sezary syndrome patients suggesting extensive re-circulation of these T-cells through skin and blood. Analysis of non-Hodgkin's T-cell lymphoma cases submitted to the national 100,000 WGS project confirmed that signature 7 was only identified in CTCL strongly implicating UV radiation in the pathogenesis of cutaneous T-cell lymphoma.

Non-Hodgkin's lymphomas (NHL) consist of distinct clinico-pathologic and molecular entities arising from nodal or extra-nodal sites<sup>1</sup>. T-cell lymphomas often develop in extra-nodal sites, constitute only 10–15% of NHL in the west, although higher rates of 20–25% are reported in Asia, and arise from tissue resident  $\alpha\beta$  or  $\gamma\delta$  T-cells<sup>2</sup>. Whilst the incidence of NHL has increased, the underlying aetiology remains poorly understood.

Recurrent chromosomal translocations are common in B-cell NHL and are due to intrinsic molecular mechanisms including aberrant V(D)J recombination in which different genes are juxtaposed with immunoglobulin promoters, somatic hypermutation and class switch recombination<sup>3</sup>. For mature T-cell NHL, the prevalence of recurrent translocations is less, with notable exceptions such as the t(2;5) in ALK+ anaplastic large cell lymphomas<sup>4</sup>, reflecting the lack of somatic hypermutation and class switch recombination in healthy mature T-cells. In contrast the identification of extrinsic factors has proved elusive, with the exception of viral transformation such as Adult T-cell leukemia-lymphoma (ATLL) associated with HTLV-1<sup>5</sup> and EBV associated NHL<sup>6</sup>.

A combination of intrinsic and extrinsic mutational processes cause a characteristic mutational spectrum. Consequently when a cell is clonally expanded through malignancy, these patterns can be detected within the tumour. The patterns of mutation are defined by assessing the six substitution types which can occur (C>A, C>G, C>T, T>A, T>C, and T>G) and considering the 5' and 3' nucleotide context giving 96 different trinucleotide mutation types<sup>7</sup>. An analysis of over 10,000 exomes and 1000 genomes across 40 types of cancer in The Cancer Genome Atlas (TCGA) identified 30 distinct signatures, some of which can be attributed to known mutagens (COSMIC mutational signatures v2, available at [https://cancer.sanger.ac.uk/cosmic/signatures\\_v2](https://cancer.sanger.ac.uk/cosmic/signatures_v2)).

<sup>1</sup>St. John's Institute of Dermatology, School of Basic and Medical Biosciences, King's College London, Guy's Hospital, London SE1 9RT, UK. <sup>2</sup>MRC Cancer Unit, Hutchison/MRC Research Centre, University of Cambridge, Box 197, Cambridge Biomedical Campus, Cambridge CB2 0XZ, UK. <sup>3</sup>Academic Laboratory of Medical Genetics, Lv 6 Addenbrooke's Treatment Centre, Addenbrooke's Hospital, Box 238, Cambridge CB2 0QQ, UK. \*A list of authors and their affiliations appears at the end of the paper. ✉email: sean.whittaker@kcl.ac.uk

	AITL	ATLL	EATL	HSTL	MF	NKTCL	PTCL	SS	
Choi 2015 <sup>16</sup>	–	–	–	–	–	–	–	33	33
da Silva Almeida 2015 <sup>17</sup>	–	–	–	–	8	–	–	26	34
Jiang 2015 <sup>12</sup>	–	–	–	–	–	25	–	–	25
Kataoka 2015 <sup>11</sup>	–	81	–	–	–	–	–	–	81
McKinney 2017 <sup>13</sup>	–	–	–	64	–	–	–	–	64
Moffitt 2017 <sup>14</sup>	–	–	58	–	–	–	–	–	58
Palomero 2014 <sup>18</sup>	3	–	1	–	–	2	6	–	12
Prasad 2016 <sup>19</sup>	–	–	–	–	–	–	–	12	12
Roberti 2016 <sup>20</sup>	–	–	15	–	–	–	–	–	15
Sakata–Yanagimoto 2014 <sup>21</sup>	3	–	–	–	–	–	3	–	6
Ungewickell 2015 <sup>22</sup>	–	–	–	–	6	–	–	5	11
Wang 2015 <sup>23</sup>	–	–	–	–	–	–	–	37	37
Woollard 2016 <sup>24</sup>	–	–	–	–	–	–	–	10	10
Yoo 2014 <sup>25</sup>	5	–	–	–	–	–	–	–	5
	11	81	74	64	14	27	9	123	

**Table 1.** Details of the studies included and the number of samples with each disease subtype.

Previous studies showed that mutational signatures in B-cell NHL including B-CLL are often associated with intrinsic abnormalities such as the COSMIC signature 9 due to somatic hypermutation associated with aberrant activation induced cytosine deaminase<sup>8</sup>. Furthermore, single cell whole genome sequencing studies of B-cells from healthy volunteers suggests that mutational signatures correlate with intrinsic factors including the age associated COSMIC signature 1 due to spontaneous deamination of methylated cytosines as well as COSMIC signature 9<sup>9</sup>. As T-cell lymphomas are not included in TCGA, there have been no comprehensive studies of mutational signatures in T-cell NHL although recent high-throughput sequencing studies have identified putative driver gene mutations targeting specific signalling pathways notably TCR, NF- $\kappa$ B and JAK-STAT signalling<sup>10–14</sup>. These findings suggest that some T-cell NHL are dependent on TCR signalling analogous to a dependence on BCR signalling in B-cell NHL<sup>15</sup>.

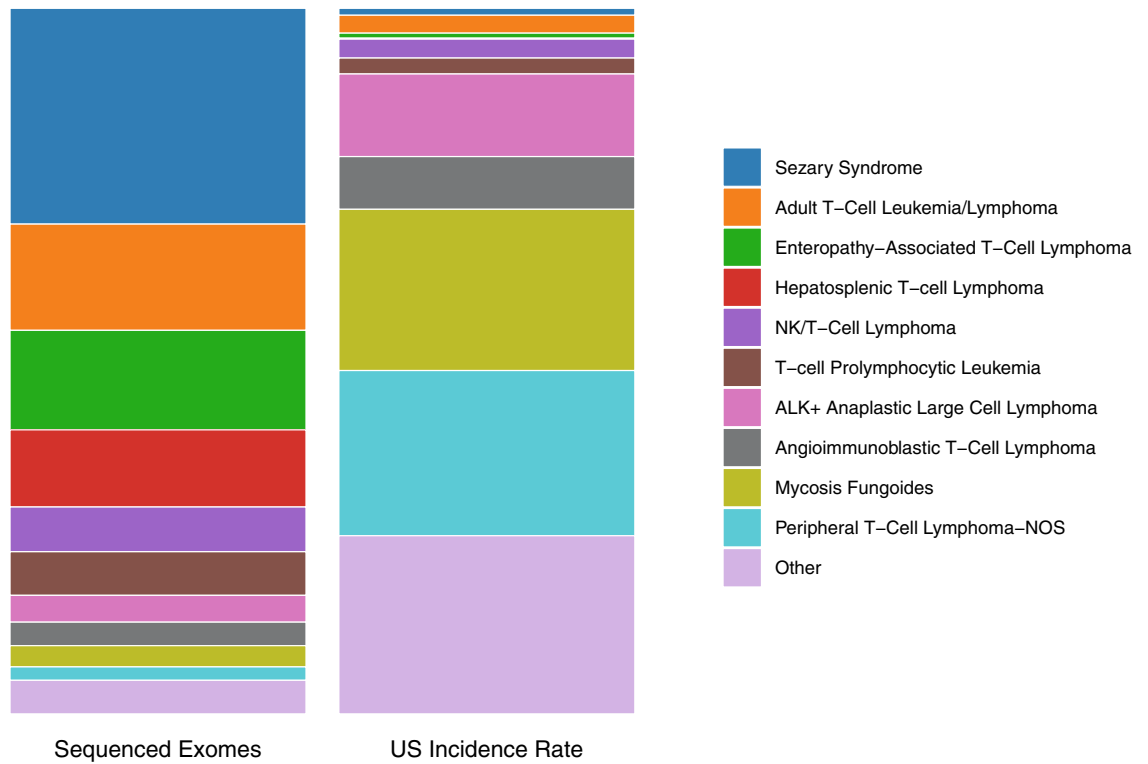
In order to identify causal factors in T-cell NHL, we performed a systematic review of the literature and identified 14 published WES studies of T-cell NHLs from which mutational data could be combined<sup>11–14,16–25</sup>. Using this dataset of over 400 whole exome sequences we deconvoluted the mutational signatures present in eight different subtypes of mature T-cell lymphomas. Whilst this analysis indicates that all mature T-cell NHLs have age related signatures, distinct signatures are present, including signature 17 associated with HTLV-1 transformed ATLL and UV signature 7 in cutaneous T-cell lymphoma (CTCL). We show that this clonotypic UV signature is present in leukemic T-cells, presumably due to recirculation from skin. These findings have important implications for the pathogenesis of CTCL and suggest that skin resident memory T-cells are susceptible to the mutagenic as well as the immunomodulatory effects of UV.

## Results

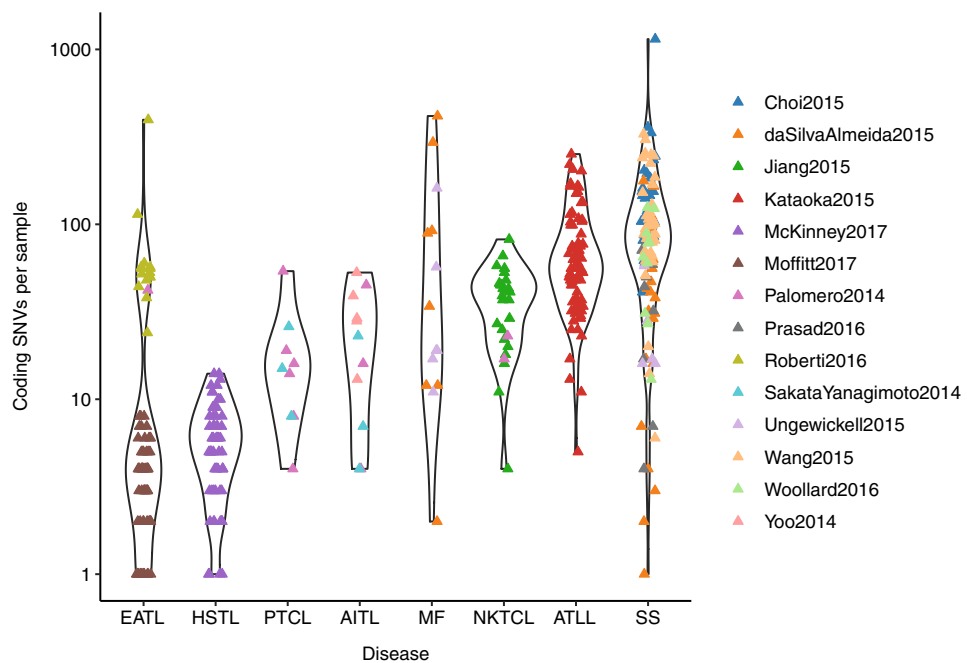
**T-cell lymphomas have a variable mutational load.** Since mature T-cell NHLs have not been included in any of the large pan-cancer sequencing projects, we performed a comprehensive review of the literature. 34 studies were identified which included whole exome data from 631 patients. Overall 14 studies comprising 403 patients met our eligibility criteria (Table 1).

There was a striking disparity in the number of exomes sequenced between the subtypes, particularly when compared to the relative incidence of each subtype (Fig. 1). Notably, rare subtypes such as Sezary syndrome (SS) and ATLL account for 45% of the exomes sequenced whilst only representing 4% of mature T-cell NHL cases in the US<sup>26</sup>. On the other hand, mycosis fungoides (MF) and peripheral T-cell lymphoma (PTCL) account for only 5% of the exomes sequenced whilst representing 45% of mature T-cell NHL cases. This is likely because both SS and ATLL are leukaemic variants facilitating isolation of clonal populations of tumour cells from peripheral blood.

Examining the distribution of mutational burden within and between subtypes of T-cell lymphoma revealed that SS has both the highest median number of coding SNVs per sample and the widest range (Fig. 2), supporting previous observations of heterogeneity in SS<sup>10</sup>. ATLL, which is driven by HTLV-1, has a high mutation rate with a much narrower distribution. Enteropathy-associated T-cell lymphoma (EATL) and hepatosplenic T-cell lymphoma (HSTL), which often arise from  $\gamma\delta$  T-cells, have the lowest mutation rate, although it is notable that the EATL samples have a bimodal distribution. This split corresponds to the study from Roberti et al.<sup>20</sup> averaging 77 SNVs and Moffitt et al.<sup>14</sup> averaging 2.5 per sample. The one sample from Palomero et al.<sup>18</sup> sits with the higher group at 42 suggesting that the filtering criteria used by Moffitt *et al.* may have been more stringent. This highlights that, because our analysis was based on published data from multiple sources, differences in sequencing depth and filtering strategy could contribute to variability in results for other subtypes as well. However, samples from multiple SS studies were analysed and showed comparable mutational burden suggesting that technical differences did not contribute significantly to variation in SS results.

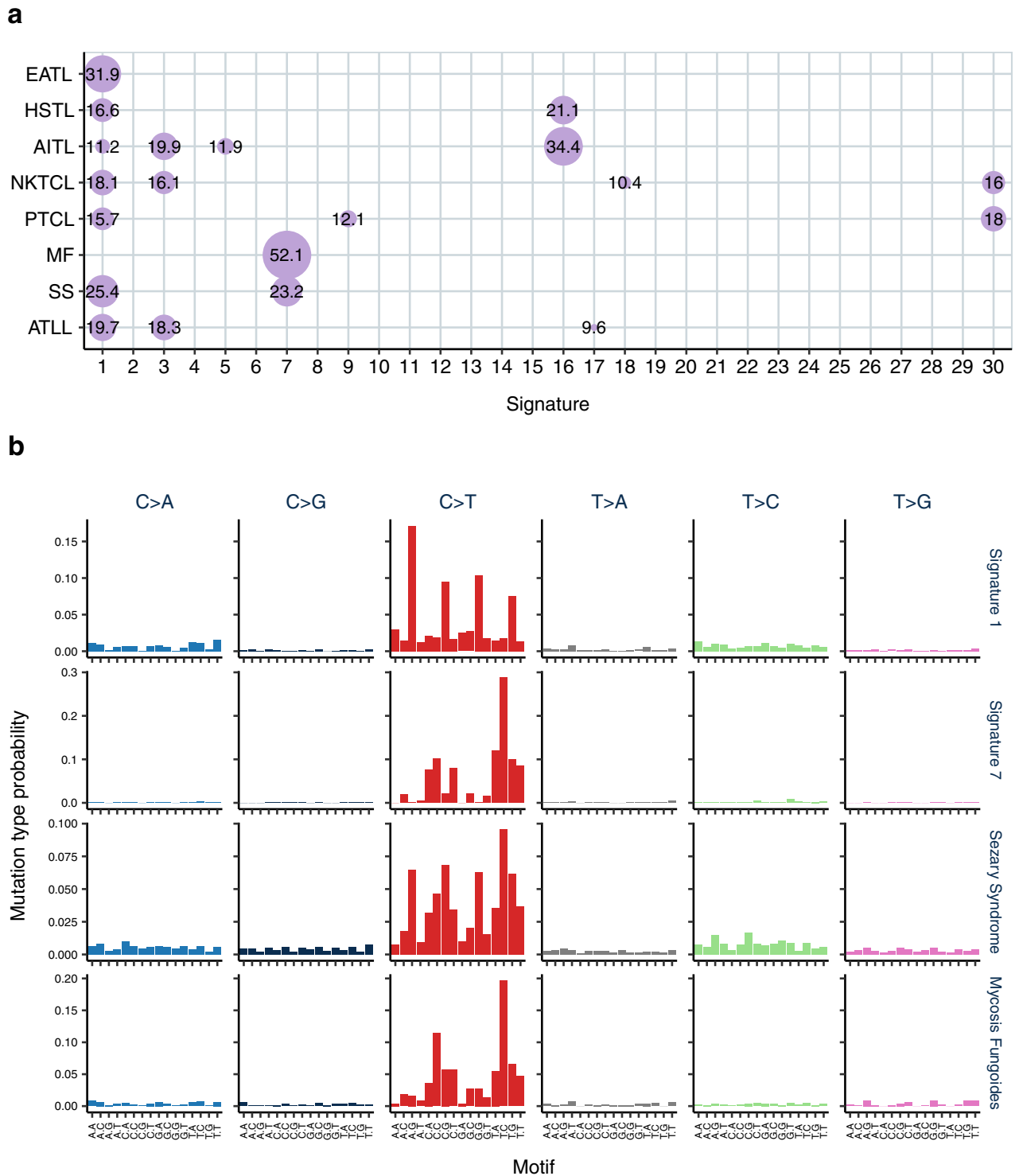


**Figure 1.** Comparison of the distribution of number of exomes sequenced with the US incidence rate of each subtype. Literature analysis identified 631 whole exome sequences performed on mature T-cell NHL samples. US Incidence rates are derived from the 2016 US lymphoid malignancy statistics<sup>26</sup> which estimated 8,380 new cases of mature T-cell NHL in 2016.



**Figure 2.** Number of coding SNVs per sample across disease subtypes. Data were collated for those studies where full mutation lists were published and for subtypes where sequencing had been performed on at least 10 patients.





**Figure 4.** Mutational signatures identified in T-cell lymphoma. **(a)** matrix showing the percentage of mutations in each disease subtype assigned to the 30 COSMIC signatures. **(b)** plots showing the likelihood of each mutation in each trinucleotide context for Signatures 1 and 7 as per COSMIC mutational signatures v2 and the observed SNV catalogues for Sezary syndrome and mycosis fungoides.

PUVA), however this information was provided in summary form and so it was not possible to allocate treatments to samples from specific individuals.

To validate this finding we examined data generated by the UK 100,000 genomes project. Six newly diagnosed and untreated CTCL samples were included and their genomes were found to have a mutational burden

of 3–8.4 somatic coding variants per Mb to which signature 7 contributed between 25 and 61%. Two cases of systemic nodal T-cell lymphoma were also sequenced revealing mutational burdens of 0.3 and 3.5 somatic coding variants per Mb with neither showing any contribution from signature 7. Notably, significant bias towards the untranscribed strand was seen for C>T mutations in cutaneous but not nodal T-cell lymphomas with frequent double nucleotide substitutions CC>TT (Fig. 5) both of which are characteristic features of UV induced mutation.

A significant contribution was noted in ATLL from signature 17. Although the aetiology of signature 17 is unknown, this signature is associated with a high mutational load and neoantigen burden subtype of oesophageal adenocarcinoma<sup>30</sup> and is enriched in metastatic tissue of colorectal cancer<sup>31</sup>. ATLL patients have a variable course consisting of indolent (smouldering and chronic) and aggressive (lymphoma and acute) subtypes. Hence, we examined whether there was an association between aggressive subtypes and signature 17. Only 1/13 (8%) indolent ATLL exhibited signature 17, whilst 9/32 (28%) aggressive ATLL exhibited signature 17 but this difference did not reach significance in a Fisher's exact test.

**Significant associations between mutational signatures and driver mutations.** We used the method of Temko et. al.<sup>32</sup> to assess the potential causative role of mutational signatures in relation to recurrent driver gene mutations. Causal peaks were used to estimate the contribution of each signature and compared between wild type and mutant groups for each recurrent driver gene mutation which occurred in five or more patients. Using the Bonferroni corrected p-value of 0.00029, two significant associations were observed, between STAT3 Y640F and signature 5 (Fig. 6) and between IRF4 K59R and signature 17 (Fig. 7). Whilst the STAT3 Y640F mutation was distributed across T-cell NHL cases, specifically 7 ATLL, 1 NK/T-cell lymphoma (NKTCL) and 1 MF, the IRF4 K59R mutation was only seen in ATLL patients, consistent with the restriction of signature 17 to these patients. For both of these associations the underlying nucleotide change was most likely to have been generated by the signature observed. No further associations were identified by examining the data on a disease subtype level.

Due to the dominant contributions of C>T peaks in signatures 1, 7 and 30, the only context in which we could clearly define a causal association for signature 7 was for C>T in a TCC context. Of the recurrent driver mutations, only PLCG1 S345F occurs in this context and there was no significant association between this recurrent mutation and signature 7. Additionally this mutation was not restricted to cutaneous lymphoma, being observed in 3 SS, 3 ATLL and 1 AITL patients suggesting it originated from another mutational process. This method of analysis has a much greater power over the whole cohort and within individual disease subtypes with highly recurrent mutations, and so the lack of significant association of specific gene variants with signature 7 causal peaks likely reflects the absence of highly recurrent mutations in CTCL.

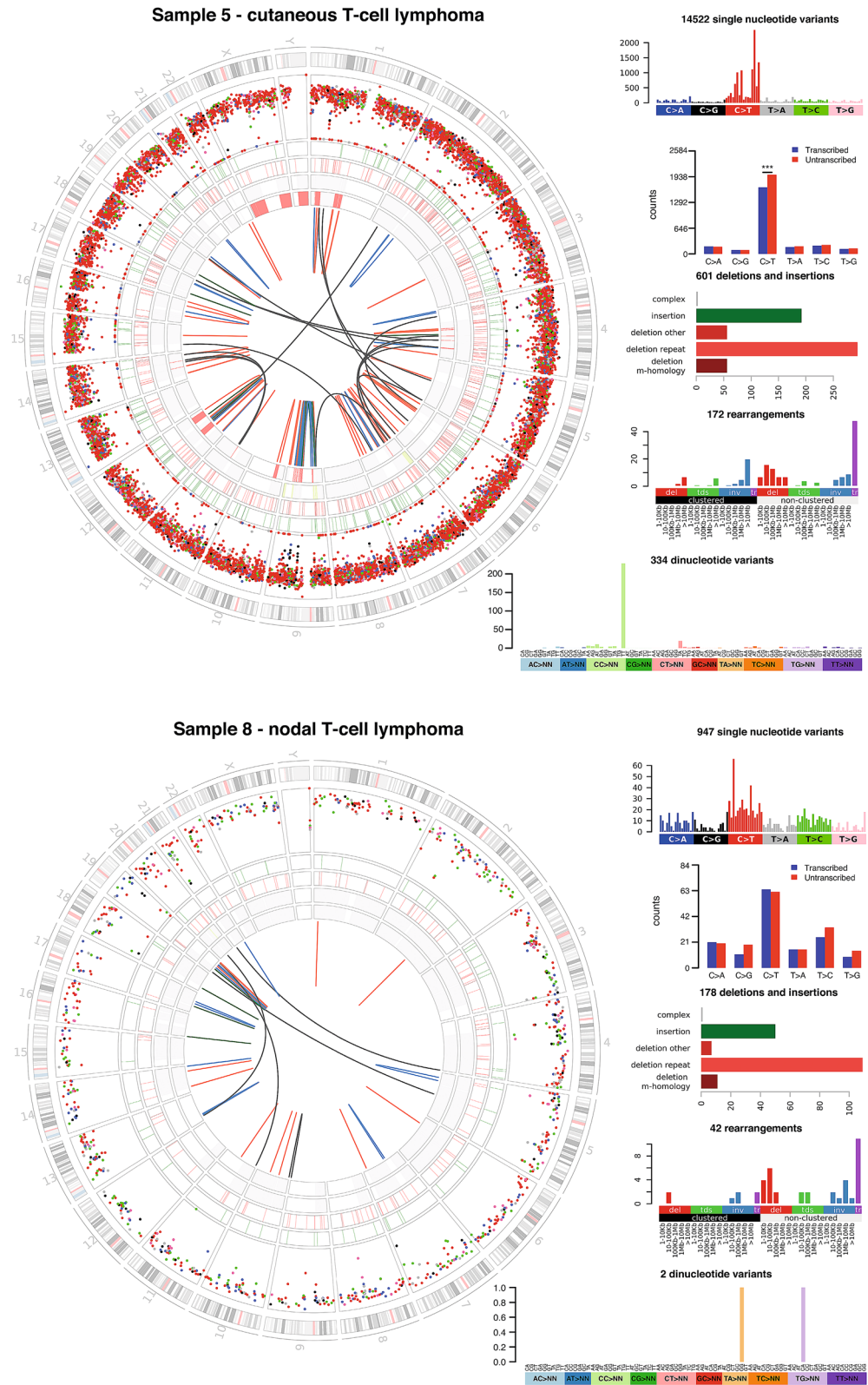
**UV mutational signatures in healthy T-cells.** If the presence of the UV signature in CTCL reflects prior exposure of the originating T-cell to UV whilst circulating in the skin, we might expect a proportion of healthy peripheral blood memory T-cells to show similar UV induced mutations. In the absence of a clonal proliferation, this can only be identified by looking at the mutational spectrum of single cells. No single cell whole genome or exome sequencing has been performed on healthy T-cells but we did identify a dataset of RNA-seq from CD4+ T-cells where the full length transcript has been sequenced which allows mutation calling<sup>33</sup>. We analysed three subjects where both T<sub>EM</sub> and T<sub>CM</sub> single cells had been sequenced. Data processing and mutation calling were performed as described in the supplementary materials. No clear UV signature was observed in 94 cells from Donor 3, 91 cells from Donor 6 or 90 cells from Donor 12, and in each donor a similar mutational burden was observed between T<sub>EM</sub> and T<sub>CM</sub> cells. It is estimated that almost twice as many T-cells are permanently resident in the skin than circulating in the blood<sup>34</sup> but the proportion of T-cells which freely re-circulate through skin and blood is not well characterised. Hence we may need to analyse a much greater number of cells to observe those exposed to UV.

## Discussion

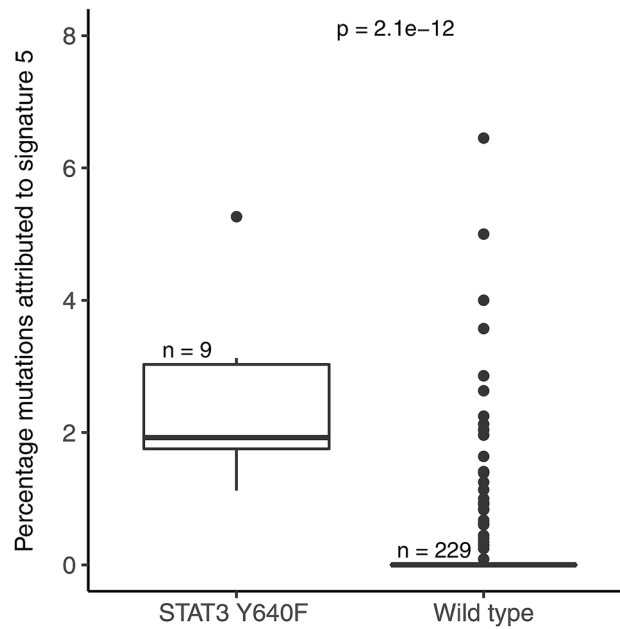
There are several findings in this study which provide insight into the pathogenesis of mature T-cell NHL and have immunobiologic and therapeutic relevance. Strikingly cutaneous T-cell lymphomas, specifically SS and MF variants, show an obvious clonotypic UV signature 7 associated with a high tumour mutational burden as expected for malignancies linked to a specific mutagen. This UV signature contributed 23% and 52% of the mutational load for the SS and MF cohorts respectively. Specifically the UV signature was identified in peripheral blood T-cells from 74% (67/90) of SS patients representing between 7.5 and 88% of the overall SNV mutational burden. Not surprisingly signature 1, associated with age related deamination of methylated cytosines, was detected in all T-cell NHL with the exception of MF in which signature 1 is likely masked by the large contribution of the UV signature. Two mutational signatures were associated with the HTLV-1 associated ATLL cohort namely signature 1, and signature 17 which is of unknown cause but associated with gastric cancer<sup>35</sup> and oesophageal adenocarcinoma<sup>30</sup>. Our computational power to deconvolute mutational signatures reflects the total number of tumours compared to the mutational load<sup>36</sup> and therefore the smaller sample size and low mutational burden of the other T-cell NHL cohorts limits an accurate interpretation. Future whole genome studies, particularly of non-cutaneous T-cell NHL subtypes, could yield additional insight into mutational processes as greater coverage would allow increased power for signature extraction.

MF is derived from skin resident effector memory T-cells and, although previous studies suggested that leukaemic T-cells from SS patients were derived from central memory T-cells<sup>37</sup>, the presence of a UV signature in DNA derived from peripheral blood CD4+ T-cells of SS patients supports data that T<sub>CM</sub> exhibit marked plasticity contributing to circulation from skin through peripheral blood<sup>38</sup>. Indeed, studies have shown that low dose solar simulated irradiation of whole skin causes alterations of gene expression in peripheral blood derived T-cells<sup>39</sup>

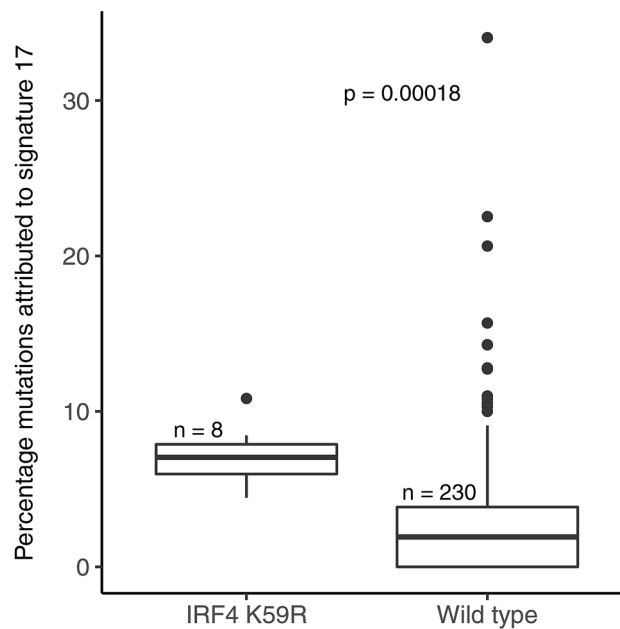




**Figure 5.** Characteristic features of UV induced mutation are observed only in patients with cutaneous T-cell lymphoma. Whole genome sequencing data from the UK 100,000 genomes project comparing representative cutaneous and nodal T-cell lymphoma patients. The highlighted track on the circos plots is a rainfall diagram of SNVs where distance from the centre is proportional to the logarithm of the distance between adjacent SNVs. Cutaneous T-cell lymphoma patients have a characteristic inner ring of C>T variants due to CC>TT substitutions, which are also clear in the light green bar on the dinucleotide variants plot. Single nucleotide variants occurring within gene bodies were assessed to determine whether they affect the transcriptional or non-transcriptional strand. Significant transcriptional strand bias was observed for C>T mutations in cutaneous T-cell lymphoma but not nodal T-cell lymphoma indicating transcription-coupled nucleotide excision repair, a feature of UV induced mutation.



**Figure 6.** Patients with the STAT3 Y640F mutation showed significantly greater contribution of signature 5 to their mutational catalogue. Wilcoxon rank sum test was used to compare the contribution of each signature between the groups mutant or WT for a given mutation.



**Figure 7.** Patients with the IRF4 K59R mutation showed significantly greater contribution of signature 17 to their mutational catalogue. Wilcoxon rank sum test was used to compare the contribution of each signature between the groups mutant or WT for a given mutation.

suggesting that UVR may have a direct effect on the blood transcriptome of T-cells circulating through the skin. Our previous study of MF suggested that p53 gene mutations were characteristic of UV damage<sup>40</sup> and several whole exome/genome datasets including our own have suggested a possible UV influence<sup>23,24,41</sup>. Importantly, exposure of cultured cells to simulated solar radiation followed by sequencing of clones generates a mutational profile which replicates that of signature 7<sup>42</sup> providing molecular evidence for the link between UV and signature 7. Combining data from multiple whole exome T-cell NHL studies has allowed us to confirm that this UV signature is associated with and restricted to CTCL.

Overall 41% of SS patients were treatment naïve and a UV signature was detected in a high proportion of these patients. Other patients received various treatments including photopheresis, in which an enriched population



of white blood cells is exposed to psoralens and UVA extra-corporeally and reinfused, and PUVA phototherapy consisting of oral psoralens and UVA irradiation of skin. However the mutational signatures detected are based on the analysis of clonotypic causal peaks. Therefore critically our findings reflect the impact of environmental UV on the original skin resident T-cell prior to clonal expansion rather than subclonal evolution of the malignant T-cell as subsequent UV induced mutations would not be present in all tumour cells and therefore would not be detectable as a clonotypic signature. Furthermore PUVA phototherapy and photopheresis are characterised mostly by UVA induced cross linking of DNA by psoralens to produce pyrimidine mutations in a TpA sequence context. Until recently mutational signatures have been based on single base substitutions, but analysis of large whole genome cancer datasets has now expanded the number of distinct signatures after identification of those associated with double base substitutions and indels<sup>43</sup>. Specifically this study has decomposed the UV signature into four constituent types with two consisting of C>T at TCN and C>T at CCN, whilst the other two consist of T>A at NTT and T>C at NTT sites. In fact our analysis of T-cell NHL cases submitted to the UK national 100,000 WGS project has confirmed a high frequency of CC>TT double substitutions at diprimidine sites with very high mutational burdens confined only to untreated CTCL cases. In addition there was a significant bias towards C>T mutations in the untranscribed strand in the CTCL cases.

Epidemiological case-control studies have suggested that occupational solar UV exposure is a risk factor for NHL in Caucasian populations<sup>44</sup> however most of the studies in this meta-analysis analysed NHL as a cohort and did not divide according to specific lymphoma subtypes. One European case-control study observed a link between MF and occupational sun exposure<sup>45</sup>. In contrast population based registry data from Australia suggested that the incidence of NHL, including cutaneous lymphomas, shows an inverse relationship to latitude, suggesting that UV is protective possibly due to immuno-modulatory effects of UV<sup>46</sup>. A study using US SEER data found no overall trend between UV radiation and cutaneous lymphoma risk but did observe that ethnicity significantly modified the association of UVR and risk<sup>47</sup>. This complex picture highlights the challenge of unravelling the role of UV in NHL. Our data provide molecular evidence for a role of UV in the development of cutaneous T-cell lymphoma only, highlighting the need to examine epidemiological data on a subtype specific basis. Assessment of UV exposure is also challenging with larger population based studies relying on location based estimation whilst personal exposure estimation based on questionnaires increases accuracy but limits the study size.

UV radiation is known to have an immune modulatory effect on the skin by inducing a T<sub>reg</sub> response<sup>48</sup>. What has not been clear is the mutational impact of UV upon skin-resident T-cells. In contrast to other rapidly dividing cells, studies have shown that irradiation of memory CD8+ T-cells ex vivo rapidly induces all DNA damage response pathways and cell cycle checkpoints during the secondary expansion phase and in the absence of repair, CD8+ T-cells are driven to compulsive suicide to prevent transformation<sup>49</sup>. DNA damage repair pathways may therefore play a critical role in preventing UV induced transformation of skin-resident T-cells. Indeed, heterozygote p53+/- mice develop B-cell lymphomas after UV irradiation<sup>50</sup>. Functional studies are now required to investigate the role of UV in the pathogenesis of T-cell lymphomas.

We identified 20 genes harbouring 34 recurrent variants across different subtypes of T-cell lymphoma, the majority of which are restricted to malignancies of haematopoietic and lymphoid origin and influence T-cell signalling pathways. Our statistical analysis of putative driver gene mutations and specific causal peaks within each signature revealed several associations. The recurrent STAT3 Y640F mutation is found in different T-cell NHL including both CTCL and ATLL and confers constitutive STAT3 phosphorylation and increased transcriptional activity<sup>51</sup>. The association of the STAT3 Y640F variant with signature 5 is not surprising as this signature is considered to be a feature of ageing<sup>52</sup>. We were unable to assess the signature association of recurrent STAT5B N642H, detected almost exclusively in EATL/HSTL. This is due to the low mutational load in these subtypes which prevented the assignment of signature contributions on an individual basis. Whole genome sequencing of these subtypes would provide a clearer picture of the mutational processes involved.

The recurrent IRF4 K59R mutation in ATLL was associated with signature 17 unique to ATLL. Mutations of IRF4 are frequent in B-cell malignancies including myeloma<sup>53</sup>, DLBCL<sup>54</sup> and primary mediastinal large B-cell lymphoma<sup>55</sup> and IRF4 gene rearrangements are also found in myeloma<sup>56</sup> and PTCL<sup>57</sup>, but intriguingly the most common point mutation, the K59R variant, is exclusively associated with ATLL<sup>58</sup>. The HTLV-1 Tax gene induces activation of the NF- $\kappa$ B pathway in ATLL and subsequent induction of IRF4, a key transcription factor downstream of NF- $\kappa$ B<sup>59</sup>. Tax gene expression is subsequently downregulated<sup>60</sup> and transformed T-cells in ATLL accumulate activating mutations of genes involved in TCR, NF- $\kappa$ B, and CD28 signalling pathways including the IRF4 K59R mutant<sup>11</sup>. Functional studies have confirmed that the activating K59R mutation is a key driver in ATLL and its effects include repression of apoptosis and DNA repair<sup>58</sup>. Specifically constitutive expression of IRF4 in T-cells directly reduces expression of critical DNA repair and checkpoint genes and increases the sensitivity of T-cells to genotoxic stress<sup>61</sup>. Therefore the association of causal signature 17 peaks with the IRF4 K59R mutation in ATLL suggests a strong selection pressure for this mutation which could directly generate an increased mutation rate of other genes due to suppression of DNA repair mechanisms. The underlying cause for signature 17 remains unknown although increasing evidence suggests oxidative damage may contribute<sup>62</sup>. The role of IRF4 in repression of DNA repair mechanisms provides further insight into the aetiology of signature 17 in ATLL.

In conclusion the identification of a clonotypic UV mutational signature in MF and CD4+ peripheral blood T-cells from SS patients confirms that environmental UV exposure contributes to the mutational burden and is a likely causal factor in the transformation of T-cells which are either circulating through or resident in the skin. Importantly these findings also raise the possibility that non-malignant T-cells could acquire UV induced mutations which might modulate normal immune responses.

## Methods

**Data assembly and pre-processing.** On 1<sup>st</sup> June 2017 Pubmed was searched using the terms ("Next Generation Sequencing"[All Fields] OR "Whole Exome Sequencing"[All Fields]) AND ("T cell lymphoma"[Tiab] OR "Mycosis Fungoides"[Tiab] OR "Sezary Syndrome"[Tiab] OR "T cell prolymphocytic leukaemia"[Tiab] OR "T cell large granular lymphocyte leukaemia"[Tiab] OR "Adult T cell leukemia/lymphoma"[Tiab] OR "CD30 positive T cell lymphoproliferative disorder"[Tiab]) NOT ("review"[Publication Type]) and Scopus was searched using the terms ((ALL ("Next Generation Sequencing") OR ALL ("Whole Exome Sequencing")) AND (TITLE-ABS-KEY ("T cell lymphoma") OR TITLE-ABS-KEY ("Mycosis Fungoides") OR TITLE-ABS-KEY ("Sezary Syndrome") OR TITLE-ABS-KEY ("T cell prolymphocytic leukaemia") OR TITLE-ABS-KEY ("T cell large granular lymphocyte leukaemia") OR TITLE-ABS-KEY ("Adult T cell leukemia/lymphoma") OR TITLE-ABS-KEY ("CD30 positive T cell lymphoproliferative disorder"))) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "le")). The outputs were combined to identify 121 unique reports which were manually curated to identify those which reported whole exome sequencing. Searches were also performed of sequencing archives including NCBI Sequence Read Archive, NCBI dbGap, EMBL European Genome-phenome Archive and EMBL European Nucleotide Archive to check for data which was not covered by the literature search.

To avoid skewing the analysis, we included only studies where full mutation details were included in the published data and studies were excluded if they only listed driver mutations, or genes which were mutated in more than one patient. We also only included samples where mutation calling was performed by comparison to a paired germline control.

Choi *et al.* only included a list of mutations in potential driver genes so their full mutation list was extracted from a data resource covering all the sequencing of CTCL<sup>63</sup>. Mutations from whole exome data of 403 patients were included and a summary by study and disease subtype is provided in Table 1. Experimental details, including capture and variant calling methods, of each study are included in Supplementary Table 2.

Mutation lists were extracted from the supplementary tables of studies and combined in R<sup>64</sup>. Where the reference assembly was not hg19, co-ordinates were lifted over using the rtracklayer<sup>65</sup> package. Finally, co-ordinates were sense checked using the GenomicRanges<sup>66</sup> and BSgenome.Hsapiens.UCSC.hg19<sup>67</sup> packages to ensure that the listed reference base corresponded to the given location. The full mutation dataset generated and analysed during the study are available in the Dryad repository, <https://doi.org/10.5061/dryad.8cz8w9gkm>.

Clinical data, including age, sex, disease stage and treatment regime at sampling or prior to sample collection, were collected from each publication where provided but this information was incomplete and resultant clinical groupings were too small for subgroup analyses. Treatment data was available for 50% of the cohort of which 70% were treatment naïve at the time of sampling (Supplementary Table 3).

**Identification of mutational signatures.** Combined SNV catalogues, i.e. the count of each mutation type (e.g. C>T) in different trinucleotide contexts (e.g. A[C>T]A), were calculated for each disease subtype and mutational signatures identified using a non-negative least squares fit against the set of 30 COSMIC signatures adjusted for exome trinucleotide frequencies. As exome data is prone to false positives, a post processing algorithm was applied to each subtype to remove signatures which had a minimal impact upon the cosine similarity between the SNV catalogue and that reconstructed using the identified signature contributions (maximum allowed reduction in cosine similarity was 0.05).

For identification of signatures on an individual patient level, fitting was restricted to the subset of signatures identified on the disease level<sup>29</sup>. Additionally, signatures were removed if assigned to less than 20 mutations in a sample and no signatures were assigned if the cosine similarity between the model and the data was less than 0.6.

**Association between recurrent mutations and signatures.** A study by Park *et al.*<sup>10</sup> defined 55 driver genes in CTCL using multiple independent analytical pipelines and confirmed their relevance to the wider subset of mature T-cell NHL and so were used in this analysis. SNV catalogues were limited to those specific variants which recurred within these 55 defined driver genes. 34 recurrent variants were identified in three or more patients, however we determined that recurrence in at least five patients was required to identify signature associations as otherwise the power of the statistical test was decreased. A separate analysis of each tumour type was performed but this approach was limited with the exception of ATLL because of the requirement for both a highly recurrent variant and sufficient SNVs per patient to estimate a signature contribution. Causal peaks for each signature were assigned by comparing the profiles of the nine signatures and identifying mutation contexts strongly associated with a particular signature. Summation of the causal peaks for each signature was then used to calculate a percentage contribution of that signature in all patients with more than 20 SNVs. For each recurrent mutation and signature pair, a Wilcoxon rank sum test was used to determine if there was a significant difference in signature contribution between the wild type and mutant groups.

**Validation in UK 100,000 genomes cohort.** All samples were processed by Genomics England and analysed through their cancer pipeline. Briefly, samples were prepared using an Illumina TruSeq DNA PCR-Free preparation kit and then sequenced on a HiSeq X generating 150 bp paired-end reads. Quantity and quality of sequence and presence of cross-contamination were assessed for each sample prior to variant calling. Reads were aligned to human reference genome GRCh38-Decoy+EBV by ISAAC then Strelka was used to call small variants and perform tumour-normal subtraction. Gene annotations were used to determine whether mutations within genes affected the transcriptional or non-transcriptional strand.

## Data availability

The dataset generated and analysed during the current study are available in the Dryad repository, <https://doi.org/10.5061/dryad.8cz8w9gkm>.

Received: 17 June 2020; Accepted: 27 January 2021

Published online: 17 February 2021

## References

1. Swerdlow, S. H. *et al.* The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood* **127**, 2375–2390 (2016).
2. Vose, J., Armitage, J., Weisenburger, D. & International T-Cell Lymphoma Project. International peripheral T-cell and natural killer/T-cell lymphoma study: pathology findings and clinical outcomes. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **26**, 4124–4130 (2008).
3. Nussenzweig, A. & Nussenzweig, M. C. Origin of chromosomal translocations in lymphoid cancer. *Cell* **141**, 27–38 (2010).
4. Tsuyama, N., Sakamoto, K., Sakata, S., Dobashi, A. & Takeuchi, K. Anaplastic large cell lymphoma: pathology, genetics, and clinical aspects. *J. Clin. Exp. Hematop. JCEH* **57**, 120–142 (2017).
5. Ishitsuka, K. & Tamura, K. Human T-cell leukaemia virus type I and adult T-cell leukaemia-lymphoma. *Lancet Oncol.* **15**, e517–e526 (2014).
6. Shannon-Lowe, C. & Rickinson, A. The global landscape of EBV-associated tumors. *Front. Oncol.* **9**, 713 (2019).
7. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).
8. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
9. Zhang, L. *et al.* Single-cell whole-genome sequencing reveals the functional landscape of somatic mutations in B lymphocytes across the human lifespan. *Proc. Natl. Acad. Sci.* **116**, 9014–9019 (2019).
10. Park, J. *et al.* Genomic analysis of 220 CTCLs identifies a novel recurrent gain-of-function alteration in RLTPR (p.Q575E). *Blood* **130**, 1430–1440 (2017).
11. Kataoka, K. *et al.* Integrated molecular analysis of adult T cell leukemia/lymphoma. *Nat. Genet.* **47**, 1304–1315 (2015).
12. Jiang, L. *et al.* Exome sequencing identifies somatic mutations of DDX3X in natural killer/T-cell lymphoma. *Nat. Genet.* **47**, 1061–1066 (2015).
13. McKinney, M. *et al.* The genetic basis of hepatosplenic T-cell Lymphoma. *Cancer Discov.* **7**, 369–379 (2017).
14. Moffitt, A. B. *et al.* Enteropathy-associated T cell lymphoma subtypes are characterized by loss of function of SETD2. *J. Exp. Med.* **214**, 1371–1386 (2017).
15. Seifert, M., Scholtysik, R. & Küppers, R. Origin and pathogenesis of B Cell lymphomas. *Methods Mol. Biol. Clifton NJ* **1956**, 1–33 (2019).
16. Choi, J. *et al.* Genomic landscape of cutaneous T cell lymphoma. *Nat. Genet.* **47**, 1011–1019 (2015).
17. Da Silva Almeida, A. C. *et al.* The mutational landscape of cutaneous T cell lymphoma and Sézary syndrome. *Nat. Genet.* **47**, 1465–1470 (2015).
18. Palomero, T. *et al.* Recurrent mutations in epigenetic regulators, RHOA and FYN kinase in peripheral T cell lymphomas. *Nat. Genet.* **46**, 166–170 (2014).
19. Prasad, A. *et al.* Identification of gene mutations and fusion genes in patients with Sézary syndrome. *J. Invest. Dermatol.* **136**, 1490–1499 (2016).
20. Roberti, A. *et al.* Type II enteropathy-associated T-cell lymphoma features a unique genomic profile with highly recurrent SETD2 alterations. *Nat. Commun.* **7**, 12602 (2016).
21. Sakata-Yanagimoto, M. *et al.* Somatic RHOA mutation in angioimmunoblastic T cell lymphoma. *Nat. Genet.* **46**, 171–175 (2014).
22. Ungewickell, A. *et al.* Genomic analysis of mycosis fungoides and Sézary syndrome identifies recurrent alterations in TNFR2. *Nat. Genet.* **47**, 1056–1060 (2015).
23. Wang, L. *et al.* Genomic profiling of Sézary syndrome identifies alterations of key T cell signaling and differentiation genes. *Nat. Genet.* **47**, 1426–1434 (2015).
24. Woollard, W. J. *et al.* Candidate driver genes involved in genome maintenance and DNA repair in Sézary syndrome. *Blood* **127**, 3387–3397 (2016).
25. Yoo, H. Y. *et al.* A recurrent inactivating mutation in RHOA GTPase in angioimmunoblastic T cell lymphoma. *Nat. Genet.* **46**, 371–375 (2014).
26. Teras, L. R. *et al.* 2016 US lymphoid malignancy statistics by World Health Organization subtypes. *CA. Cancer J. Clin.* **66**, 443–459 (2016).
27. Forbes, S. A. *et al.* COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45**, D777–D783 (2017).
28. Shirole, N. H. *et al.* TP53 exon-6 truncating mutations produce separation of function isoforms with pro-tumorigenic functions. *eLife* **5**, e17929 (2016).
29. Maura, F. *et al.* A practical guide for mutational signature analysis in hematological malignancies. *Nat. Commun.* **10**, 2969 (2019).
30. Secrier, M. *et al.* Mutational signatures in esophageal adenocarcinoma define etiologically distinct subgroups with therapeutic relevance. *Nat. Genet.* **48**, 1131–1141 (2016).
31. Sun, J. *et al.* Genomic signatures reveal DNA damage response deficiency in colorectal cancer brain metastases. *Nat. Commun.* **10**, 3190 (2019).
32. Temko, D., Tomlinson, I. P. M., Severini, S., Schuster-Böckler, B. & Graham, T. A. The effects of mutational processes and selection on driver mutations across cancer types. *Nat. Commun.* **9**, 1857 (2018).
33. Patil, V. S. *et al.* Precursors of human CD4+ cytotoxic T lymphocytes identified by single-cell transcriptome analysis. *Sci. Immunol.* **3**, eaan8664 (2018).
34. Clark, R. A. *et al.* The vast majority of CLA+ T cells are resident in normal skin. *J. Immunol.* **176**, 4431–4439 (2006).
35. Li, X., Pasche, B., Zhang, W. & Chen, K. Association of MUC16 mutation with tumor mutation load and outcomes in patients with gastric cancer. *JAMA Oncol.* **4**, 1691–1698 (2018).
36. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* **3**, 246–259 (2013).
37. Campbell, J. J., Clark, R. A., Watanabe, R. & Kupper, T. S. Sezary syndrome and mycosis fungoides arise from distinct T-cell subsets: a biologic rationale for their distinct clinical behaviors. *Blood* **116**, 767–771 (2010).
38. Gehad, A. *et al.* A primary role for human central memory cells in tissue immunosurveillance. *Blood Adv.* **2**, 292–298 (2018).
39. Bustamante, M. *et al.* The acute effects of ultraviolet radiation on the blood transcriptome are independent of plasma 25OHD3. *Environ. Res.* **159**, 239–248 (2017).
40. McGregor, J. M. *et al.* Spectrum of p53 gene mutations suggests a possible role for ultraviolet radiation in the pathogenesis of advanced cutaneous lymphomas. *J. Invest. Dermatol.* **112**, 317–321 (1999).
41. McGirt, L. Y. *et al.* Whole-genome sequencing reveals oncogenic mutations in mycosis fungoides. *Blood* **126**, 508–519 (2015).
42. Kucab, J. E. *et al.* A compendium of mutational signatures of environmental agents. *Cell* **177**, 1–16 (2019).

43. PCAWG Mutational Signatures Working Group *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
44. Lu, D. *et al.* Occupational ultraviolet exposure and risk of non-Hodgkin's lymphomas: a meta-analysis. *Oncotarget* **8**, 62358–62370 (2017).
45. Morales-Suárez-Varela, M. M. *et al.* Occupational sun exposure and mycosis fungoides: a European multicenter case-control study. *J. Occup. Environ. Med.* **48**, 390–393 (2006).
46. van Leeuwen, M. T. *et al.* Latitude gradients for lymphoid neoplasm subtypes in Australia support an association with ultraviolet radiation exposure. *Int. J. Cancer* **133**, 944–951 (2013).
47. Cahoon, E. K. *et al.* Relationship between ambient ultraviolet radiation and non-Hodgkin lymphoma subtypes: a U.S. population-based study of racial and ethnic groups. *Int. J. Cancer* **136**, 432–441 (2015).
48. Schwarz, T. The dark and the sunny sides of UVR-induced immunosuppression: photoimmunology revisited. *J. Invest. Dermatol.* **130**, 49–54 (2010).
49. Galgano, A., Barinov, A., Vasseur, F., de Villartay, J.-P. & Rocha, B. CD8 memory cells develop unique DNA repair mechanisms favoring productive division. *PLoS ONE* **10**, e0140849 (2015).
50. Puebla-Osorio, N. *et al.* Induction of B-cell lymphoma by UVB radiation in p53 haploinsufficient mice. *BMC Cancer* **11**, 36 (2011).
51. Butler, R. M. *et al.* Contribution of STAT3 and RAD23B in primary sézary cells to histone deacetylase inhibitor FK228 resistance. *J. Invest. Dermatol.* **139**, 1975–1984.e2 (2019).
52. Alexandrov, L. B. *et al.* Clock-like mutational processes in human somatic cells. *Nat. Genet.* **47**, 1402–1407 (2015).
53. Walker, B. A. *et al.* Mutational spectrum, copy number changes, and outcome: results of a sequencing study of patients with newly diagnosed myeloma. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **33**, 3911–3920 (2015).
54. Mareschal, S. *et al.* Whole exome sequencing of relapsed/refractory patients expands the repertoire of somatic mutations in diffuse large B-cell lymphoma. *Genes. Chromosomes Cancer* **55**, 251–267 (2016).
55. Mottok, A. *et al.* Integrative genomic analysis identifies key pathogenic mechanisms in primary mediastinal large B-cell lymphoma. *Blood* **134**, 802–813 (2019).
56. Iida, S. *et al.* Deregulation of MUM1/IRF4 by chromosomal translocation in multiple myeloma. *Nat. Genet.* **17**, 226–230 (1997).
57. Feldman, A. L. *et al.* Recurrent translocations involving the IRF4 oncogene locus in peripheral T-cell lymphomas. *Leukemia* **23**, 574–580 (2009).
58. Cherian, M. A. *et al.* An activating mutation of interferon regulatory factor 4 (IRF4) in adult T-cell leukemia. *J. Biol. Chem.* **293**, 6844–6858 (2018).
59. Sharma, S. *et al.* Regulation of IFN regulatory factor 4 expression in human T cell leukemia virus-I-transformed T cells. *J. Immunol.* **169**, 3120–3130 (2002).
60. Azran, I., Schavinsky-Khrapunsky, Y. & Aboud, M. Role of tax protein in human T-cell leukemia virus type-I leukemogenicity. *Retrovirology* **1**, 20 (2004).
61. Mamane, Y. *et al.* Repression of IRF-4 target genes in human T cell leukemia virus-1 infection. *Oncogene* **21**, 6751–6765 (2002).
62. Tomkova, M., Tomek, J., Kriaucionis, S. & Schuster-Böckler, B. Mutational signature distribution varies with DNA replication timing and strand asymmetry. *Genome Biol.* **19**, 129 (2018).
63. Chang, L.-W. *et al.* An integrated data resource for genomic analysis of cutaneous T-cell lymphoma. *J. Invest. Dermatol.* **138**, 2681–2683 (2018).
64. R Core Team (2017). *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>).
65. Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* **25**, 1841–1842 (2009).
66. Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).
67. BSgenome.Hsapiens.UCSC.hg19. *Bioconductor* <http://bioconductor.org/packages/BSgenome.Hsapiens.UCSC.hg19/>

## Acknowledgements

We are grateful to Antony Young for providing insightful advice on the photobiology of UV radiation. CLJ was supported by an educational grant from Galderma. SNZ, AD and TDA are supported by CRUK Pioneer Award (C60100/A23433), CRUK Advanced Clinician Scientist Fellowship (C60100/A23916) and NIHR-BRC Cambridge core grant, and SNZ is supported by Wellcome-Beit Prize, Wellcome Trust Strategic Award (WT100126/B/13/Z) and CRUK PRECISION Grand Challenge award. This research was supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London and/or the NIHR Clinical Research Facility. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. This research was made possible through access to the data and findings generated by the 100,000 Genomes Project. The 100,000 Genomes Project is managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care). The 100,000 Genomes Project is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure. The 100,000 Genomes Project uses data provided by patients and collected by the National Health Service as part of their care and support.

## Author contributions

C.L.J., T.J.M., S.N.Z. and S.J.W. designed the work; C.L.J. and G.E.R.C. performed acquisition of data; C.L.J., A.D., T.D.A., G.E.R.C. and S.N.Z. analysed the data; C.L.J., A.D., V.G., T.J.M., S.N.Z. and S.J.W. performed interpretation of data; C.L.J., V.G. and S.J.W. drafted the work; C.L.J., A.D., V.G., T.J.M., S.N.Z. and S.J.W. revised the work.

## Competing interests

Research support was provided by Galderma. The funder had no role in the conceptualization, design, data collection, analysis, decision to publish, or preparation of the manuscript.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-83352-4>.

**Correspondence** and requests for materials should be addressed to S.J.W.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

## Genomics England Research Consortium

John C. Ambrose<sup>4</sup>, Prabhu Arumugam<sup>4</sup>, Emma L. Baple<sup>4</sup>, Marta Bleda<sup>4</sup>, Freya Boardman-Pretty<sup>4,5</sup>, Jeanne M. Boissiere<sup>4</sup>, Christopher R. Boustred<sup>4</sup>, Helen Brittain<sup>4</sup>, Mark J. Caulfield<sup>4,5</sup>, Georgia C. Chan<sup>4</sup>, Clare E. H. Craig<sup>4</sup>, Louise C. Daugherty<sup>4</sup>, Anna de Burca<sup>4</sup>, Andrew Devereau<sup>4</sup>, Greg Elgar<sup>4,5</sup>, Rebecca E. Foulger<sup>4</sup>, Tom Fowler<sup>4</sup>, Pedro Furió-Tari<sup>4</sup>, Adam Giess<sup>4</sup>, Joanne M. Hackett<sup>4</sup>, Dina Halai<sup>4</sup>, Angela Hamblin<sup>4</sup>, Shirley Henderson<sup>4,5</sup>, James E. Holman<sup>4</sup>, Tim J. P. Hubbard<sup>4</sup>, Kristina Ibáñez<sup>4,5</sup>, Rob Jackson<sup>4</sup>, Louise J. Jones<sup>4,5</sup>, Dalia Kasperaviciute<sup>4,5</sup>, Melis Kayikci<sup>4</sup>, Athanasios Kousathanas<sup>4</sup>, Lea Lahnstein<sup>4</sup>, Kay Lawson<sup>4</sup>, Sarah E. A. Leigh<sup>4</sup>, Ivonne U. S. Leong<sup>4</sup>, Javier F. Lopez<sup>4</sup>, Fiona Maleady-Crowe<sup>4</sup>, Joanne Mason<sup>4</sup>, Ellen M. McDonagh<sup>4,5</sup>, Loukas Moutsianas<sup>4,5</sup>, Michael Mueller<sup>4,5</sup>, Nirupa Murugaesu<sup>4</sup>, Anna C. Need<sup>4,5</sup>, Peter O'Donovan<sup>4</sup>, Chris A. Odhams<sup>4</sup>, Andrea Orioli<sup>4</sup>, Christine Patch<sup>4,5</sup>, Mariana Buongiorno Pereira<sup>4</sup>, Daniel Perez-Gil<sup>4</sup>, Dimitris Polychronopoulos<sup>4</sup>, John Pullinger<sup>4</sup>, Tahrira Rahim<sup>4</sup>, Augusto Rendon<sup>4</sup>, Pablo Riesgo-Ferreiro<sup>4</sup>, Tim Rogers<sup>4</sup>, Mina Ryten<sup>4</sup>, Kevin Savage<sup>4</sup>, Kushmita Sawant<sup>4</sup>, Richard H. Scott<sup>4</sup>, Afshan Siddiq<sup>4</sup>, Alexander Sieghart<sup>4</sup>, Damian Smedley<sup>4,5</sup>, Katherine R. Smith<sup>4,5</sup>, Samuel C. Smith<sup>4</sup>, Alona Sosinsky<sup>4,5</sup>, William Spooner<sup>4</sup>, Helen E. Stevens<sup>4</sup>, Alexander Stuckey<sup>4</sup>, Razvan Sultana<sup>4</sup>, Mélanie Tanguy<sup>4</sup>, Ellen R. A. Thomas<sup>4,5</sup>, Simon R. Thompson<sup>4</sup>, Carolyn Tregidgo<sup>4</sup>, Arianna Tucci<sup>4,5</sup>, Emma Walsh<sup>4</sup>, Sarah A. Watters<sup>4</sup>, Matthew J. Welland<sup>4</sup>, Eleanor Williams<sup>4</sup>, Katarzyna Witkowska<sup>4,5</sup>, Suzanne M. Wood<sup>4,5</sup> & Magdalena Zarowiecki<sup>4</sup>

<sup>4</sup>Genomics England, London, UK. <sup>5</sup>William Harvey Research Institute, Queen Mary University of London, London EC1M 6BQ, UK.