

Chapter 6: *PiggyBac* Insertional Mutagenesis of the Mature B Cell Compartment

6.1 Introduction

Multiple myeloma (MM) is a plasma cell malignancy that is incurable with conventional therapy and causes nearly 2% of cancer deaths (Jemal et al.). It is preceded by the asymptomatic presence of a monoclonal protein in serum/plasma; termed the monoclonal gammopathy of uncertain significance (MGUS). MGUS occurs in 3% of people over the age of 50 (Kyle et al., 2006) and transforms to MM at a rate of approximately 1% per year (Kyle et al., 2002), but the molecular mechanisms that drive progression are largely unknown.

MM is a heterogeneous disease. A hyperdiploid karyotype occurs in approximately 50% of cases, but the driver for chromosome accumulation is not known (Chng et al., 2007). Recurrent chromosomal translocations involving the immunoglobulin loci are found in approximately 70% of non-hyperdiploid tumours (Avet-Loiseau, 2007). These translocations are thought to represent primary oncogenic events that occur in normal B cells during germinal center development. Breakpoints are usually within or near the immunoglobulin switch regions or VDJ sequences (Chng et al., 2007). Recurrent translocation partners include Cyclin D, MAF and MMSET/FGFR3 (Chng et al., 2007). These genetic sub-groups of MM can be used to predict clinical response and guide treatment decisions (Avet-Loiseau et al., 2007; Palumbo and Rajkumar, 2009).

Massive parallel sequencing studies have recently highlighted the remarkable molecular heterogeneity of multiple myeloma and described several additional molecular abnormalities. Deep sequencing of 38 tumour-normal pairs revealed frequent mutations in genes involved in protein translation, histone methylation and blood coagulation (Chapman et al., 2011). Along with previously reported recurrent mutations in MM such as *KRAS*, *NRAS* and *TP53*, this paper described several additional point mutations which may act as driver lesions including *CCND1*, *DIS3*, *FAM46C*, *BRAF* and *IRF4* (Chapman et al., 2011). Another group described further

candidate driver genes including truncating mutations of *SP140*, *LTB*, *ROBO1* and missense mutations in *EGR1* (Bolli et al., 2014). The striking features in all of the sequencing studies performed to date is the heterogeneity of mutational spectra between cases and the large burden of variants within each tumour.

In 2008 Chesi and colleagues described a novel mouse model, which was the first to accurately recapitulate many of the clinical features of human MM (Chesi et al., 2008). In this model the human *c-MYC* transgene was placed under the transcriptional control of the V_k promoter (V_k^*MYC) and expressed in late B-cells. The third codon of the lead V-kappa exon was mutated to a stop codon (figure 6.1), so although it spliced in frame to the human *MYC* exons, translation of the MYC protein did not occur in tested tissues. The stop codon was engineered to overlap with a preferential target sequence for endogenous Activation Induced Deaminase (AID) and the native kappa light chain gene regulatory elements were maintained to invite targeting by AID. Therefore, it was anticipated that the stop codon would be mutated in a small proportion of B cells during germinal centre development allowing expression of the *MYC* transgene in these cells (Betz et al., 1994; Papavasiliou and Schatz, 2000). With age all mice developed progressive monoclonal plasma cell expansion and as the tumours did not show intra-clonal heterogeneity of B cell receptor sequences, the authors concluded they were not subject to ongoing somatic hypermutation (SHM).

MYC is a global transcriptional regulator that controls cell proliferation, differentiation, growth and survival (Larsson and Henriksson; Meyer and Penn, 2008). Translocations involving *c-MYC* that lead to its inappropriate expression are an initiating event in human Burkitt lymphoma, an aggressive, mature B cell non-Hodgkin lymphoma, but are not unique to this disease (Dalla-Favera et al., 1982) (Au et al., 2004; Kanungo et al., 2005). Notably the incidence of Burkitt lymphoma was low in the V_k^*MYC mice and no cases of aggressive pro-B lymphoma were reported (Chesi et al., 2008). In mice, *Myc*-Immunoglobulin translocations are an initiating event in plasmacytoma (Ohno et al., 1979; Shen-Ong et al., 1982), whilst constitutive *MYC* expression in early B cells often leads to aggressive pro-B or diffuse high-grade blastic B-cell lymphoma (Adams et al., 1985; Butzler et al., 1997; Chesi et al., 2008; Harris et al., 1988; Kovalchuk et al., 2000; Palomo et al., 1999; Park et al., 2005; Refaeli et al., 2008; Zingone et al.). However, transgenic expression of *MYC* alone does not appear to be sufficient for lymphoma

development as *Eu-Myc* transgenic mice initially demonstrate a benign expansion in pre-B cells (Harris et al., 1988).

Translocations involving *MYC* do occur in human MM, but they are thought to be late progression events (Chng et al., 2007) and are rare in MGUS. In contrast, they occur in 15% of MM tumours and nearly 90% of human myeloma cell lines (Chng et al., 2007). However, mutations or translocations involving the *MYC* locus are not required for *MYC* activation (Meyer and Penn, 2008) and *c-MYC* over-expression due to stimulation by IL-6 and other mechanisms, occurs early in MM (Chesi et al., 2008). Compared to normal bone marrow plasma cells *MYC* expression is significantly higher in cells from MGUS and even higher in MM (Chesi et al., 2008; Zingone et al.). Gene expression profiling data suggests that patients with MM that expresses *N-MYC* or very high levels of *c-MYC* have worse survival (Chng et al., 2007; Janz, 2008).

Transposon insertional mutagenesis (IM) provides a powerful approach for the identification and validation of cancer drivers that compliments human sequencing efforts. In order to further investigate genes involved in the pathogenesis of multiple myeloma I adapted the *Vk*MYC* model to target (*hyper*)*piggyBac* (*hPB*) IM to the mature B cell compartment (figure 6.1).

The *PB* transposon system was chosen because of its efficiency, extensive access to the genome, lack of excision footprint and high rate of intragenic insertions (Cadinanos and Bradley, 2007; Liang et al., 2009; Rad, 2010; Wang et al., 2008). The hyperactive *PB* (*hPB*) cDNA was used, which is an enhanced version of the mouse codon optimised *PB* (*mPB*) but with a ten-fold higher transposition efficiency versus *mPB*, which is itself much more efficient than the native *Trichoplusia Ni* version (Cadinanos and Bradley, 2007; Liang et al., 2009; Yusa et al., 2011).

Two related constructs were generated. In the first, the coding exons of *MYC* were replaced by the *hPB* transposase (figure 6.1B). In the second, *MYC* and *hPB* were expressed together from the same cistron using a *Thosea asigna* 2A (T2A) linker peptide that is hydrolysed very quickly after translation to generate equimolar amounts of the *MYC* and *hPB* proteins (Szymczak et al., 2004). In the *Vk*MYC-TA-hPB* mice there was minimal disruption of the original *Vk*MYC* construct as the latter generated highly penetrant plasma cell tumours (Chesi et al., 2008). *T2A-hPB* cDNA

was introduced in frame, after the penultimate codon of *MYC*(thus removing the stop codon), but the remainder of the *Vk*MYC* construct was left intact.

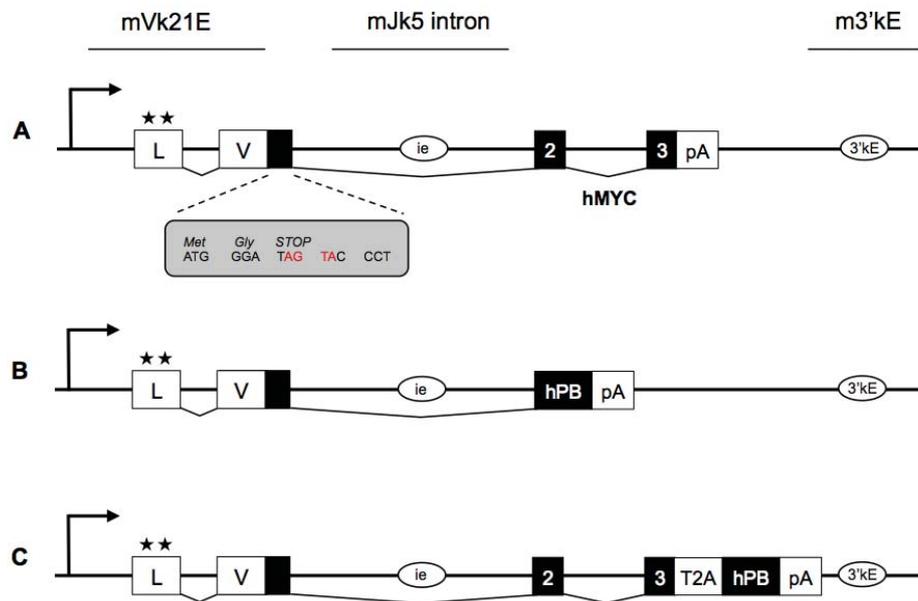


Figure 6.1: The *Vk*MYC*, *Vk*hPB* and *Vk*MYC-TA-hPB* constructs. This figure is repeated from Chapter 2 (figure 2.1) for ease of reference

(A) In the *Vk*MYC* construct published by Chesi et al, the Jk5 exon in the rearranged mouse Vk21 kappa light chain gene was replaced by a short coding exon containing a Kozak ATG(Chesi et al., 2008). Human *MYC* exons 2 and 3 replaced the C1k region. Transcription initiates at the Vk21e proximal promoter (\blacktriangleright), extends to the leader (L) and Vk (V) exons, splices in frame to human *MYC* (*hMYC*) and terminates at the endogenous polyA signal (PA). ATG codons (*) in L were mutated to ACG to stop initiation of translation at these positions. Intronic (ie) and 3'kappa (3'kE) enhancers are maintained. The DNA sequence immediately downstream of the Vk21 ATG is depicted. Nucleotides in red letters fit the DGYW consensus for AID targeting.

(B) In *Vk*hPB*, *hMYC* is replaced by the *hPB* cDNA, carrying a splice acceptor signal that leads to splicing of *hPB* mRNA in-frame with the reading frame “opened” by AID mutation of the upstream TAG stop codon.

(C) In *Vk*MYC-TA-hPB* the cDNA for the self-cleaving peptide T2A links *hPB* in-frame to *hMYC*. The chimaeric polypeptide produced from a single cistron is predicted to spontaneously dissociate into hMYC and hPB proteins.

The decision to generate two models was taken to enable comparisons between tumours derived in the presence and absence of MYC and because it was uncertain whether the *MYC* transgene would be critical for the development of MM. As the *Vk* regulatory elements in combination with the early stop codon requiring mutation by SHM should have, in principle, ensured that transgene activation was restricted to late B-cells, it was considered probable that *hPB* could drive the development of MM as well as other mature B cell tumours such as follicular or Burkitt lymphoma (figure 6.2 (Weigert and Weinstock, 2012)). Information on the role of *MYC* in initiating and driving these tumours would therefore be derived by comparing the tumour phenotype, latency and transposon integrations between the *Vk*MYC-TA-hPB* and *Vk*hPB* cohorts.

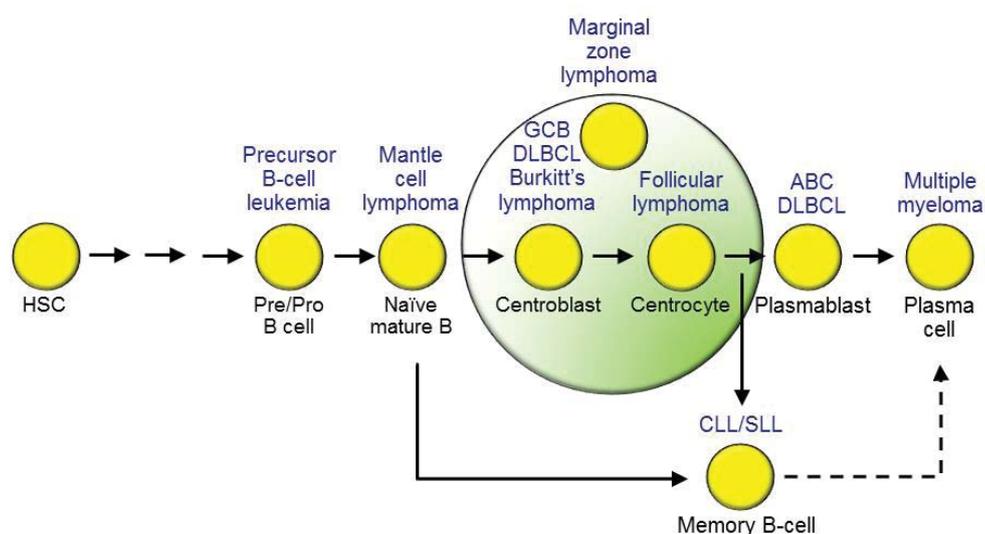


Figure 6.2: B cell maturation and lymphoma phenotypes. The stages of normal maturation are indicated in black and the corresponding neoplasm in blue text. The germinal centre is represented by the area shaded in green. GCB = germinal centre B cell like; ABC = activated B cell like; DLBCL = diffuse large B cell lymphoma; CLL/SLL = chronic lymphocytic leukaemia/small lymphocytic lymphoma. From Weigert and Weinstock, 2012.

6.2 Results

6.2.1 Cloning *Vk*HPB* and *Vk*MYC-TA-HPB*

The *Vk*MYC-TA-HPB* and *Vk*HPB* constructs were generated as described in Materials and Methods, sequenced by Sanger sequencing, linearised and sent for pronuclear injection at PolyGene AG (Switzerland). The *hPB* sequence in *Vk*MYC-TA-hPB* was identical to the expected sequence. The *Vk*hPB* construct had an S to G substitution at position 1520 in *hPB*. This corresponds to an intermediate stage of *hPB* development between *mPB* and the final *hPB*, and is expected to have a very slightly reduced transposition efficiency compared to the final *hPB* version, but still higher than *mPB* (Kosuke Yusa personal communication).

Three of the 46 tail samples received after pro-nuclear injection were positive for the *Vk*MYC-TA-hPB* transgene and four of 40 samples were positive for the *Vk*hPB* construct by PCR analysis. The seven founder (F0) mice were imported to our quarantine facility and colonies generated for re-derivation inside the WTSI animal facility. One line from each construct was selected for expansion and mating with low copy *GrOnc* transposon lines. This decision was primarily based on the fecundity of initial matings.

6.2.2 Validation of splicing in the transgenic constructs

Correct splicing of the constructs was initially confirmed by RT-PCR of RNA from U266 cells transiently transfected with each of the constructs. PCR products using primers annealing to exon 2 of *MYC* and to *hPB cDNA* are shown in figure 6.3. For the *Vk*HPB* construct, the *V* exon forward primer (*VkMycexon1F*) and the reverse primer in *hPB* (*SAHPBR*) also generated the expected RT-PCR product (190bp) (image not shown). The splicing was subsequently confirmed on mouse tumour RNA samples using the same method and further verified using capillary sequencing of PCR products.

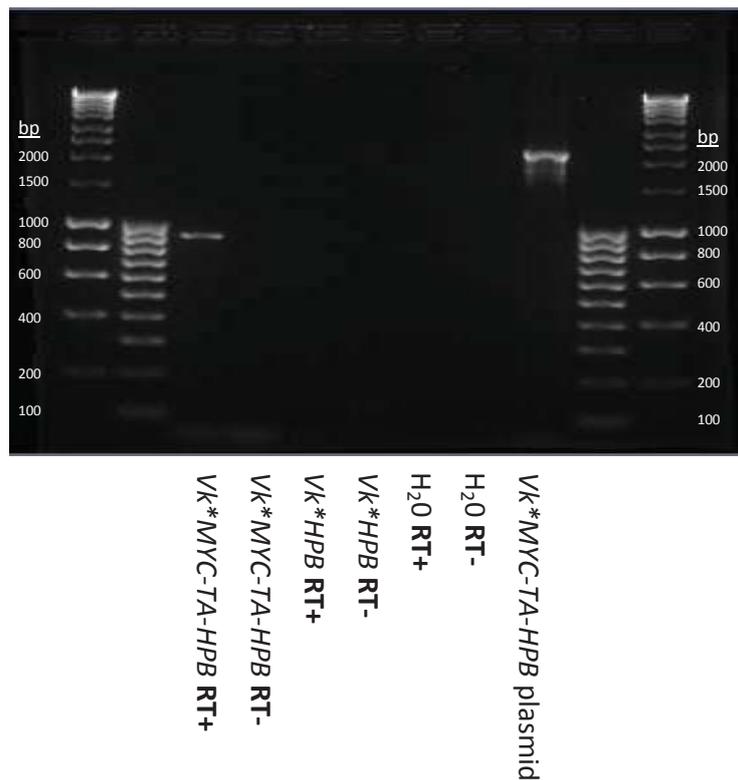


Figure 6.3: RT-PCR using primers *VkMycexon2F* and *SAHPBRCORTAHPBR*. RNA from the *Vk*MYC-TA-hPB* transfected U-266 cells generated the expected 900bp RT-PCR product. By contrast when *Vk*MYC-TA-hPB* plasmid DNA was used as a template a 2200bp PCR product was generated. As expected, no RT-PCR product was amplified from non-reverse transcribed RNA (RT-) or from RNA from *Vk*HPB* transfected cells.

6.2.3 The *Vk*MYC-TA-HPB* construct generates an active *PB* transposase: HAT resistance assay

After hydrolysis, the T2A linker leaves a single proline at the 5' end of *hPB* (Szymczak *et al.*, 2004). To test if this affected *PB* transposase activity, I performed a HAT resistance assay. In this assay, when an active transposase removes a *PB* transposon from within the X-linked *Hprt* gene locus of male ES cells, *Hprt* activity is restored permitting growth in HAT media. Transfection with the *Vk*MYC-TA-hPB* cDNA resulted in growth of HAT resistant colonies, in numbers proportional to the amount of transfected DNA. Colony numbers were comparable to those seen after transfection with *hPB* cDNA (positive control), confirming that an active transposase is generated from the bi-cistronic *MYC-TA-hPB* construct (Figure 6.4).

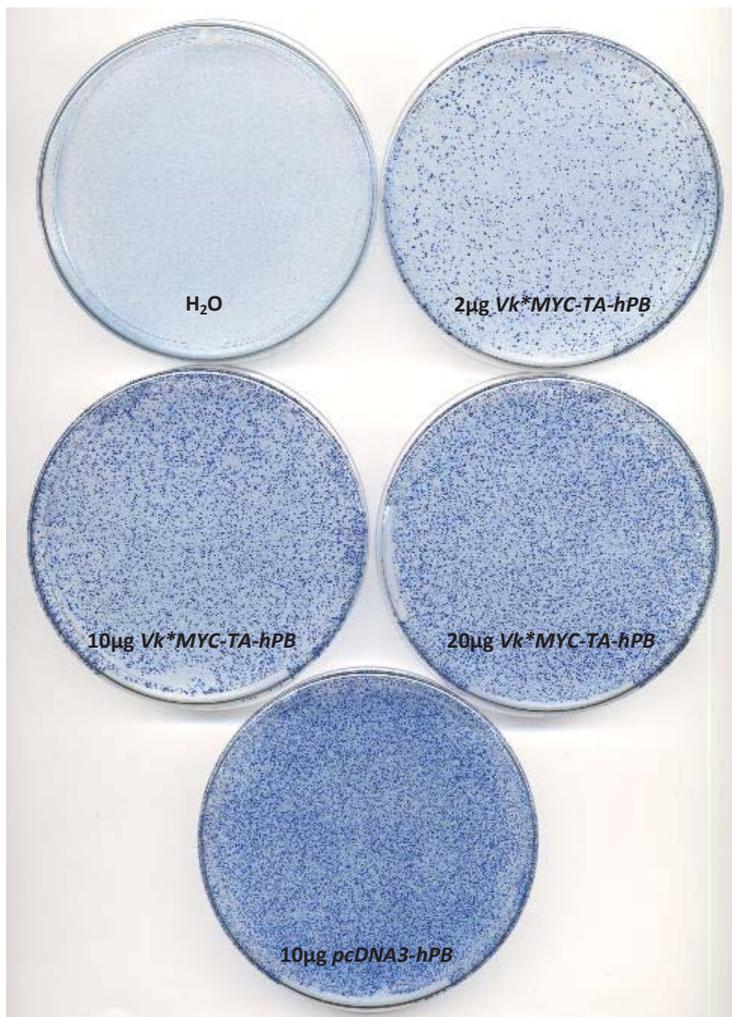


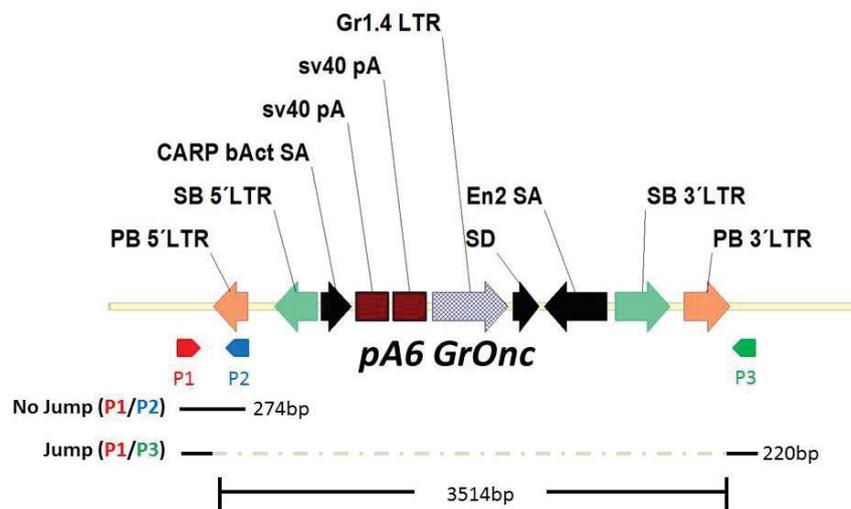
Figure 6.4: HAT resistance assay. Male ES cells harbouring a *PB* transposon in the *Hprt* locus were electroporated with H₂O or constructs *Vk*MYC-TA-hPB* or *pcDNA3-hPB* (positive control) using the indicated amounts of plasmid. Photographs are of colony growth from 1/10 platings (1/10th of 1x10⁷ electroporated cells plated).

6.2.4 The *hPB* transposase is active *in vivo*, although transposon mobilisation is not limited to the mature B cell compartment

The *Vk*MyC-TA-hPB* and *Vk*hPB* constructs were designed to specifically express *hPB* protein in the mature B cell compartment. This specificity would be imparted by the *Vk* regulatory elements and the presence of the early stop codon, designed to prevent expression of the *hPB* transposase in the absence of mutation by SHM. The stop codon was positioned such that it created a preferential target sequence for AID and it was anticipated this would be reverted in a small percentage of B cells during germinal centre development as previously described (Chesi et al., 2008). However, using “jump” PCR, we found that mobilisation of the transposon was also occurring in non-haematopoietic tissues (figure 6.5) and some insertional mutagenesis mice were observed to develop non-haematopoietic tumours. Furthermore, when spleen and bone marrow cells from IM mice were flow sorted and DNA was extracted separately from CD34+ (progenitor), CD3+ (T), Gr1+/Mac1+

(granulocytes), B220+/CD19+ (pro/mature B) and B220+/CD19- (pre-pro B) cells, the 'jump' PCR was positive in all lineages tested in the *MYC-TA-hPB* mouse and in all cells except granulocytes in the *hPB* mouse tested.

A



B

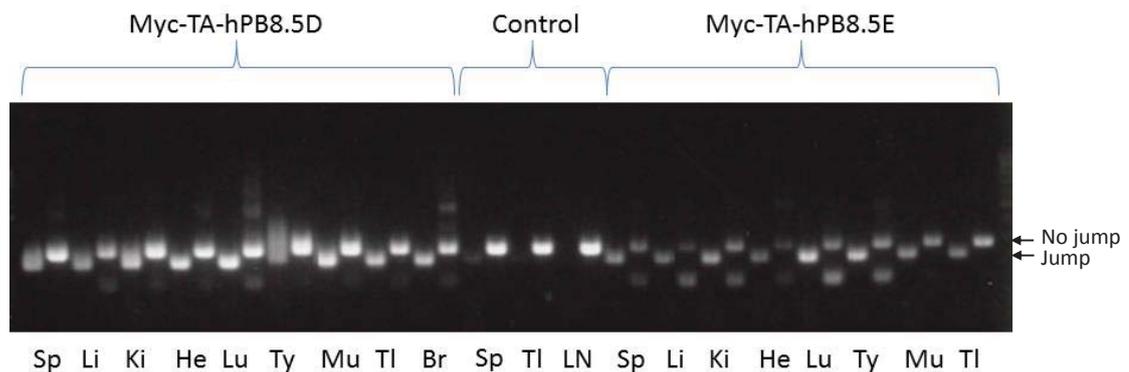


Figure 6.5: Tissue specificity of the *hPB* transposon. (A) Design of 'jump' and 'no jump' PCRs. The yellow line represents the plasmid backbone. When one or more transposons mobilise, the region flanked by the PB repeats is removed and a 220bp product is generated by the P1 and P3 primers (Jump PCR). In the absence of jumping these primers are separated by the full length of the *GrOnc* transposon (3.5kb) and no PCR product is generated. By contrast, the P1 and P2 primers will produce a 274bp product only when one or more transposons do not mobilise (No jump PCR). (B) Results from alternating 'jump' and 'no jump' PCRs on DNA extracted from various tissues from two *Vk*Myc-TA-hPB* IM mice (8.5D and 8.5E) and a control mouse which carried the *GrOnc* transposon, but no transposase. Sp-spleen, Li-liver, Ki-kidney, He-heart, Lu-lung, Ty-thymus, Mu-muscle, TI-tail, Br-brain, LN-lymph node.

6.2.5 Insertional mutagenesis mice have increased lymphoma-associated mortality

The $Vk^*MYC-TA-hPB$ and Vk^*hPB IM mice were born at expected Mendelian ratios. They died at a similar rate ($p=0.57$), which was significantly accelerated compared to non-IM mice (median survival 75.4, 71.9 and 91.1 weeks respectively) (figure 6.6). Of note, the survival of the $Vk^*MYC-TA-hPB$ only mice was also no different to the GRL (GrOnc) only mice which lacked the *MYC* transgene (median survival 94.3 v 93.2 weeks, $p = 0.61$ Mantel-Cox test). Intraperitoneal injection of sheep red blood cells did not alter the survival or tumour spectrum and mice were considered together in the results, regardless of whether or not they received antigen stimulation.

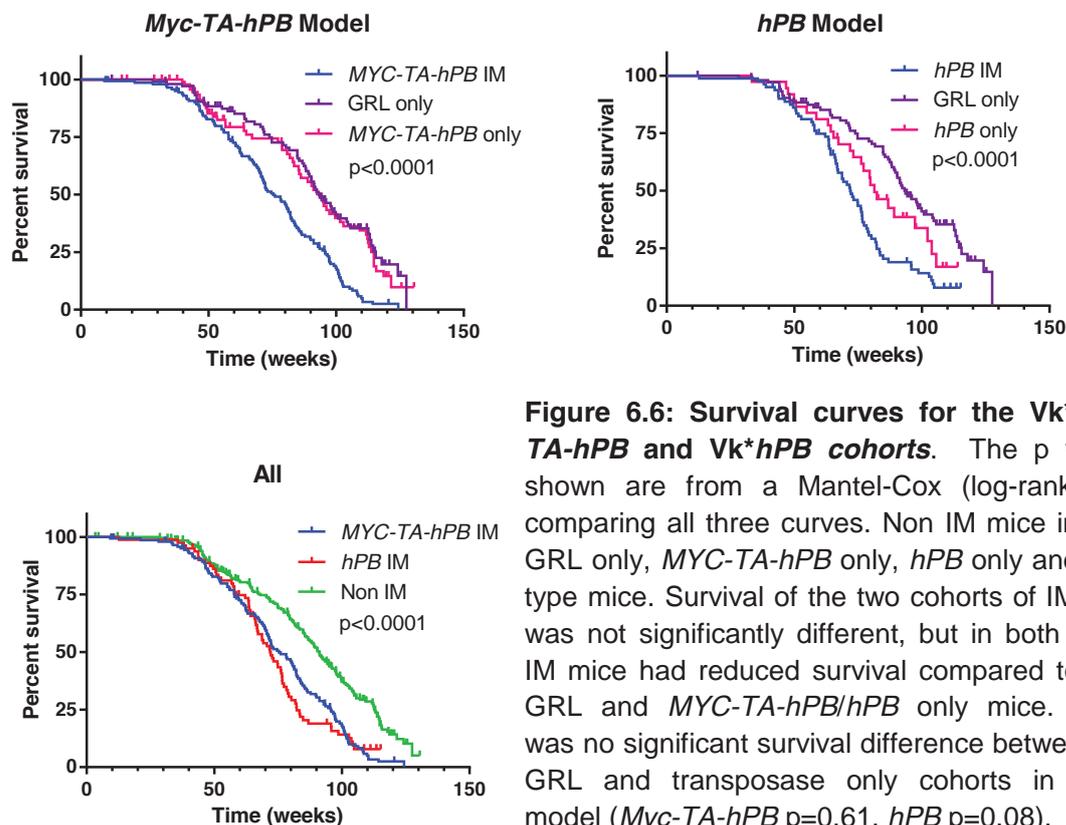


Figure 6.6: Survival curves for the $Vk^*MYC-TA-hPB$ and Vk^*hPB cohorts. The p values shown are from a Mantel-Cox (log-rank) test comparing all three curves. Non IM mice include GRL only, *MYC-TA-hPB* only, *hPB* only and wild-type mice. Survival of the two cohorts of IM mice was not significantly different, but in both cases IM mice had reduced survival compared to both GRL and *MYC-TA-hPB/hPB* only mice. There was no significant survival difference between the GRL and transposase only cohorts in either model (*Myc-TA-hPB* $p=0.61$, *hPB* $p=0.08$).

Histopathology assessment was performed on 170 of the first 184 mice in the $Vk^*MYC-TA-hPB$ cohort culled due to illness or found dead. Of the fourteen cases that were not reviewed, eight mice were found dead and considered too decomposed for useful histological analysis. Histopathology samples were collected in three cases but blocks and slides could not be located. No histopathology or

necropsy records were identified for two mice. One mouse was culled due to an ulcerated eye but had no other abnormality. Of the 170 cases reviewed 100 were insertional mutagenesis mice, 40 had *Vk**Myc-TA-hPB** only and 30 had *GRL* only. The spectrum of tumours is shown in figure 6.7.

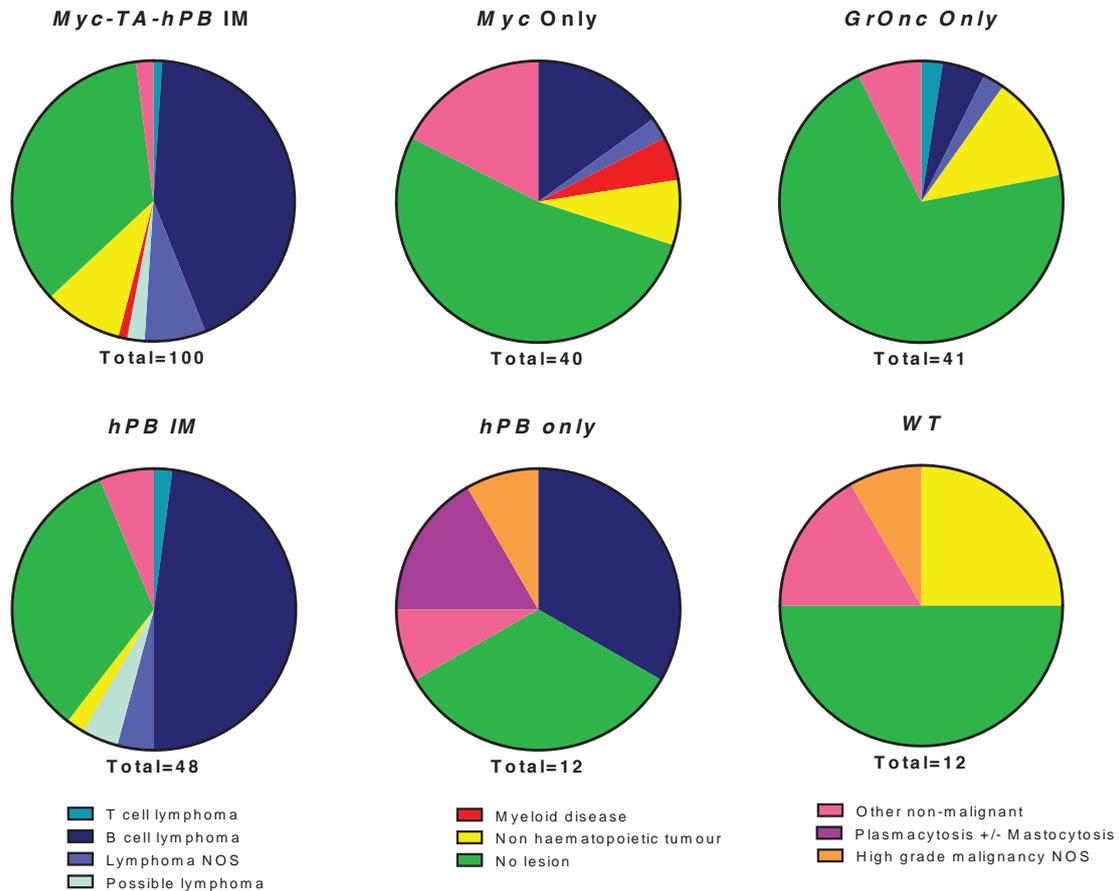


Figure 6.7: Diagnoses in the *Vk*MYC-TA-hPB* and *Vk*hPB* mice. The disease classification is based on independent review by a histopathologist (who was blinded to genotype) and immunophenotyping where this was performed. The *GRL* (*GrOnc*) only mice were pooled between the two cohorts. The diagnoses in twelve wild-type littermates which were aged along with the study animals are also shown. The *hPB* group includes one mouse that had a benign lesion only that was culled at 97.7 weeks for experimental reasons. Similarly, one of the *hPB IM* mice was culled for experimental reasons at 94.1 weeks of age but was found to have lymphoma.

In the *Vk*-hPB* cohort histopathology examination was performed on 69 of the first 96 mice to be culled sick or found dead and a further two mice which were culled for experimental reasons. Of the mice that did not have histopathology examination, two

were not necropsied (died aged 12.7 and 48.4 weeks). A further thirteen were found dead and considered too decomposed for useful histological analysis. The histopathology blocks and slides could not be found in four cases. In the remaining eight mice, histopathology samples were not recorded as received by our tissue bank.

Vk-Myc-TA-hPB mice*

Of the 100 *Myc-TA-hPB* IM mice on which histopathology was performed, 51 were reported to have lymphoma. Of these, six had additional non-haematopoietic tumours (carcinomas (4), sarcoma (1), probable neural or smooth muscle tumour(1)), two had focal increases in plasma cells in the bone marrow and one had a lymph node plasmacytosis of uncertain significance. Overall, there was no evidence of multiple myeloma in any of the *Vk*Myc-TA-hPB* IM mice.

Immunohistochemistry (IHC) staining for B220 and CD3 was performed on 46 of the mice with lymphoma and 43 were determined to have B cell lymphoma based on this staining. The other three lymphomas were of uncertain lineage. Flow cytometry was performed on spleen cells from two of these and one was found to be a T cell tumour, while in the other the spleen was not definitely involved with lymphoma and flow cytometry was unhelpful. The third tumour was reported to have a sclerotic pattern with giant cells on histopathology, more similar in morphology to human Hodgkin lymphoma. IHC was also performed on a further two mice that were called 'possible lymphoma' on initial histopathology review, but a definitive diagnosis could still not be made. Both of these mice were culled sick; one had a swollen abdomen with pale liver and splenomegaly (0.6g), the other was culled due to respiratory difficulties and had marginal thymomegaly with a spleen size of 0.34g.

Of the remaining 47 *Myc-TA-hPB* IM mice, only nine were diagnosed with other malignancies. These included myeloid leukaemia (1), squamous cell carcinoma (2), skin appendage carcinoma and hydronephrosis with probable carcinoma in the urinary tract (1), sarcoma (2) and papillary adenocarcinoma (3) of which one was also reported to have evidence of a myeloproliferative disease in the bone marrow. The remaining 38 mice were reported as having no (35) or benign lesions only (3). Of these, nine mice were found dead at ages of between 33 and 63.1 weeks. In one of these cases the spleen was enlarged (0.66g) and a large abdominal mass was

found at necropsy which is suggestive that this mouse also had lymphoma, however the tissue was too autolysed to allow a histopathological diagnosis. In the remaining eight mice that were found dead, the spleen was 0.1g or less and no masses or lymphadenopathy were identified at necropsy. The other 27 mice in which no abnormality was diagnosed on histopathology were culled due to illness. The commonest reason was a swollen abdomen (13 mice) but at necropsy the only abdominal finding was enlarged seminal vesicles in the majority of these cases. Some had a very full bladder and one had a distended bowel without an overt mass. The other mice were culled sick due to tachypnoea, severe scratch marks, piloerection, being hunched and inactive and one mouse each with limping and anal prolapse. Only one was reported as moribund by the animal technicians and this animal was found to have a left inguinal mass and splenomegaly (0.63g) at necropsy but no diagnosis could be reached on histopathology assessment. The spleen was less than 0.22g in weight in all of the remaining mice that were culled sick and reported as having 'no lesion' on histopathology. Figure 6.8 shows the spleen weight at death for the various cohorts. A bimodal distribution of spleen sizes across the *Vk*MYC-TA-hPB* cohort is evident, which largely represents cases with and without lymphoma.

Vk-hPB mice*

Histopathology assessment was performed on 48 IM mice from the *Vk*hPB* cohort, including one which was culled for experimental purposes and had features of lymphoma at death. Of these, 26 were found to have lymphoma on review of the H&E slides. Immunohistochemistry, flow cytometry or both was performed on 25 of these samples and the lineage was confirmed as B cell in 23 cases. In the most striking case of B cell lymphoma the spleen was 11.8g and there were also liver lesions, thymomegaly, and a mesenteric mass (figure 6.8 and 6.9). One mouse with B cell lymphoma was also noted to have a plasmacytosis in the thoracic and mesenteric lymph nodes and the liver. One of the remaining cases showed double staining for CD3 and B220 on IHC but the flow cytometry of the spleen suggested this was a T cell tumour (66% CD4+, 32% CD19+/B220+). Of the two lymphoma cases not assigned a lineage one was thought to be a B cell lymphoma based on morphology and distribution, however it has not been assessed by immunophenotyping. The flow cytometry of the other case is most suggestive of a T

cell tumour, but around 40% of spleen cells were negative for all markers used (CD3, CD19, B220, Mac1 and Gr1). There were two further mice which were reported as 'possible lymphoma', but this diagnosis was not confirmed even after flow cytometry and IHC. These mice both had enlarged spleens (2.07g and 1g) and one had widespread lymphadenopathy. Of the remaining IM cases, no lesion was identified in sixteen and none of these had macroscopic features of lymphoma. There were two cases with papillary adenomas of the lung and one with extensive invasive adenocarcinoma of the lung. The final case had unusual morphology with reactive and apoptotic changes and it was concluded this was most likely reactive.

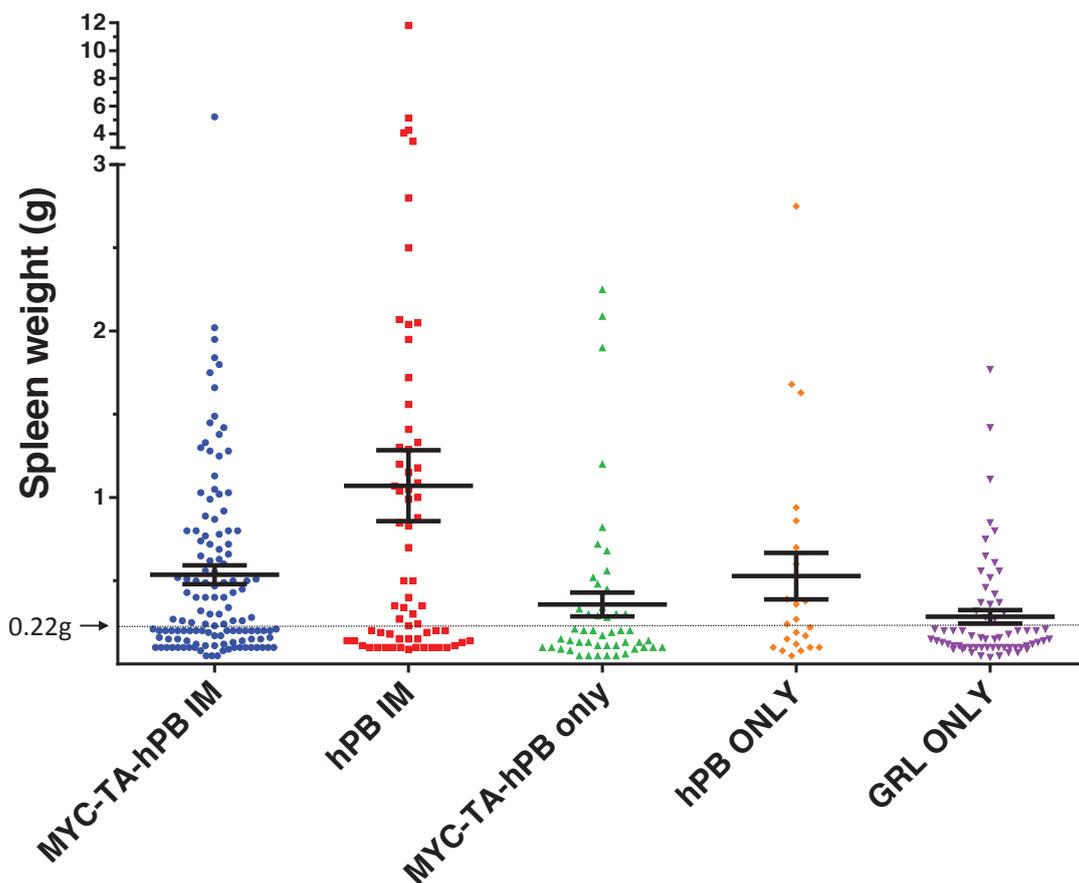


Figure 6.8: Spleen weight for the various colonies. Each mouse is represented by a point. The mean and the standard error of the mean for each colony are shown in black. The difference between *hPB* IM and *MYC-TA-hPB* IM is statistically significant ($p=0.0022$). For the *hPB* IM and *MYC-TA-hPB* IM mice the difference versus *GRL* only mice was statistically significant ($p=0.0004$ and $p=0.003$ respectively), but not quite so versus the *hPB* only ($p=0.15$) or *MYC-TA-hPB* only ($p=0.0785$) mice.

In total there were 41 GRL only (*GrOnc*) mice that had histopathology assessment between these two cohorts. Of these, no lesion was identified on histopathology in 29 cases, including one which was culled due to an eye defect and was not considered to have reached the survival endpoint. Of the cases in which no lesion was identified, nine were found dead. One of these had an enlarged spleen (0.52g) with no other findings at necropsy, but the tissues showed autolysis making it unsuitable for histopathology. Four of the remaining 12 *GrOnc* mice had lymphoma; two B cell, one T cell and one of uncertain lineage which was negative for B220 and CD3 on IHC but involved the thymus, heart, lung, kidney, spleen, liver, lymph nodes and blood. One of the mice with B cell lymphoma also had a focal increase in plasma cells in the BM and protein deposits in the glomeruli. Five of the GRL mice had non-haematopoietic tumours, one a lung adenoma and one fatty liver only. The final mouse had thoracic lymphadenopathy and splenomegaly, but these changes were thought to be reactive based on histopathology and IHC.

Of the 40 mice with the *Vk*MYC-TA-hPB* transgene but no *GrOnc*, only seven were found to have lymphoma. IHC was performed in six of these and confirmed B cell lymphoma in all cases. Two of these had focal areas of plasmacytosis in the bone marrow in addition to the lymphoma. No lesion was identified on histopathology in a further 21 mice and all had spleens that weighed less than 0.35g at death. Diagnoses in the remaining twelve mice were; i) probable myeloproliferative disorder (MPD) ii) chronic inflammation and immune complex glomerulopathy iii) groin abscess iv) periarteritis in lung and muscle v) lung adenoma only (n=3) vi) lung adenocarcinoma (n=3) vii) transitional cell papilloma of kidney and viii) possible myeloid leukaemia.

To date, histopathology examination has been performed on twelve mice with the *Vk*hPB* transposase alone. Of these four had B cell lymphoma, four had no lesion, and one each had plasmacytosis with mastocytosis, a skin lesion of uncertain aetiology with a plasmacytosis in a nearby lymph node, a high grade malignant mass in the abdomen of uncertain lineage and a lung adenoma. One of the mice with lymphoma was also noted to have plasmacytosis in the thoracic lymph nodes.

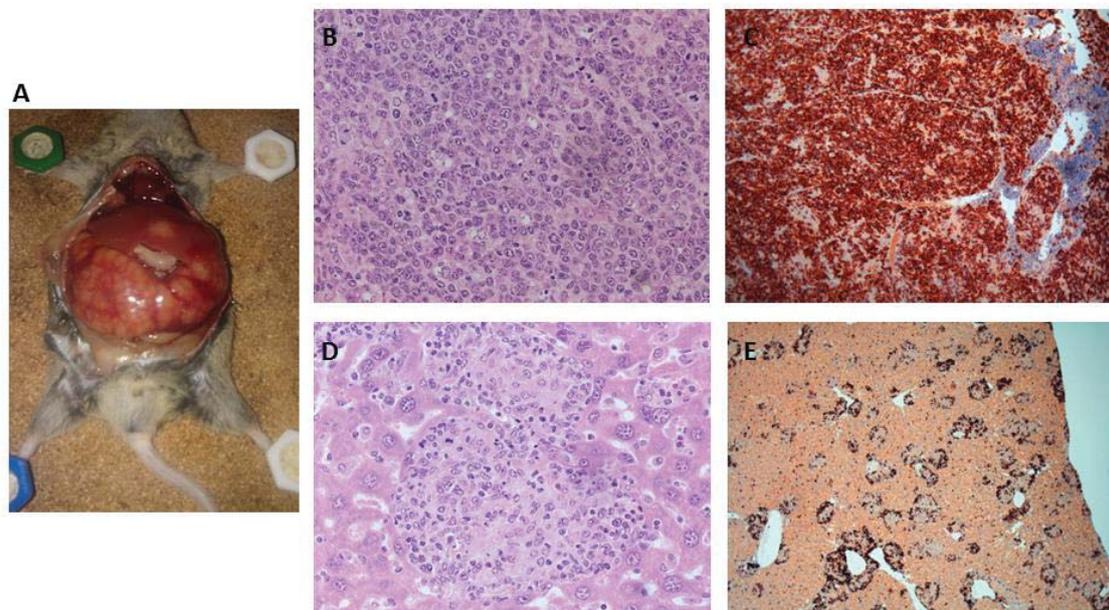


Figure 6.9: Mouse *Vk*HPB IM 4.1c*. (A) Necropsy findings included an 11.8g spleen and 4.75g liver. (B – E) Histopathology shows a diffuse infiltrate of B cells in the spleen [x200 H&E (B) and x 100 B220 IHC (C)] and peri-portal infiltrates in the liver [x400 H&E (D) and x 50 B220 IHC(E)].

Twelve wild-type littermates were aged along with these mice and tissues were taken for histopathology at death. Of these only one had a possible haematopoietic malignancy; an undifferentiated malignant infiltrate involving the liver, lymph node, lung, spleen, blood and kidney, which was negative for B220, CD3 and MPO on IHC.

6.2.6 Immunophenotyping to determine developmental stage of the B cell tumours

The histopathology and immunophenotyping confirmed an increase in B cell malignancies in the two insertional mutagenesis cohorts. However, the morphology of these lymphomas was variable and included follicular, diffuse and sclerotic tumours and some that varied in pattern between different regions of the same tumour. The cell size also varied from small to large cell both between and within tumours. Some lymphomas also had leukaemic changes, including some with blastic morphology.

With the assistance of Dr George Giotopolos (Huntly laboratory, University of Cambridge) I performed flow cytometry on CD19 and B220 positive tumours using CD24, CD43, AA4.1 (CD93), BP-1, IgM and IgD to determine what stage of B cell development these tumours correspond to (figure 6.10). Nine *Vk*MYC-TA-hPB* and 12 *Vk*hPB* IM tumours were selected for further analysis along with one *Vk*MYC-TA-hPB* only, one *Vk*hPB* only and two *GrOnc* only tumours. These included B cell lymphomas which were described on morphology as mature and immature, small and large cell and follicular, nodular and diffuse in pattern.

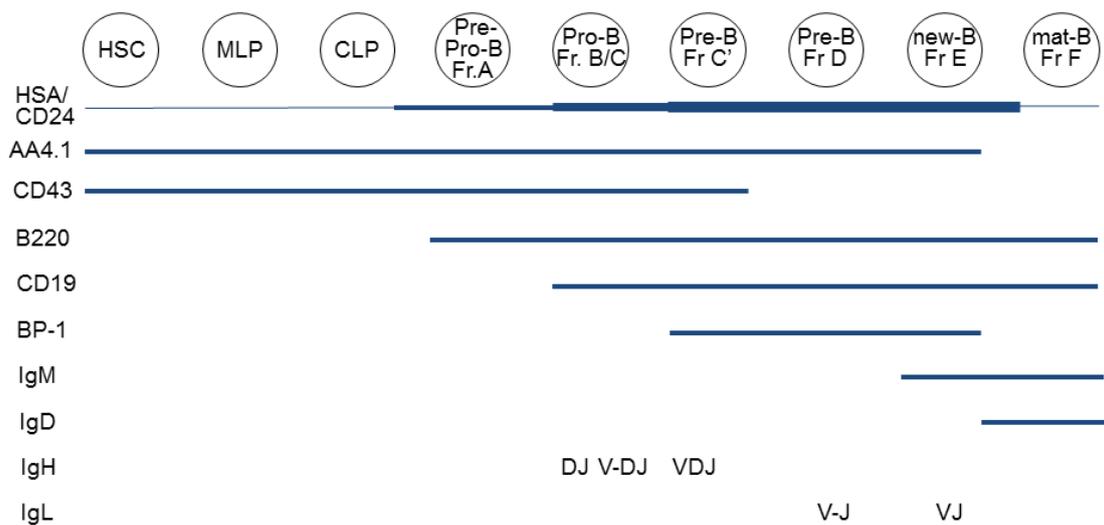


Figure 6.10: Stages of B cell development in mouse bone marrow. B cell development stages are shown according to the Hardy classification system (Hardy and Hayakawa, 2001). Corresponding surface antigens are indicated along with the timing of immunoglobulin gene re-arrangement for the heavy (IgH) and light (IgL) chain genes. Adapted from Hardy and Hayakawa, 2011.

Three samples described as ‘blastic’ or ‘poorly differentiated’ were selected for flow cytometry; two from the *Vk*MYC-TA-hPB* and one from the *Vk*hPB* IM cohorts. Of these, two were clearly positive for B220 and CD19, while one of the *Vk*MYC-TA-hPB* samples was positive for B220 but not for CD19 (figure 6.11A). Unfortunately, further flow cytometry could not be performed on the *Vk*MYC-TA-hPB* 12.4a and *Vk*MYC-TA-hPB* 10.1h samples due to a lack of viable cells. The other sample with blastic morphology *Vk*hPB* 5.1e was strongly positive for CD24, predominantly negative for BP-1 and showed variable staining for AA4.1, CD43, IgM and IgD. This

suggests that a significant proportion of the B cells in this tumour had a mature B cell phenotype (figure 6.11B).

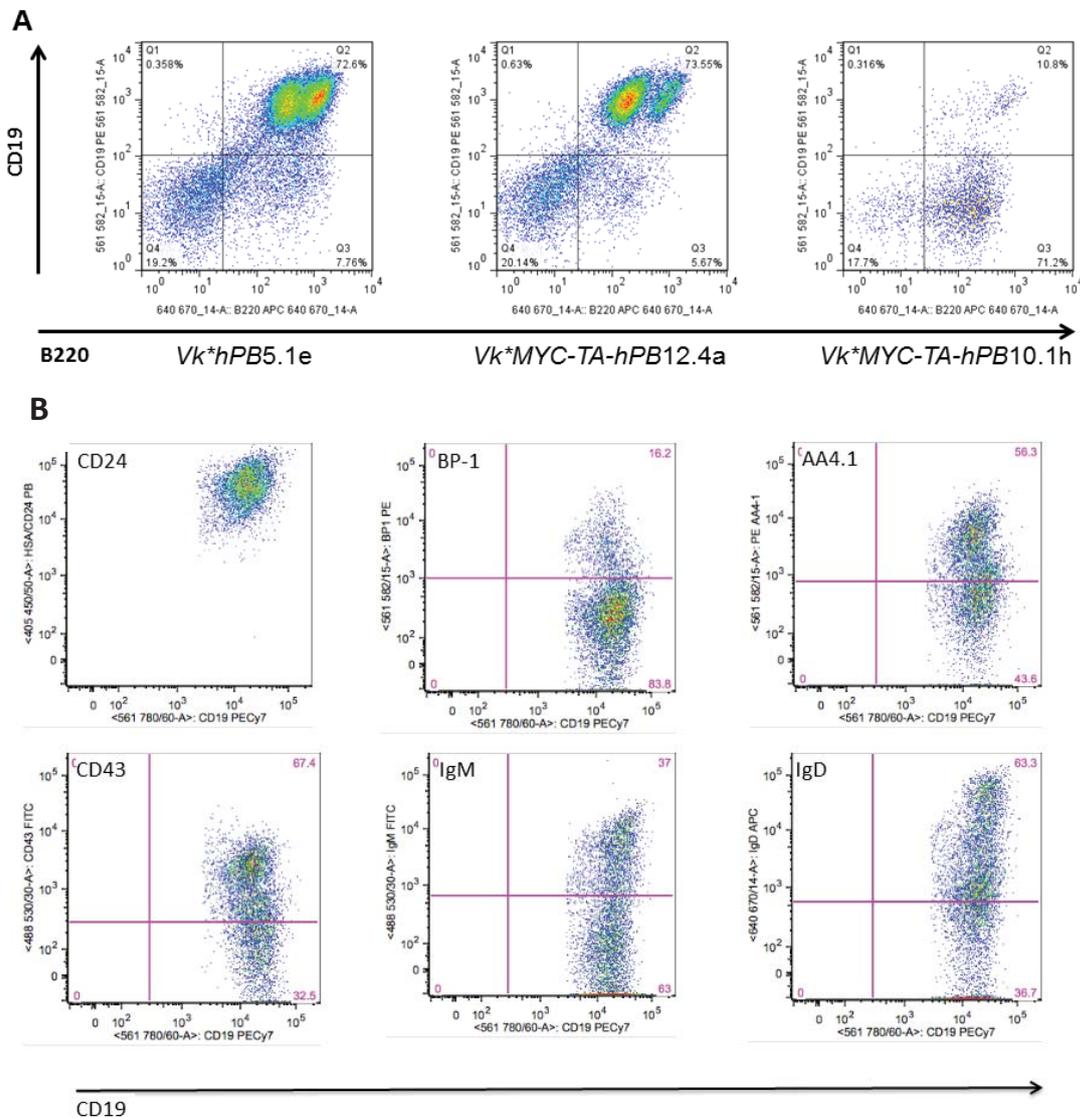


Figure 6.11: Flow cytometry of spleen samples from three mice with blastic morphology (A) CD19 (vertical axis) and B220 (horizontal) staining on the three mice (B) Further analysis of *VkhPB5.1e**. Clockwise from top left CD24, BP1, AA4.1, IgD, IgM and CD43 on the vertical axis; all shown against CD19 (horizontal).**

The nine *Vk*MYC-TA-hPB* IM and one *Vk*MYC-TA-hPB* only tumours investigated using the extended flow panel were uniformly negative for AA4.1 and BP-1 and positive for CD24. Representative flow cytometry results are shown in figure 6.12. IgM was positive in most cases and the IgD staining varied from weak to strong but was positive in all cases, with the exception of *Vk*MYC-TA-hPB* 10.4f (figure 6.13),

which was reported as a large cell high grade lymphoma on morphology. The CD43 was negative in most cases, including tumours classified as both small and large cell, and low and high grade on morphological appearance. However, in samples *Vk*MYC-TA-hPB* 14.1G, the *Vk*MYC-TA-hPB* only mouse, and *Vk*MYC-TA-hPB* 12.1C (figure 6.12), there was some weak CD43 staining. A third sample also had a small number of B cells with weak CD43, although the majority were negative. All of these tumours were described as follicular lymphomas on histopathology, although 12.1C was also noted to have sclerotic changes. Despite these differences, in general, the flow patterns were similar across the various morphological subtypes. Therefore, flow cytometry analysis showed that the majority of *Vk*MYC-TA-hPB* B cell malignancies had a mature B cell phenotype.

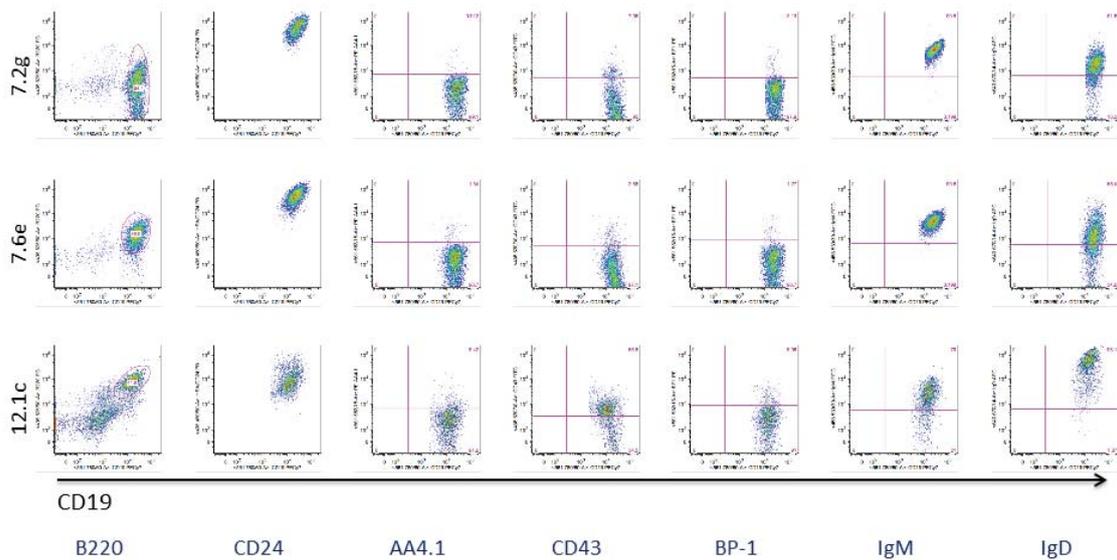


Figure 6.12: Representative flow cytometry from *MYC-TA-hPB* IM cases. The samples are *Vk*MYC-TA-hPB* 7.2g (top), *Vk*MYC-TA-hPB* 7.6e (middle) and *Vk*MYC-TA-hPB* 12.1c (bottom). Each tumour is represented in a separate row. From left to right the columns show; i) gating on B cells by B220 and CD19, ii) CD24, iii) AA4.1, iv) CD43, v) BP-1, vi) IgM and vii) IgD. CD19 is shown on the horizontal axis in all.

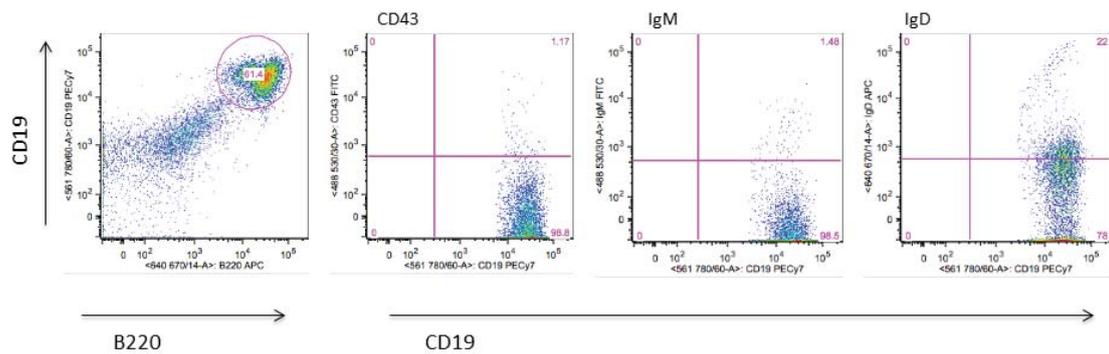


Figure 6.13: MYC-TA-hPB14.3f. Like most *Vk**MYC-TA-hPB samples 14.3f was negative for CD43, but In contrast to the other samples it was negative for IgM and had negative/weak IgD.

The flow cytometry of the twelve *Vk**hPB IM samples was more variable. Although these samples were uniformly positive for CD24, there was a spectrum of CD43 staining, with some samples clearly positive (e.g. *Vk**hPB 4.2b, *Vk**hPB 6.2l), others negative (e.g. *Vk**hPB 5.1d, hPB 7.7d) and most showing a mixture of positive and negative cells (figure 6.14). *Vk**hPB 4.2b and 6.2l were both uniformly positive for AA4.1 and 6.5c and 5.1e were positive in a significant proportion of cells, however all the other samples were negative. BP-1 was negative in the majority of B cells in all cases, but two had BP-1 positivity in up to 30% of cells (4.2b and 6.5c). Regardless of their CD43 expression, the *Vk**hPB IM tumours were mostly IgD positive. Even in samples 4.2b and 6.2l, a number of cells were IgD positive (figure 6.15). The fact that these tumours are positive for IgD suggests they have a mature B cell phenotype, but CD43 is usually negative from the late pro-B stage of development. Therefore these tumours are not easy to classify in the spectrum of normal B cell maturation (figure 6.10). The finding of CD43 with sIgM and IgD may reflect an aberrant phenotype or is consistent with these being B1 cells(Wells et al., 1994).

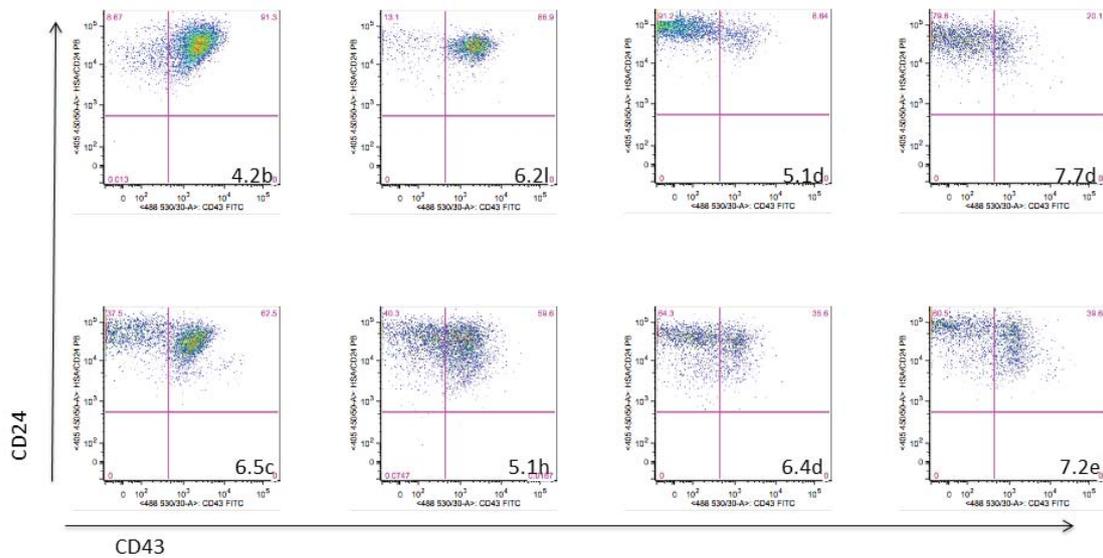


Figure 6.14: CD24 and CD43 flow cytometry on *hPB* IM samples. All analyses were performed on CD19 and B220 double positive cells. The mouse ID is indicated in each. Although all samples were CD24 positive there was significant variation in CD43.

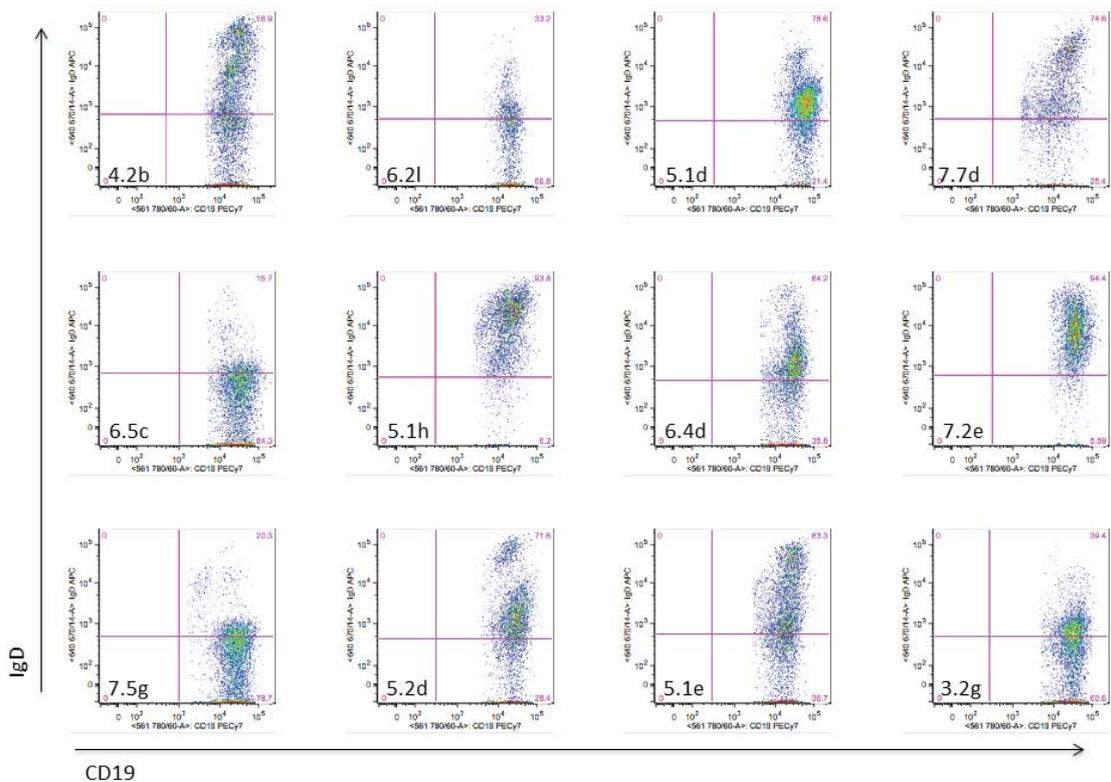


Figure 6.15: IgD expression in the *Vk* hPB* tumours. The data and mouse ID are shown for each of the twelve *Vk* hPB* IM mice analysed with the extended flow panel.

6.2.7 The MYC-TA-hPB tumours are not universally MYC dependent

In order to establish if the tumours in the *Vk*MYC-TA-hPB* cohort were MYC-dependent we performed Western blotting using various anti-Myc antibodies. This work was performed with Nicla Manes. Although both the human and mouse cMyc were detected by the appropriate antibodies in positive control lysates, neither was detected in the mouse samples despite abundant protein (figure 6.16). Western blotting for the *PB* transposase was also attempted using an in-house antibody, but this was unsuccessful.

Immunohistochemistry was performed using the same antibody as described in the Chesi paper, which detects both mouse and human Myc (Chesi et al., 2008). *MYC* was found to be expressed in many, but not in all of the *Vk*MYC-TA-hPB* IM tumours (figures 6.17 – 6.19). The *Vk*MYC-TA-hPB* IM samples which were described as having immature morphology typically had a high proportion of Myc positive cells (e.g. 10.1h, 12.4a and 3.4c) (figures 6.17 and 6.18). Many of the *GrOnc* only control mice or IM mouse from the *Vk*hPB* cohort did not stain strongly for Myc, but some of these tumours were positive (figures 6.20). Eight samples from *Vk*MYC-TA-hPB* mice that lacked the *GrOnc* transposon were also stained. Four of these had no lesion detected on histopathology and none of these had increased Myc staining. Of the four samples that were reported as lymphoma, only one had a noticeable increase in Myc staining (figure 6.21). Attempts to perform IHC using an antibody specific to human Myc were unsuccessful.

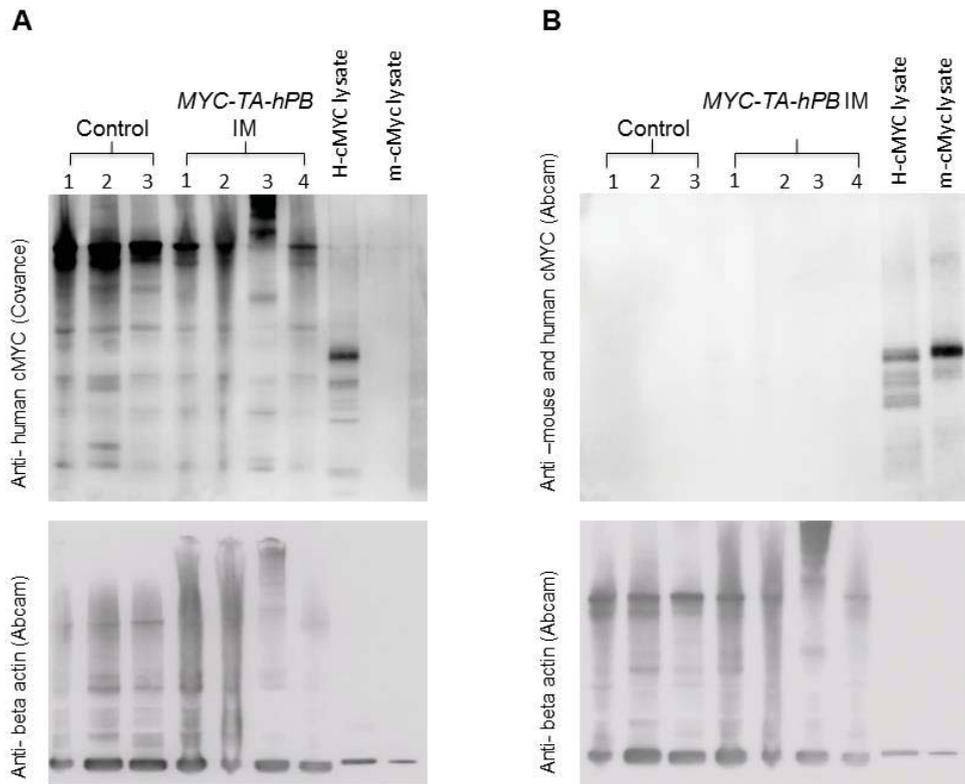


Figure 6.16: Western Blot for human and mouse Myc protein. (A) The Covance antibody is specific for human Myc. Despite a strong signal in the control lysate, no human MYC protein was detected in the transgenic IM mice. However, neither human or mouse MYC protein was detected with the non-specific Myc antibody in any of the tumour samples, although it was positive in control lysates (B). The bottom panels show the beta-actin staining, indicating adequate protein was present. The control samples were a lymphoma from a *Ras*^{G12D} GRL insertional mutagenesis mouse (1), and tumours from GrOnc only mice from this study (2 and 3). The *Vk***MYC-TA-hPB* IM lysates were all from spleen samples from mice with lymphoma, including three with follicular morphology (1-3) and one with blastic morphology (4). These blots were generated by Nicla Manes.

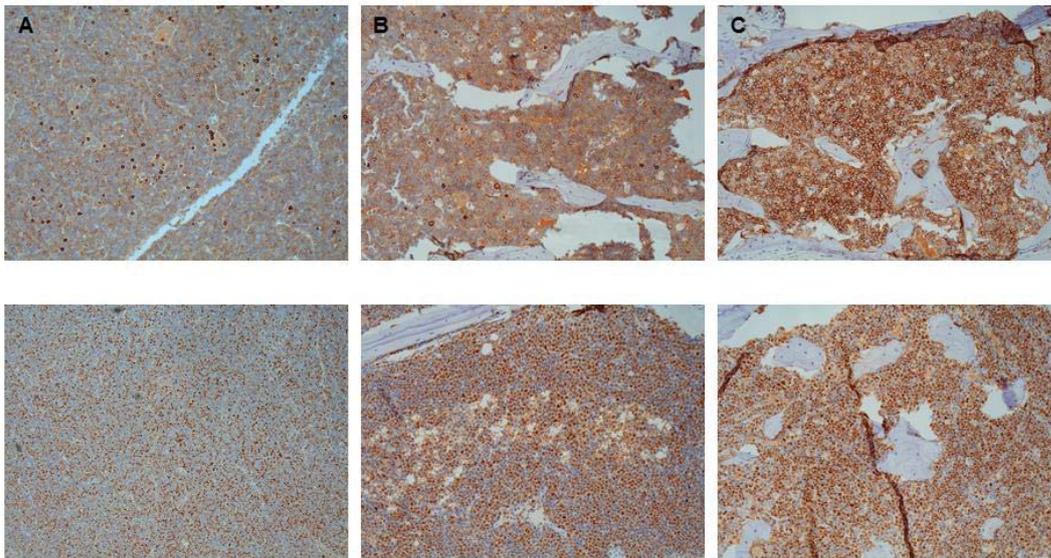


Figure 6.17: IHC staining of three *Vk*MYC-TA-hPB* IM tumours. The top row shows B220 staining and the bottom c-Myc. (A) Axillary lymph node from *Vk*MYC-TA-hPB* 10.1e, which has been described as a large cell lymphoma by our histopathologist. The majority of cells are positive for B220 and many are also positive for c-Myc. Bone marrow samples from *Vk*MYC-TA-hPB*10.1h (B) and *Vk*MYC-TA-hPB*3.4c (C) showing almost complete replacement of the marrow with B220 and c-Myc positive cells. Both of these tumours were described as blastic on morphological appearance.

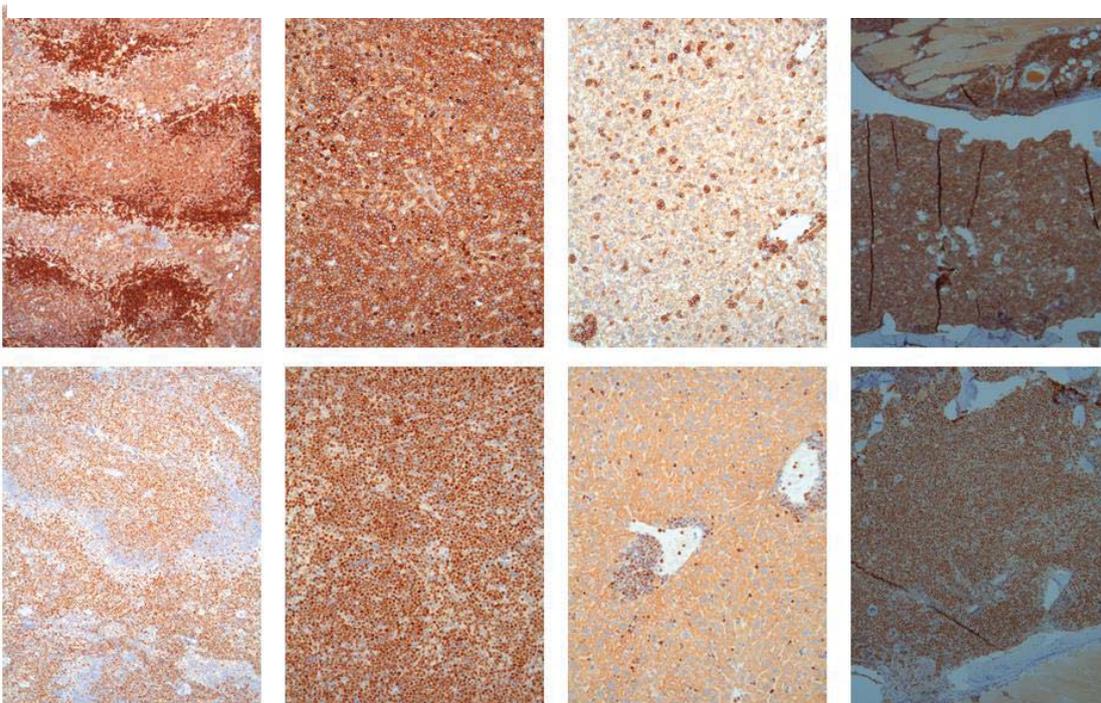


Figure 6.18: IHC of *Vk*MYC-TA-hPB* 12.4a, an IM tumour with blastic morphology. From left to right the tissues are spleen, lymph node, liver and bone marrow. Top panel: B220. Bottom panel: c-Myc.

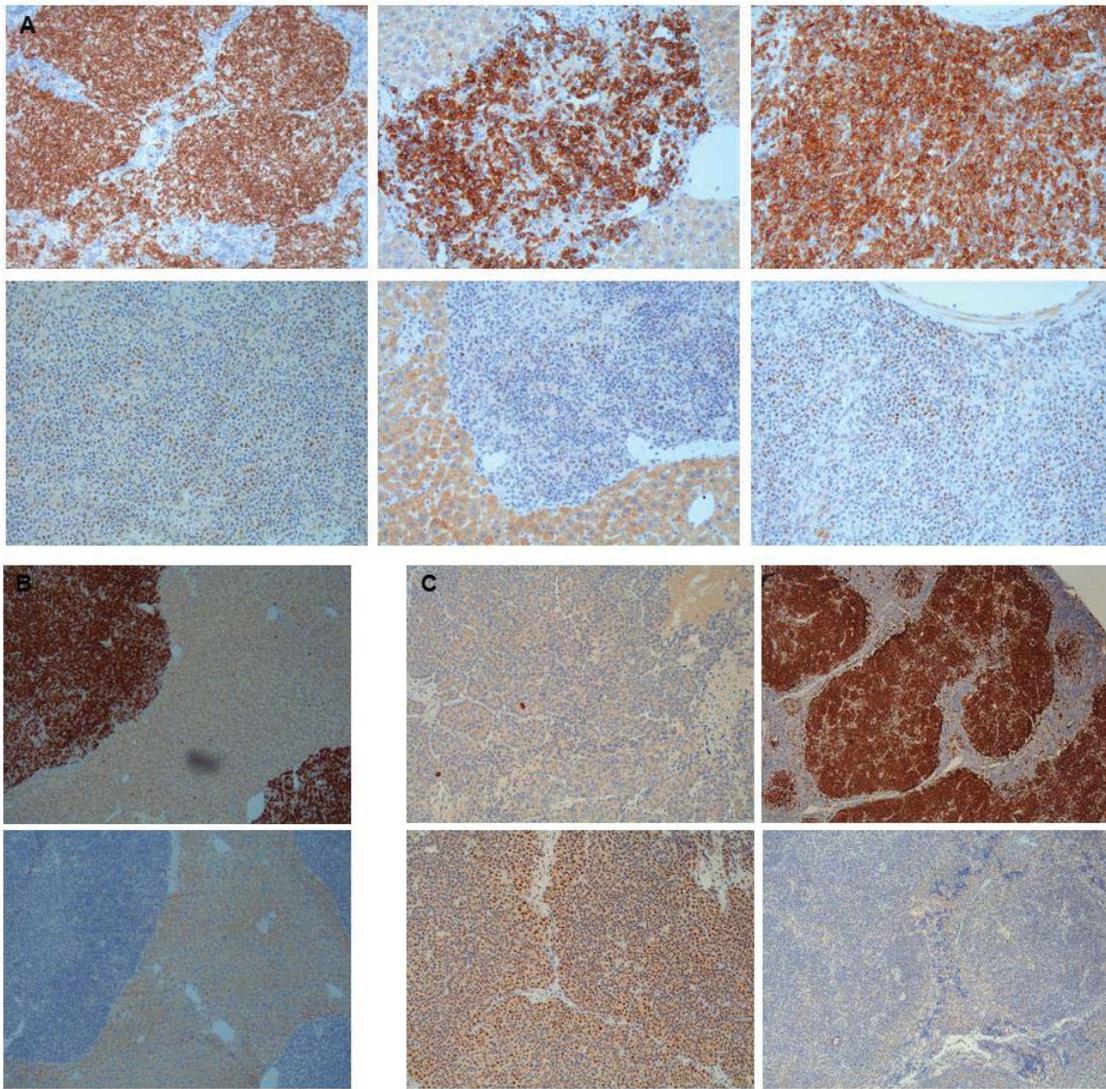


Figure 6.19: IHC of tumours from *Vk*MYC-TA-hPB* IM mice diagnosed with lymphoma, which were negative for *Myc*. B220 staining (top row) and c-Myc (bottom row) in each. **(A)** The malignant B cells in *Vk*MYC-TA-hPB14.1h* are negative for *Myc* in the spleen (left), liver (center) and mesenteric lymph node (right). **(B)** The findings were similar in *Vk*MYC-TA-hPB1.1b*. Liver immunohistochemistry is shown. **(C)** *Vk*MYC-TA-hPB12.3g* had an axillary tumour (left) (thought to be a carcinoma of a skin appendage or breast) in addition to a B cell lymphoma which involved the spleen (right). The axillary tumour is positive for *Myc*, but the B220 positive spleen cells are all negative.

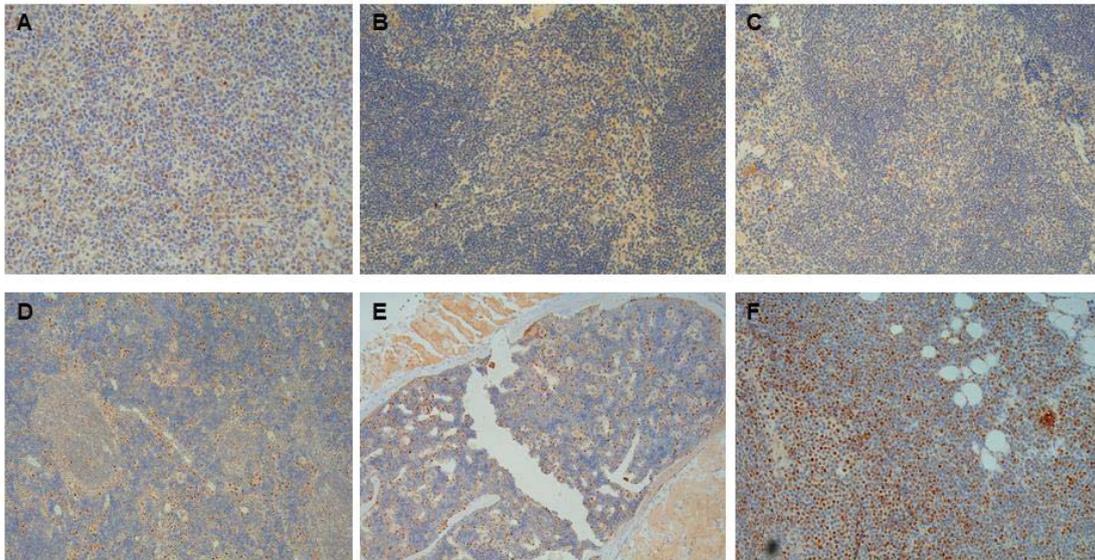


Figure 6.20: Myc IHC shows variable numbers of MYC positive cells in control mice. (A) Spleen sample from *Vk**hPB5.1a**, an IM mouse without the *MYC* transgene that developed lymphoma (x200). (B/C) Spleen samples from *GrOnc* only mice 4.1f (x200) and 2.5d (x100) showing only occasional Myc positive cells. (D) Spleen (x100) and (E) bone marrow (x100) from a wild type mouse showing scattered Myc positive cells. (F) A large number of Myc positive cells are seen in the bone marrow from *Vk**hPB5.1e** (x200), an IM mouse without the *MYC* transgene which developed lymphoma that was described as blastic in morphology.

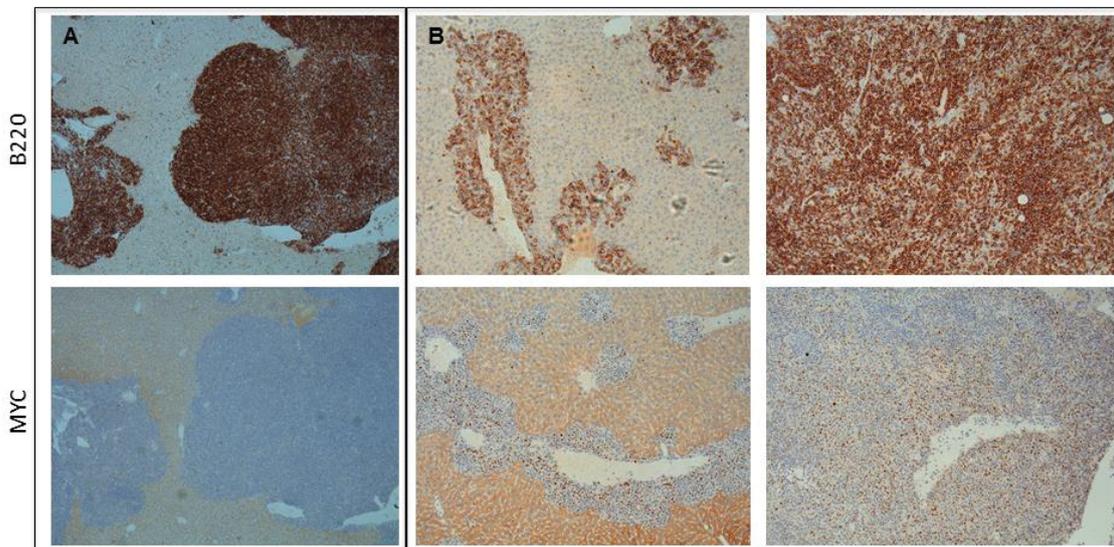


Figure 6.21: Myc IHC in *VkMYC-TA-hPB** mice that lacked the *GrOnc* transposon.** (A) Of the four tumours tested three had only occasional Myc positive cells as seen here in *Vk**MYC-TA-hPB14.1g**. (B) *Vk**MYC-TA-hPB10.1g** was the only one to have some increase in Myc staining. Shown is the IHC performed on the liver (left) and spleen (right). This was reported as a high grade large cell lymphoma. Top panel B220; bottom Myc.

6.2.8 Stop codon reversion was not seen in *Vk*hPB* and *Vk*MYC-TA-hPB* tumours

In order to investigate if the stop codon was reverted in the *hPB* and *MYC-TA-hPB* mice I performed RT-PCR followed by capillary sequencing on RNA from tumour samples in three *Vk*hPB* and four *Vk*MYC-TA-hPB* mice using the same primers as in 6.2.2, which were used to validate splicing from the *Vk* exon into *MYC* or *hPB*. The results showed clear sequence with maintenance of the stop codon in all tumour samples. Reversion of the stop codon in a minor sub-clone could not be excluded using this capillary sequencing approach. However, as the construct was designed such that activation of the transposase should have been dependent on reversion of the stop codon, we would expect this to be evident in the major tumour clone if it had occurred.

To further investigate reversion of the stop codon in tumour samples I performed the same RT-PCR in eight *Vk*hPB* and ten *Vk*MYC-TA-hPB* samples followed by sequencing of the PCR products using the MiSeq platform. The RNA was derived from tumour samples in 16 of 18 cases. There was one spleen sample from each IM cohort where the histopathology showed 'no lesion'. At least 350 000 reads were mapped to the target sequence in each sample, which represented between 67 and 86% of the total reads. Although there was evidence of mutation around the stop codon, this was always in a minority of reads (table 6.1 and 6.2). In every case at least 88% of the reads had the wildtype sequence which resulted in a protein sequence of ATMG-YPYDV around the stop codon. Typically this wild type sequence accounted for over 95% of reads (table 6.1).

<i>Vk*MYC-TA-hPB</i>	Annotated reads	Top hit (% of reads)	<i>Vk*hPB</i>	Annotated reads	Top hit (% of reads)
1.1b spleen	569878	95.37	3.1d spleen	670795	88.36
1.6e spleen	452051	97.10	3.2c spleen	845021	97.31
7.6e mass	490535	96.03	3.2g spleen	833492	97.45
8.1d spleen	361669	97.53	5.2d spleen	633469	97.50
10.1g	435312	94.84	6.2b spleen	865972	97.37
10.4f spleen	371716	96.79	6.2l spleen	531630	96.76
12.3g	399241	96.87	7.2e spleen	754682	97.36
13.2f spleen	350713	96.54	7.7d spleen	899094	97.18
14.3f spleen	542318	96.85			
14.3f spleen	475863	96.77			
14.5f spleen	408151	96.89			
14.5f spleen	381979	96.79			

Table 6.1: Total reads around the stop codon in each sample from sequencing the RT-PCR products on the MiSeq platform. In each case the top hit was the unmutated sequence and this accounted for at least 94% of reads in all but one case.

6.2.9 The *hPB* and *MYC-TA-hPB* IM tumours are clonal and have undergone somatic hypermutation

In order to determine if these tumours were clonally re-arranged and had undergone SHM, BCR repertoire analysis was performed by Rachael Bashford-Rogers using the method she has established (Bashford-Rogers et al., 2013b). This analysis was performed on DNA from spleen, lymph node or abdominal mass in samples from five *Vk*hPB* and five *Vk*MYC-TA-hPB* IM mice. All samples except one returned more than 21000 reads that passed filtering, which is sufficient to derive robust B cell repertoires. The sample that failed was *Vk*MYC-TA-hPB1.1b*, which returned only 635 reads after filtering (16% of the all reads with its unique barcode). The DNA from this sample was taken from an abdominal mass, which turned out to be a non-B cell tumour, with morphology consistent with a neural or smooth muscle tumour. The mouse also developed lymphoma in the spleen and liver, however the BCR repertoire analysis was not performed on these samples. The remaining samples gave a mean of 130732 reads post filtering (range 21123-239144).

1.1b spleen	1.6e spleen	7.6e mass	8.1d spleen	10.1g	10.4f spleen	12.3g	13.2f spleen	14.3f spleen	14.5f spleen
B cell lymphoma	Large cell lymphoma with low mitotic rate	Small cell lymphoma	No lesion	Large cell lymphoma	Large cell high grade lymphoma	Lymphoma	Follicular lymphoma	Follicular large cell lymphoma	Blastic
ATMG-YPYDV	ATMG-YPYDV	ATMG-YPYDV	ATMG-YPYDV	ATMG-YPYDV	ATMG-YPYDV	ATMG-YPYDV	ATMG-YPYDV	ATMG-YPYDV	ATMG-YPYDV
543519	438951	471083	352739	412840	359781	386745	338589	525218	395458
1847	1259	1058	1240	1004	2113	826	559	1257	704
53938	1136	979	817	853	744	677	409	1055	701
4154	905	979	560	775	677	635	359	989	636
2434	865	726	528	618	562	576	320	958	590
1520	829	716	506	563	554	534	309	840	497
1255	596	666	396	539	532	512	300	810	454
1224	586	646	366	525	506	480	246	662	454
961	547	599	357	473	477	448	241	570	452
864	540	556	320	372	384	360	229	515	373
856									
Lymphoma?	3.2c spleen	3.2g spleen	5.2d spleen	6.2b spleen	6.2i spleen	7.2e spleen	7.7d spleen		
ATMG-YPYDV	Pleomorphic B cell lymphoma	Follicular lymphoma	B cell lymphoma	No lesion	Burkitt like lymphoma	Large cell high grade	Follicular large cell high grade		
592695	822303	812238	617610	843166	514425	734725	873706		
53938	1808	1890	1348	1883	2209	1759	2021		
4154	1777	1758	1210	1727	1274	1576	1739		
2434	1769	1469	1198	1431	1202	1567	1723		
1520	1375	1239	1008	1316	1088	1330	1347		
1255	1354	1233	996	1272	918	1239	1304		
1224	1244	1205	980	1250	820	1056	1288		
961	1194	1161	953	1178	772	994	1224		
864	1009	1121	696	1170	754	969	1081		
856	823	969	648	893	747	923	1054		

Table 6.2: Results from RT-PCR and sequencing around the stop codon in 10 *Vk*MYC-TA-hPB* and 8 *Vk*hPB* samples. The diagnosis is shown in each case along with the protein translation of the major sequences identified using MiSeq sequencing. Although some sequences were identified in which the stop codon is reverted, these were always in a minority of reads. The bioinformatic analysis was performed by Rachael Bashford-Rogers.

The *Vk*hPB* samples included two controls that had no clinical features of lymphoma at death and no identifiable abnormality on histopathology. The spleen DNA from these samples showed diverse B cell repertoires without evidence of clonal expansion (figure 6.22). The largest cluster in the samples without an identified tumour was 2.1%. By contrast, the most monoclonal of the samples was *Vk*MYC-TA-hPB10.1h*, the sample which was positive for B220 but negative for CD19 on flow cytometry (figure 6.11a) and was positive for Myc on IHC (figure 6.17). In this sample 95% of the B cell repertoire was represented by a single cluster. This cluster shared the V gene IGHV14-2*02 and the J gene IGHJ4*01 and there were also an average of 7.94 mutations in the V gene, indicating the B cells had undergone SHM.

Of the six other tumour samples that were analysed four had a dominant cluster that accounted for at least 50% of the repertoire (table 6.1). In sample *Vk*hPB3.3a* there were two clusters that accounted for 16% and 10% of the repertoire and this was reported as a diffuse large cell lymphoma based on H&E stained tissue sections. However, IHC showed double staining for CD3 and B220 and on flow cytometry only 33% of viable cells were CD19 and B220 positive, while 66% were CD4 positive. On the basis of the flow cytometry results this was called a T cell tumour, but these results indicate there was also a clonal expansion of B cells. In the remaining sample analysed for B cell repertoires (*Vk*MYC-TA-hPB13.2f*) the largest two clusters accounted for 4% of reads each. This sample was reported as follicular lymphoma based on the H&E stained specimens, but IHC was not performed. Flow cytometry on the spleen showed 42% B cells and 35% T cells and no definite lineage was assigned to this tumour. The B cell repertoire results on the spleen are consistent with a small clonal expansion of B cells. Repeat analysis was subsequently performed using BM RNA from the same mouse and the largest cluster identified was 18%, which provides further evidence that this mouse did have a clonal B cell disorder.

It is important to note that as our main analysis was performed on DNA it is possible that two clusters relate to a single B cell clone, with one representing an abortive non-functional rearrangement. This is less likely when the clusters are clearly different in size such as for *Vk*hPB 6.4d*, than in cases such as *Vk*hPB 6.5c*. However, on flow cytometry of *Vk*hPB6.5c* two distinct B cell populations were

evident (figure 6.23) and it is also possible that these clusters came from separate B cell clones.

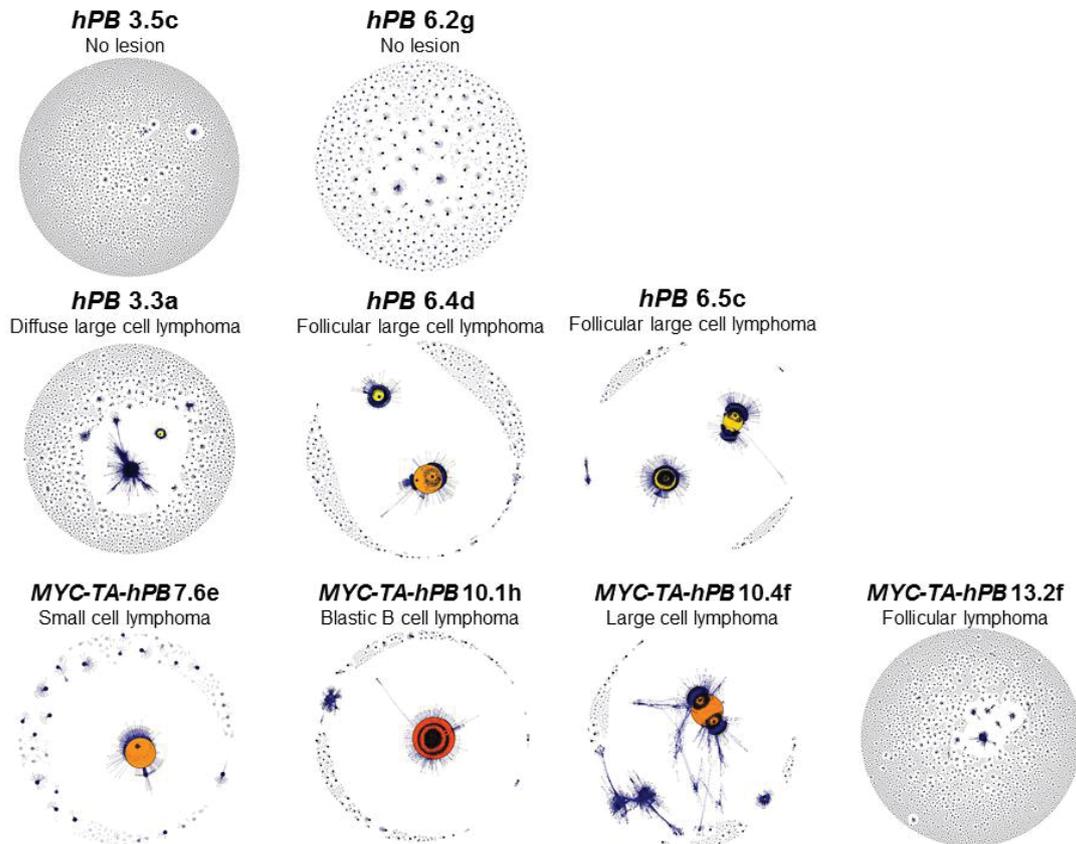


Figure 6.22: B cell receptor repertoires from different samples. Each vertex represents a unique sequence and the relative size of the vertex is proportional to the number of reads that shared this identical sequence. Vertices that differ by one base are connected by edges and groups of connected vertices are grouped into clusters. The colour of the major vertex indicates the proportion of reads that share that sequence (yellow <40%, orange 40-90%, red>90%)(Bashford-Rogers et al., 2013b). The sample ID and histopathology diagnosis are indicated for each sample. The total sequence reads and proportion assigned to the major clusters are shown in table 6.3. Plots courtesy of R. Bashford-Rogers.

Sample ID	Cluster ID	% of Repertoire	Number of Reads	Number of Vertices	V gene	J gene	Mean mutations in V
<i>Vk*hPB</i> 3.5c spleen	1	1.05	1753	235	IGHV3-6*01	IGHJ4*01	12.00
<i>Vk*hPB</i> 6.2g spleen	1	1.93	706	155	IGHV5-15*01	IGHJ1*03	1.37
	2	2.12	773	152	IGHV5-9-1*02	IGHJ4*01	1.18
	3	1.20	439	77	IGHV3-1*02	IGHJ4*01	5.06
	4	1.06	386	72	IGHV3-1*02	IGHJ4*01	6.11
	5	1.20	438	68	IGHV15-2*02	IGHJ4*01	1.03
	9	1.16	424	58	IGHV3-1*02	IGHJ4*01	5.21
<i>Vk*MYC-TA-hPB</i> 13.2f spleen	1	3.83	5901	745	IGHV11-2*01	IGHJ1*03	1.44
	2	3.99	6148	421	IGHV3-1*02	IGHJ2*01	5.92
	3	1.03	1582	202	IGHV11-2*01	IGHJ4*01	4.89
<i>Vk*hPB</i> 3.3a spleen	1	10.04	12740	3463	IGHV1-15*01	32	31.00
	2	15.57	19746	1126	IGHV2-4-1*01	IGHJ3*01	6.33
<i>Vk*hPB</i> 6.5c spleen	1	51.44	123026	6061	IGHV5-4*03	IGHJ3*01	1.93
	2	46.15	110354	5791	IGHV1-55*01	IGHJ3*01	5.70
<i>Vk*hPB</i> 6.4d spleen	1	63.76	88237	4352	IGHV15-2*01	IGHJ1*03	1.63
	2	27.92	38637	2191	IGHV14-2*02	IGHJ3*01	7.60
<i>Vk*MYC-TA-hPB</i> 7.6e mesenteric mass	1	68.74	14519	1220	IGHV1-15*01	IGHJ2*01	3.38
	2	2.82	596	90	IGHV3-1*02	IGHJ4*01	7.13
	3	2.54	537	79	IGHV3-3*01	IGHJ4*01	1.23
	4	1.98	419	75	IGHV12-3*01	IGHJ1*03	1.19
	5	2.25	476	66	IGHV3-6*01	IGHJ1*03	1.12
	6	1.28	270	49	IGHV3-1*02	IGHJ3*01	11.94
	7	2.27	479	47	IGHV11-2*01	IGHJ1*03	0.91
	8	1.35	285	47	IGHV11-2*01	IGHJ4*01	4.00
	9	1.12	236	45	IGHV5-6-2*01	IGHJ4*01	41.36
	10	1.30	274	37	IGHV12-3*01	IGHJ3*01	2.92
	12	1.08	228	30	IGHV5-2*02	IGHJ4*01	1.17
	<i>Vk*MYC-TA-hPB</i> 10.4f spleen	1	82.34	72794	5986	IGHV1-53*01	IGHJ2*01
2		7.85	6940	1757	IGHV1-53*01	IGHJ2*01	27.32
3		4.54	4013	470	IGHV12-3*01	IGHJ1*03	1.35
<i>Vk*MYC-TA-hPB</i> 10.1h LN	1	94.95	193947	7615	IGHV14-2*02	IGHJ4*01	7.94

Table 6.3: Cluster assignment and B cell repertoires for samples depicted in Figure 6.22. All clusters which accounted for at least 1% of the reads for a sample are shown. The V and J genes assigned to the cluster and the mean number of mutations in the variable region of sequences in that cluster are shown.

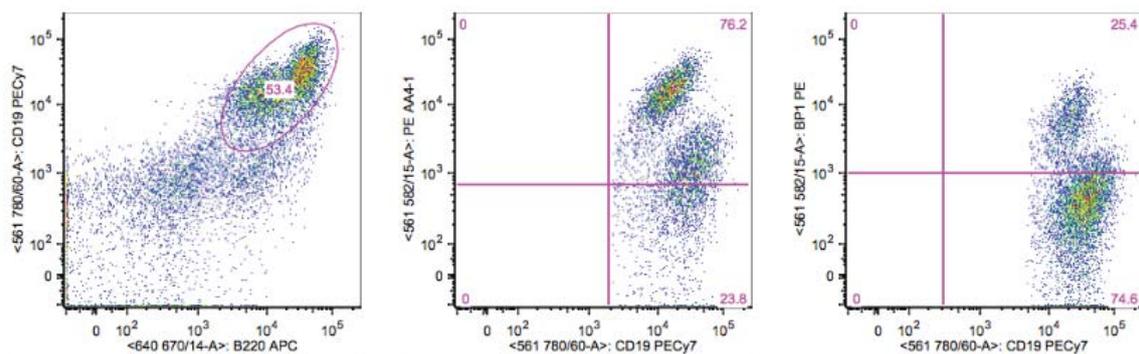


Figure 6.23: Flow cytometry on *hPB* 6.5c. Two distinct B cell populations are evident.

6.2.10 Serum protein electrophoresis of *Vk*MYC-TA-hPB* and *Vk*hPB* mice

Serum samples were collected at monthly intervals from a cohort of mice to look for the development of a paraprotein. In addition, where possible, a serum sample was taken from all mice at the time of death. The rate of detection of monoclonal proteins was lower than that reported by Chesi et al who found that over 50% of *Vk*MYC* mice, but less than 10% of controls had a detectable paraprotein at 30 weeks of age (Chesi et al., 2008). In that study monoclonal proteins were first detectable from 20 weeks of age and increased in intensity over time. By 50 weeks 80% of *Vk*MYC* mice had a monoclonal band (Chesi et al., 2008). Although I detected occasional monoclonal bands these did not always increase in intensity and were sometimes lost on subsequent sampling (figure 6.24). For example *Vk*MYC-TA-hPB* 10.1b had a clear monoclonal band on the week 37 serum sample, but it was negative when the mouse died of lymphoma at 84.5 weeks of age.

Protein electrophoresis was performed on blood samples taken at death from 49 *Vk*MYC-TA-hPB* mice (all genotypes) of which only eleven had a definite monoclonal band. Of these, no diagnosis of malignancy was made based on histopathology in four, while the other seven were found to have lymphoma. Only one was found to have a bone marrow plasmacytosis on histopathology examination and this mouse also had lymphoma. Of the 17 *Vk*MYC-TA-hPB* only mice (i.e. no IM), which had protein electrophoresis performed on the final blood sample, four had a clearly detectable paraprotein of which one had a bone marrow plasmacytosis on histopathology. Of 16 *GRL* only or WT mice, four had a clear paraprotein at death.

Protein electrophoresis was performed on serum samples taken at death from seventeen *Vk**hPB** mice (including two non-IM) and six had a paraprotein detected. The one IM mouse reported to have a plasmacytosis in the lymph nodes and liver on histopathology had an increase in gamma globulins but no discrete monoclonal band.

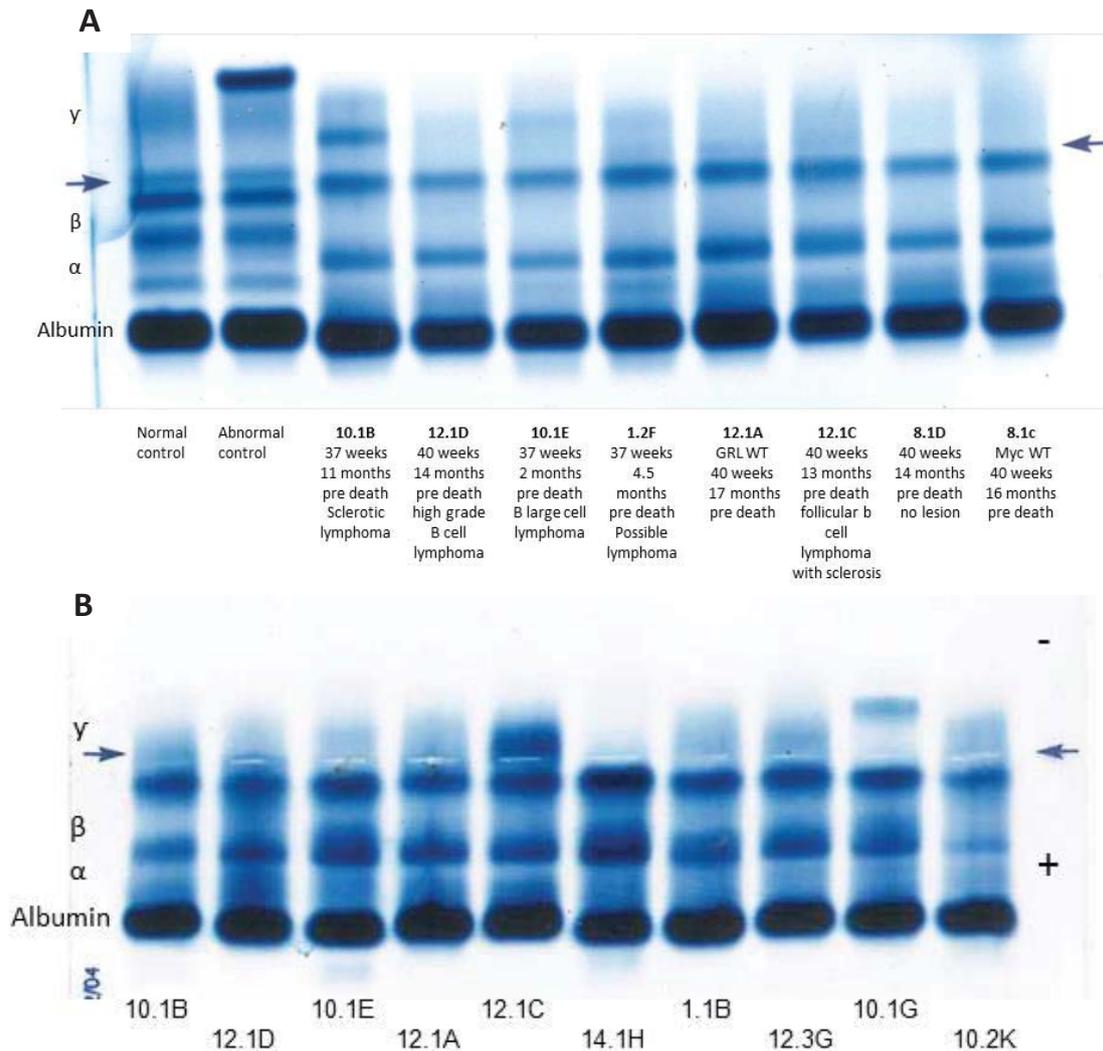


Figure 6.24: Serum protein electrophoresis on *Vk*MYC-TA-hPB* mice. A: Week 37-40 results from a selection of live mice which were serially bled. **B:** Results from sera obtained at death from ten mice including five mice from A (left). In these terminal blood samples only 10.1G has a clear monoclonal band. This mouse was a *Vk*MYC-TA-hPB* only mouse (No IM) that died with large cell high grade lymphoma. Mouse 12.1C had a lymph node plasmacytosis on histopathology and shows increased gamma globulins without a discrete band. 10.1B, 14.1H, 1.1b, 12.3G and 10.2K were all IM mice that died with lymphoma. Note that a monoclonal band was evident in mouse 10.1B at week 37 (as well as in several other serial samples). This was lost at death when the mouse had lymphoma. The arrows indicate the point of sample application and the position of α , β and γ globulins are indicated. The controls are Kemtrol Normal and abnormal serum controls.

6.2.11 Common integration site analysis identifies known and novel lymphoma genes

CIS were analysed using the top 10, 25 and 100 integrations from each tumour sample. Samples included in the analysis were from IM mice which had been diagnosed with lymphoma based on histopathology analysis or, where histopathology results were not available, from mice with necropsy findings which were suggestive of lymphoma (splenomegaly ($\geq 0.4g$) and/or significant lymphadenopathy). Only a single tissue sample from each mouse was used in the analysis and non-lymphoid tumours were excluded. The CIS identified using the top 25 hits in each cohort are shown in tables 6.3 and 6.4. Similar lists for the top 10 and 100 hit analysis are shown in appendix 6A.

The CIS identified by CIMPL analysis showed significant overlap between the *Vk**hPB** and *Vk*MYC-TA-hPB* IM cohorts (figure 6.25). The overlapping integrations between the two cohorts using the top 10 hits from each tumour included five of the top hits in the *Vk*MYC-TA-hPB* cohort and four of the top six in the *Vk**hPB** cohort. The number of CIS sites identified increased with the number of integrations from each sample used in the analysis (figure 6.26). The most frequently hit CIS were detected regardless of the cut-off used and there were 26 CIS integrations found in all three analyses on *Vk*MYC-TA-hPB* cohort and 22 for the *Vk**hPB** tumours (figure 6.26). The samples that harboured integrations contributing to these CIS are shown in tables 6.5 and 6.6.

Chr	Start	End	Genes in CIS	Gene nearest to peak	Number insertions	Number insertions (no hop)	p value	Scale
13	37608797	38048760	Rreb1 Ssr1 Cage1 Riok1	Rreb1	35	25	0	10-100
16	23824262	24334780	Sst Rtp2 Bcl6	Bcl6	19	15	0	10-100
4	44552079	44884132	Pax5 Mir5120 Gm12462 Gm12463 Zcchc7 Gm22639	Gm12463	8	7	0	10-30, 50-100
4	32211207	32533941	Bach2 D130062J21Rik Gm11932 Gm24371	Bach2	9	6	0	10-40, 60-100
11	44649490	44834554	Ebf1 Gm12158	Ebf1	6	6	4.16E-11	40, 60, 80-100
1	58595725	58840495	Ndufb3 Gm10068 Als2cr12 Cflar Casp8	Cflar	5	5	0	10-100
7	24973525	25245544	Atp1a3 Grik5 Zfp574 Pou2f2 D930028M14Rik Dedd2 Zfp526 Gsk3a 9130221H12Rik Erf	Pou2f2	5	5	0	10-100
11	62337694	62503277	Ncor1 Plgl Gm12278	Ncor1	6	4	0	10-60, 80, 100
2	18620014	18835286	Gm13355 Gm13352 Commd3 Bmi1 Gm13334 BC061194 Gm20539 Gm13333 RP23-396N6.8	BC061194	4	4	0	50, 70, 100
6	98999663	99220090	Foxp1	Foxp1	4	4	0	50-100
18	65351483	65529809	Alpk2 Gm22567 Malt1 Gm26114	Malt1	4	3	0	10-100
18	80544846	80723172	Nfatc1	Nfatc1	4	3	0	10-100
X	52821588	53000206	Gm14607 Gm6539 Rps2-ps13 Phf6 Hprt	Phf6	4	3	0	10-100
1	9647221	9732895	Mybl1 Vcpip1	Mybl1	3	3	0	10, 100
2	170043831	170229747	Tshz2 AL731822.1 Zfp217 AL844576.1	Zfp217	3	3	0	10-100
6	129115904	129259023	Klrb1-ps1 Gm26180 Clec2d AC142191.1	Clec2d	3	3	3.106E-12	10, 20, 40, 60-80, 100
10	18920536	19091825	Tnfrsf3	Tnfrsf3	3	3	0	10-100
11	20037469	20142449	Actr2	Actr2	3	3	0	10, 30-70
17	47684419	47814375	Frs3 Gm14873 Pgc Tfeb	Tfeb	3	3	0	40, 50, 70-100
18	60673185	60937918	Ndst1 Rps14 Gm8731 Cd74 Mir5107 Tcof1 Arsi Camk2a	Cd74	3	3	0	10-100
5	147333730	147365844	Flt3	Flt3	3	2	0	10-30, 50, 60, 90
14	115020837	115042641	intergenic	Mir17hg	3	2	0	20, 30, 70, 90, 100
2	98663779	98667685	Gm10801 Gm10800	Gm10800	2	2	0	10
2	104072250	104098503	Cd59b	Cd59b	2	2	0	10, 20, 40, 60, 80, 90
2	163223462	163288022	Tox2	Tox2	2	2	0	10, 30, 40, 60, 80, 90
2	163598659	163692531	Ttpal Serinc3 0610039K10Rik Pkig Gm16316	0610039K10Rik	2	2	1.792E-09	80
2	165953289	166047161	Gm11462 Gm11463 Gm11464 Ncoa3	Ncoa3	2	2	0	20, 40-60, 80, 90
3	101227944	101368445	Gm12490 Cd2 Gm10355	Cd2	2	2	0	10-100
4	138035572	138067992	Elf4g3	Elf4g3	2	2	0	10-40, 60, 70, 90
5	24779372	24802663	intergenic	Rheb	2	2	0.0002003	80
7	80139008	80250546	Gm24012 Mir1965 Sema4b Cib1 Gdpap1 Gm15504 Tll13	Sema4b	2	2	0	20, 30, 50-100
7	126136157	126229106	Xpo6 Gm17137 Gm17136	Xpo6	2	2	0.0011571	80
11	44993577	45002073	Ebf1	Ebf1	2	2	6.277E-05	40, 60
14	121914942	121993177	Ubac2 Gpr18 Gpr183	Ubac2	2	2	0	20, 40-100
15	97756840	97814214	Rappel3 Gm16257 Gm16256 Slc48a1 Hdad7	Slc48a1	2	2	5.632E-05	50
17	45513141	45606050	Aars2 Tcte1 Tmem151b Nfkbie Slc35b2 Hsp90ab1 Slc29a1 Gm17080 Gm7325	Nfkbie	2	2	0	10-100
17	80424772	80439617	Sos1	Sos1	2	2	6.359E-05	80
18	50018926	50029554	Tnfrsf8	Tnfrsf8	2	2	2.987E-05	10, 40, 90
18	54955971	54955971	Zfp608	Zfp608	2	2	0.0021753	90
X	38527423	38569712	Cul4b	Cul4b	2	2	1.863E-13	30, 60-90
3	94969980	94964305	Rfx5 B230398E01Rik	B230398E01Rik	1	1	5.056E-05	40
3	100479915	100483577	Fam46c	Fam46c	1	1	9.02E-10	40
8	84907402	84908305	intergenic	Dnase2a	1	1	6.443E-06	10
9	82971653	82972578	Phip	Phip	1	1	2.891E-06	10
10	80108736	80112531	intergenic	Stk11	1	1	3.692E-06	40
14	76672486	76720129	intergenic	Serp2	1	1	9.473E-08	50
15	97731212	97738858	Endou	Endou	1	1	0.0018575	40
17	56598174	56605596	Safb	Safb	1	1	0.0016065	80
X	7811171	7813045	Gripap1	Gripap1	1	1	3.688E-08	20
X	12133903	12138595	Bcor 2908C10Rik	Bcor	1	1	0	50
X	48414879	48419570	Elf4	Elf4	1	1	1.11E-16	50

Table 6.3: Complete list of CIS generated by CIMPL analysis using the top 25 hits in each of the *Vk*HPB* IM lymphoma samples. The start and end boundaries encompass all analysis windows in which each locus was identified as a CIS. The gene shown as nearest to peak was the central gene in the majority of kernel windows (scales) detecting the CIS, but is not necessarily the target gene for the CIS. Due to local hopping the total number of insertions occasionally included multiple integrations from the same tumour, so both the total number and the number after correction for local hopping are shown. The smallest p value identified at any scale is shown along with the analysis scales at which the CIS was detected (x1000).

Chr	Start	End	Genes in CIS	Gene nearest to peak	Number insertions	Number insertions (no hop)	p value	Scale
13	3760904	38076948	Rreb1 Ssr1 Caga1 Rlok1	Rreb1	45	36	0	10-100
16	23859730	24308844	Ssl Rtp2 Bol6	Bol6	19	16	0	10-100
11	3018303	3350841	Pisd-ps1 Sfl1 Gm11399 Gm11400 Drg1 Gm12735 Fau-ps2 Efa4en11 Patz1 Gm12592 Gm11844 Pk3p1 Lim2	Sfl1	19	15	0	10-100
1	58523329	58882113	Fam126b Ndufb3 Gm10068 Als2cr12 Cflar Casp8	Cflar	14	14	0	10-100
4	32215887	32588511	Bach2 D130062J21Rik Gm11932 Gm24371 BC024582	Bach2	13	10	0	10-100
5	147199841	147483174	Gm24556 2210019111Rik Pdx1 RP24- 510G5.4 Cdx2 Pmoxnb Fl3 AC134441.1 Gm054 Pan3	Flt3	10	7	0	10-100
3	135522156	135794340	Manba Oaz2-ps Nkb1 Gm9799	Nkb1	7	7	0	10-100
9	32456146	32777632	Fl1 Ets1	Ets1	7	7	0	30-100
3	102923643	103141390	Syp1 Nr1h5 Gm22826 Sika1 Cade1 Nras Ampd1 Gm23820 Dend2c	Csde1	6	6	0	10-100
11	108273985	109479376	Rgs9 Gm11696 Gna13 9930022D16Rik Amz2 Gm15642 Slc16a6 Gm25540 Arng	Gna13	6	5	0	10-100
16	36530447	36730895	Casr Cd86 Ildr1	Cd86	6	5	0	10-100
16	5728468	5937755	Gm26369 Nfbiz Nvpe3 Cep97	Nfbiz	5	5	0	10-100
15	61857394	62060660	Myc Pxf1	Myc	6	4	0	16-100
6	98915128	99120608	Foxp1	Foxp1	5	4	0	20-100
14	114940488	115120931	Gm24073 Mir17hg Mir17 Mir18 Mir19a Mir20a Mir19b-1 Mir92-1 Gpc5	Mir17hg	5	4	0	10-100
3	51341797	51447464	Mgapr Ndufc1	Naa15	4	4	1.145E-11	40-70, 90, 100
6	99277164	99463074	Foxp1 Gm20696 Gm20705	Foxp1	4	4	0	30-100
10	68079912	68275650	Arnd5b	Arnd5b	4	4	0	10-100
11	62305055	62437030	Ttc19 Gm12275 Ncor1 Gm12276	Ncor1	4	4	0	10-30, 50-100
16	4486960	4590661	Srl Gm15885 Tlap4	Tlap4	4	4	0.0004811	60-100
17	46715492	46941700	CT030702.1 Pex8 Gm1 Cnpy3 P1ora Z310039H08Rik Rpf711 Gltscr11 A330017A19Rik Tlcc Gm23797 Pph2 Ubr2	Gltscr11	4	4	0	10-100
X	18090577	18280335	Dusp21 Kdm6a	Kdm6a	4	4	0	10, 30-100
12	58923991	59079189	AC163296.1 Sec23a Gemin2 Gm22973 Trappc8b Pm	Sec23a	5	3	0	10-90
11	89052813	89089993	Dgke Gm24974 A930013810Rik Gm525	Dgke	4	3	0.000567	70
1	175819610	175931858	Exo1 Gm23805	Exo1	3	3	0	20-80, 100
2	45048065	45075395	Zeb2	Zeb2	3	3	0.0001208	70, 80
5	64572308	64721730	Gm25306 RP24-448C16.1 RP24-448C16.2	Gm25306	3	3	0	30-100
6	100333917	100373056	Intergenic	Rybp	3	3	0.000447	100
6	148905409	148980057	Fam60a 3010003L21Rik Gm23395 Gm25539	Fam60a	3	3	5.598E-11	10, 20, 40, 80-100
7	3186234	3265222	Mir290 Mir291a Mir292 Mir291b Mir293 Mir294 Mir295 Nirp12	Nirp12	3	3	4.186E-08	60, 90
7	28054513	28140292	Grik5 Zfp574 Pou2f2	Pou2f2	3	3	0	20, 80-100
10	128758739	128885989	Wlbg Rpsa-ps2 Mmp19 Tmem198b Dnajc14 Ormdl2 Samp Gdf11	Mmp19	3	3	9.983E-08	10, 40, 100
11	61072660	61168564	Tnfrsf13b Gm12269 Usp22	Tnfrsf13b	3	3	0	10-100
13	48131563	48256184	AC140293.1	AC140293.1	3	3	0.0002004	80-100
14	121864043	121981443	Ubac2 Gpr18 Gpr183	Ubac2	3	3	0	70, 80
16	10433615	10587629	Tvp23a Cilia Dexl Clec16a	Cilia	3	3	0	10-100
16	38485797	38539866	Cd80 Timmdc1 Gm15953 Poglul1	Timmdc1	3	3	0.0007035	70
17	5336726	5478112	Arid1b Tmem242 Gm22475	Tmem242	3	3	0.0005588	90, 100
17	24338911	24395466	Abca17 Abca3 Gm25618	Abca3	3	3	0.0005634	100
17	75407283	75510966	Rasgrp3	Rasgrp3	3	3	0	30-100
18	65400269	65488624	Gm22567 Malt1 Gm26114	Malt1	3	3	0	10-80, 100
19	4297380	4471645	Adrbk1 Kdm2a Rhod A930001C03Rik Syt12	Kdm2a	3	3	0.0001512	10-100
19	34171155	34283560	Stambpl1 Acta2	Acta2	3	3	0.0010686	90, 100
19	44300363	44474648	Sod2 Mir5114 Sod4 Sod1	Sod1	3	3	0	10-100
19	6324570	6471336	Cdc42bpg Men1 Map4k2 Gm14966 Gm22278 Sfl1 Ppym Rasgrp2 Gm14965 Nrx2 Gm25470	Rasgrp2	4	2	0	10-60, 80-100
7	110230822	110242509	Swap70 Gm22185	Swap70	3	2	0.0005304	60
X	11961414	12181172	Gm14512 Bcor 2900008C10Rik	Bcor	3	2	0	10-50, 70-100
1	11345739	11352185	Intergenic	A830018L16Rik	2	2	4.477E-08	10-30
2	32635236	32664468	AK1 Eng Mir1954	Eng	2	2	0	10, 30, 70, 90
2	168568329	168609254	Nfatc2	Nfatc2	2	2	7.648E-09	20, 30
2	179522970	179542465	Cdh4	Cdh4	2	2	5.577E-09	10, 30, 40
3	9610064	9651853	Intergenic	Zfp704	2	2	7.179E-07	30-50, 70, 80
4	154885055	154990673	Mme11 Fam213b Tnfrsf14 Gm20421 Hes5 Pank4 Pch2	Tnfrsf14	2	2	0	10, 20, 40-70, 90, 100
6	86905261	86907211	Aak1	Aak1	2	2	1.694E-11	20
6	87825230	87855075	Isy1 Cnbp	Cnbp	2	2	4.707E-06	60, 70, 90
7	43339452	43359983	Siglec5	Siglec5	2	2	0	10, 20, 60
8	72398249	72362298	Klf2 Eps1h1	Klf2	2	2	0.0009992	70
8	105164301	105174567	Ctbf Gm22063	Ctbf	2	2	1.038E-07	50, 60, 90
9	88438513	88464196	Gm20537 4932427H20Rik Syncr1p	Gm20537	2	2	8.416E-06	10, 20
10	81312878	81424445	Cactin Tbx2r Gipc3 Hmg20b Misd12 4930404N11Rik Fz1 Dohh Z210404O07Rik Nfic Gm16104	Dohh	2	2	0	10-100
11	103231757	103243849	Map3k14	Map3k14	2	2	1.998E-15	20-40
11	107932992	107936888	Prkca	Prkca	2	2	3.25E-09	20
12	54975064	54997395	Baz1a Gm20403	Baz1a	2	2	4.465E-05	20, 40, 60, 90
12	72673685	72696616	Intergenic	Dhrs7	2	2	2.154E-05	40, 90
12	92880073	92883895	Intergenic	Gm23249	2	2	0.0003176	40
12	98973696	98994165	Ttc8	Ttc8	2	2	7.88E-05	30, 90
17	23584741	23603145	Zfp13	Zfp13	2	2	0	20, 50, 60
18	2966613	3021996	Intergenic	Vmn1r-ps151	2	2	0	20, 30, 50, 70, 100
18	34933618	34947414	Hspa9 Gm22200 Gm26109	Hspa9	2	2	6.965E-11	10, 50
18	46917651	46930883	Arl14ep1	Arl14ep1	2	2	0	10, 20, 50
18	60705221	60898148	Nds1 Rps14 Gm8731 Cd74 Mir5107 Tcof1	Cd74	2	2	0	10-100
19	6046035	6117281	Sym1 Mrpl49 Fau Znh12 Tm7af2 Vps51 Zfp11 BC048609 Cdca5 Naalad1 Sac3d1	Naalad1	2	2	0.0007123	60, 90, 100
19	26532985	26543910	Intergenic	Smarca2	2	2	1.439E-07	10, 30, 50
X	136022361	136111302	5730412P04Rik	5730412P04Rik	2	2	3.593E-07	50, 90
3	88483631	88509602	Lmna	Lmna	1	1	1.102E-07	60, 70
7	101407282	101408254	Arap1	Arap1	1	1	4.123E-06	10
10	127576664	127579639	Lrp1	Lrp1	1	1	4.693E-08	10
14	31153984	31158222	Stab1	Stab1	1	1	1.11E-16	10, 20

Table 6.4: Complete list of CIS generated by CIMPL analysis using the top 25 hits in each of the *Vk*MYC-TA-hPB* IM lymphoma samples. Columns are as described in table 6.3.

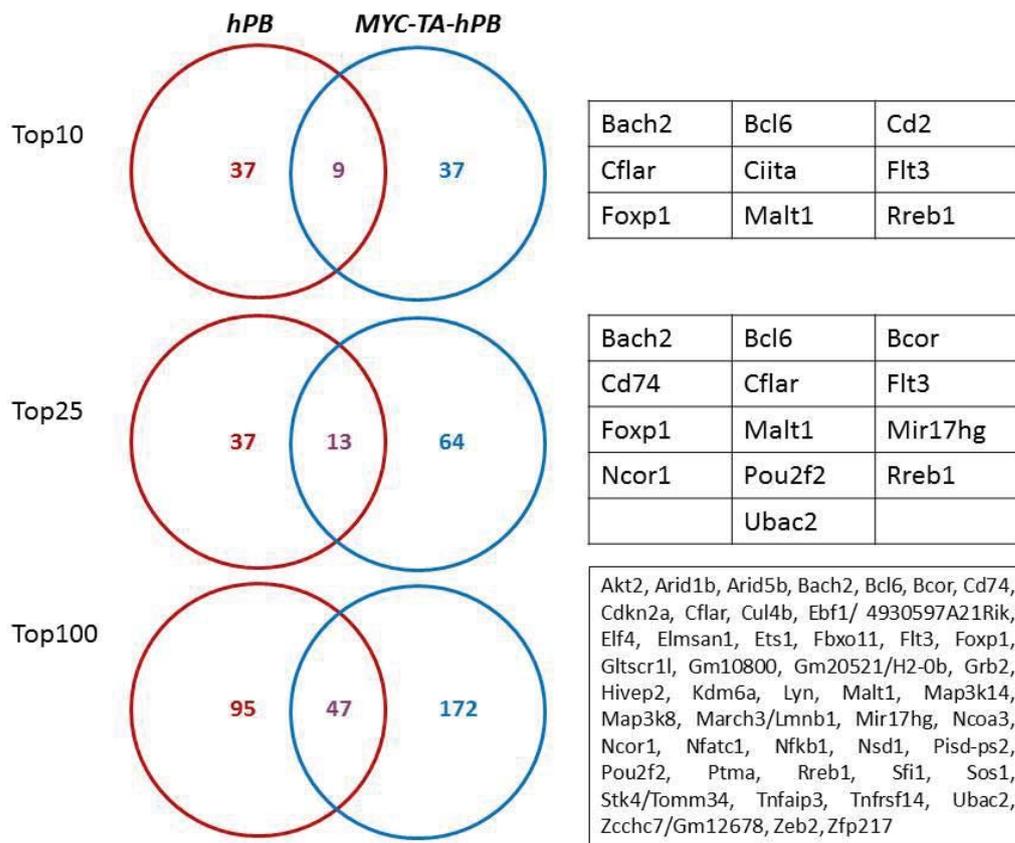


Figure 6.25: Overlapping integrations between the two cohorts on analysis using the top 10, 25 or 100 integrations in each. The identity of the central gene in the CIS is shown on the right, for each of the shared integrations.

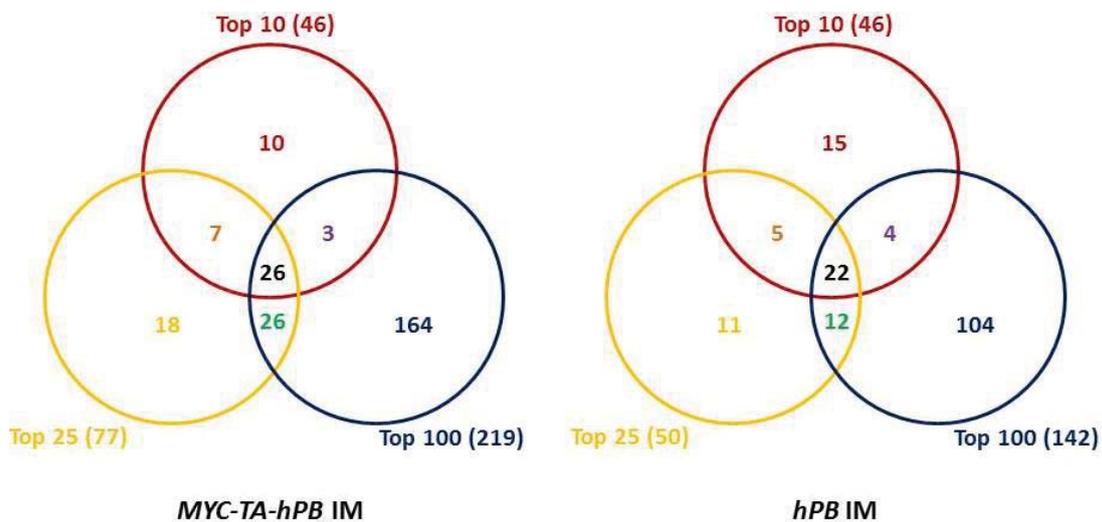


Figure 6.26: Number of integrations shared between analysis using different cut-offs for the number of integrations per tumour included in the analysis. The total number of CIS detected in each analysis is shown in brackets. This includes all CIS including those detected only at a single kernel window.

The CIS with the highest number of integrations in both cohorts was at the position of the gene for *Rreb1* (Table 6.3-6.6). *Rreb1* integrations were in the forward orientation within intron 1 in the vast majority of cases suggesting this was an activating integration. Another frequently hit gene in both cohorts was the known B cell lymphoma oncogene *Bcl6*. Although the larger CIS windows report a single CIS spanning the *Bcl6* gene and its flanking regions, the smaller windows report multiple CIS around *Bcl6*. The integrations are localised within intron one and the intergenic region 5' of *Bcl6*. For example, in the *Vk*hPB* CIS data, using the top 25 integrations the 10kb kernel scale reports three discrete CIS; i) centred at 23988419, with 10 hits, ii) centred at ~24144166 with four hits and iii) at ~24188921 with four hits. The first location involves the integrations within intron 1 of *Bcl6*, but the latter two sites are 5' of the gene. It is noteworthy that although the integrations in intron one are almost universally in the forward orientation, which suggests this is an activating integration, those 5' to the gene occur in both orientations.

6.2.12 Read depth and correlation with sample clonality

In addition to sequencing the *Vk*hPB* and *Vk*MYC-TA-hPB* IM mice that developed tumours I also included many spleen samples from mice which were not found to have an abnormality on histopathology analysis. Although the read coverage was generally lower in these samples compared to mice with lymphoma (mean 55491 vs 113460 reads, $p=0.0048$) this still provided deep coverage of transposon integrations. A mean of 92 unique transposon integration sites were mapped with 2 or more reads after removal of duplicates in these samples compared to 466 in the *Vk*hPB* and *Vk*MYC-TA-hPB* IM mice ($p<0.0001$). It is notable that the proportion of reads assigned to the top integrations in these samples is significantly lower than in the mice with lymphoma (4.6% vs 18% $p=0.0008$ in the non-duplicate filtered data) (figure 6.28) although there were a few outlier samples. The sample with the lowest read coverage (8933 reads including duplicates) was unremarkable on histopathology, but over 23% of reads mapped to a single integration (6:71585574, intergenic). It is possible this was an artefact related to the low read coverage.

The drop off in the read count in data without removal of the PCR-duplicates was reviewed for the top 50 integrations in samples which also underwent BCR repertoire analysis (figure 6.29). In the two samples in which the largest clone was less than 2.5% on B cell repertoire analysis, the top transposon integration accounted for

under 1% of total reads. For *Vk*hPB* 3.3a, which had two clusters of 10 and 15%, the top integration accounted for 19% of reads, but the fall off in read counts for the next most common integrations was marked. It is evident that the pattern of fall in read count is different between samples and this did not appear to directly correlate with the number and size of clusters on B cell repertoire analysis. Therefore, from the transposon integration read counts it is not possible to tell if the major integrations are shared in a single or multiple clones although it is possible to infer which samples are unlikely to contain a clonal expansion.

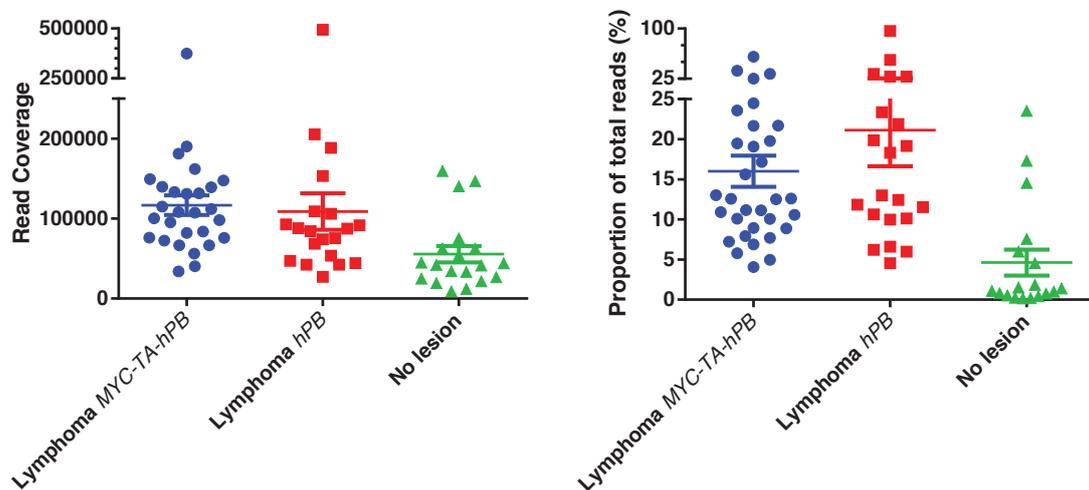


Figure 6.28: Sequence coverage and proportion of reads assigned to the top insertion
 Read number (left) and proportion of reads assigned to the top hit (right) in IM mice with either no malignant lesion or a diagnosis of lymphoma. Only mice sequenced on the first two (of three) TraDIS sequencing runs are represented. Each dot represents a single mouse. The mean and standard error of the mean are indicated. PCR duplicates were not removed for this analysis.

Multiple samples from a single mouse were analysed by TraDIS in 22 *Vk*MYC-TA-hPB* and nine *Vk*hPB* cases. In those where lymphoma was the only tumour identified on histopathology the major transposon integrations were similar between the different tissue samples. In samples where a non-haematopoietic tumour was also detected, transposon integrations were often detected in these tumours, but the integrations typically differed to those in the lymphoma samples.

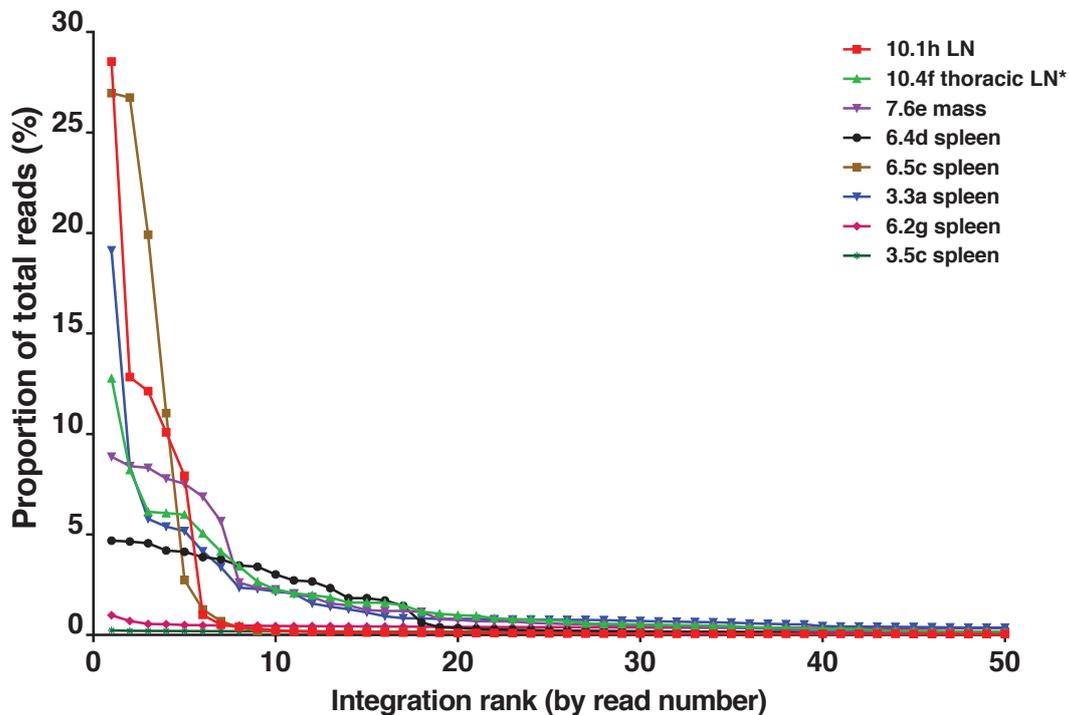


Figure 6.29: The proportion of reads assigned to each of the top 50 transposon integrations in mice that had B cell repertoire analysis. The mice are listed in order of size of the dominant clone as determined by BCR analysis. The tissue that was sequenced is shown and was the same tissue used for B cell repertoire analysis except for 10.4f (*) in which thoracic LN was used for integration site analysis and spleen for the BCR analysis. Mice 6.2g and 3.5c had no abnormality identified on histopathology.

6.3 Discussion

Two *PB* IM mouse cohorts based on *Vk*MYC* mice, which were reported to develop highly penetrant plasma cell malignancies (Chesi et al., 2008), had significantly reduced survival largely due to the development of mature B cell lymphomas, but did not develop multiple myeloma. Although the *PB* transposase was clearly active, the mechanism of activation was not through reversion of the stop codon by SHM, as had been anticipated given the design of the construct. Furthermore, there was evidence of jumping in multiple tissues analysed by PCR, which suggests that activation of *PB* was not specific to the mature B cell compartment. Although it is possible that positive jump PCRs reflect the presence of mature B cells in non-haematopoietic tissues, the finding of large numbers of transposon integrations in

malignant samples which were not B cell lymphomas, and the positive jump qPCR in different haematopoietic lineages is evidence that the transposase was active in cells other than B cells. Nevertheless, the significant incidence of B cell lymphomas suggests that transposition was most active in B lymphocytes.

In the original *Vk*MYC* model as published in *Cancer Cell* (Chesi et al., 2008), it was predicted that translation of the transgene would be stopped by the engineered stop codon in the third codon of the V-kappa exon. However, this stop codon was also engineered as part of a DGYW motif, a preferential target sequence for AID-mediated SHM, in order for it to be sporadically reverted to allow *MYC* translation in a small proportion of germinal and post-germinal centre B cells. In the original paper transfection experiments in 293T cells were used to confirm that translation of *MYC* was absent, and that it did not initiate from a downstream AUG in the setting of the engineered stop codon (Chesi et al., 2008). Furthermore, *hMYC* mRNA expression was only detectable in spleen and BM and was up-regulated by LPS stimulation to induce plasma cell differentiation. The authors reported that 100% of *Vk*MYC* mice developed a monoclonal expansion of bone marrow plasma cells with age, with approximately 80% of necropsied mice found to have more than 5% (range 2%-62%). In contrast, transgenic mice with *MYC* but without the engineered stop codon developed aggressive pro-B lymphomas (Chesi et al., 2008).

To evaluate reversion of the stop codon, Chesi et al performed single colony sequencing on the heavy chain locus VDJ fragment and the transgenic *Vk* region from CD138 selected tumour plasma cells. They found evidence of SHM at both loci, with a median of 2.6% mutations in the *VH* gene and 2.4% in the transgenic *Vk* region, but no evidence of SHM in the *MYC* exons. Of the ten mice on which data was presented, reversion of the stop codon was detected in at least one colony in six, including both mice with Burkitt-like lymphoma. Although not every colony showed reversion of the stop codon, they thought this was because multiple copies of the transgene were expressed (either 20 or 8 copies) and concluded that reversion did occur in every plasma cell because they expressed *Myc* by IHC. However, although it was raised against amino acids 1-262 of the human protein, the anti-*Myc* antibody used (sc-764_N-262) is not specific for human *Myc* (<http://datasheets.scbt.com/sc-764.pd>).

In the *PB* IM mouse cohorts I did not identify reversion of the stop codon in any of the tumours analysed using PCR and capillary sequencing of DNA from the bulk tumour. RT-PCR and deep sequencing on MiSeq, showed that over 94% of reads in each tumour had an identical sequence, with the stop codon intact. Although a small number of reads with reversion of the stop codon were detected in most tumours, these accounted for less than 1% of the total reads in each case, and therefore could not be responsible for activation of the transposase in the main tumour clone.

The precise mechanism by which the transposase is activated in these transgenic mice is yet to be determined. Alternate start codons in the leader sequence were mutated to prevent premature initiation and there are no in-frame initiation codons downstream of the engineered stop codon. The presumption is that there is a cryptic initiation codon and alternate splicing which allows bypass of the stop codon. This may occur from a distant methionine, but translation initiation at non-AUG triplets has also been described in mammalian cells (Peabody, 1989; Starck et al., 2008). These issues were discussed with Leif Bergsagl, senior author on the *Vk*MYC* paper. He acknowledged they are also now suspicious that reversion of the stop codon is not the main mechanism for transgene expression and this may only be true in the infrequent mice which develop Burkitt lymphoma (2/122 mice in the original paper) (Leif Bergsagl, personal communication). Their evidence for this is that the Myc protein is frequently truncated in the plasma cell tumours. Furthermore, in a modified *Vk*MYC* model in which *Cre* was expressed instead of *MYC*, *Cre* activation occurred in 15% of plasma cells, which seems too high for SHM (Leif Bergsagl, personal communication).

The other striking feature of our *PB* IM mice cohorts is the similar median survival and tumour rate between the *Vk*MYC-TA-hPB* and *Vk*hPB* mice, suggesting that the *MYC* transgene did not have a strong additional tumourigenic effect. In keeping with this, the *Vk*MYC-TA-hPB* IM mice without the *GRL* transposon did not have shortened survival compared to the *GRL* only mice. Less than a quarter of these mice died with lymphoma and there were no cases of plasma cell malignancy at death. Therefore, even in the absence of transposition, the *MYC* transgene did not have a powerful tumorigenic effect in the B cell compartment. One possible explanation is that the T2A linker adversely effected Myc function, as after cleavage it is expected to leave a 17 amino acid tail on the C terminal end of Myc (Szymczak et al., 2004).

Interestingly, we identified a CIS centred at *Myc* in the *Vk*MYC-TA-hPB* cohort, but not in the *Vk*hPB* IM mice. Although there were some integrations around this site detected in *Vk*hPB* mice, these were always in low number (<0.5% of reads). In three *Vk*MYC-TA-hPB* mice (7.3h, 8.1b and 16.2a) this was amongst the top 10 hits by read number. The transposon integration site was at position 15:61983792, 2kB upstream of the *Myc* gene in all three cases. If the *MYC* transgene was having a strong lymphomagenic effect, it would be surprising to have further selection for integrations at the *Myc* locus during disease development.

Although the *Vk*MYC-TA-hPB* and *Vk*hPB* mice did not develop the spectrum of tumours we anticipated, insertional mutagenesis was still active in the B cell compartment and resulted in tumour formation. Furthermore, these tumours were predominantly mature B cell lymphomas. CIS analysis performed on integrations from these tumours identified several known lymphoma associated genes (e.g. *Bcl6* and *Malt1*) and novel genes of interest. In this regard, the study was successful in identifying putative cancer genes of relevance to mature B cell malignancies, which are candidates for further study.

The CIS genes included several genes with known roles in lymphomagenesis. The CIS with the second highest number of hits in both cohorts was at the *Bcl6* locus. *Bcl6* is a transcription factor and a proto-oncogene that is expressed in normal germinal centre (GC) B cells and follicular helper T cells (Wagner et al., 2011). *Bcl6* deficient mice display normal B cell development except they are unable to form GC (Dent et al., 1997; Ye et al., 1997). Within the B lymphocyte lineage, expression of *Bcl6* is restricted to GC B cells and its down-regulation is thought to be required for differentiation into plasma cells or memory B cells. Normally, expression of *Bcl6* is regulated by T cell induced CD40 signalling, which down-regulates *Bcl6* via *NF-κB* and *IRF4* activation (Saito et al., 2007), and B cell receptor signalling which leads to *Bcl6* phosphorylation by MAPK and targets it for degradation by the ubiquitin proteasome pathway (Niu et al., 1998). *Bcl6* is expressed in cases of diffuse large B cell lymphoma (DLBCL), follicular, Burkitt's, primary mediastinal B cell, and some Hodgkin lymphomas. Overall about 20-35% of DLBCL have chromosomal translocations involving *BCL6* (Iqbal et al., 2007; Lo Coco et al., 1994; Offit et al., 1994). These translocations are more common in the activated B cell type, than germinal centre cases of DLBCL, although gene expression studies have associated

overexpression of *Bcl6* with a GC signature(Wagner et al., 2011). Mouse studies have demonstrated an auto-regulatory site in the first exon of *Bcl6* and around 10-15% of DLBCL have mutations in this auto-regulatory sequence which is another mechanism of *Bcl6* overexpression (Iqbal et al., 2007; Wang et al., 2002). The 5' non coding region of *BCL6* is also a target for SHM and mutations are reported in around 60% of cases of DLBCL, although the functional significance of many of these mutations is not known (Iqbal et al., 2007).

Bcl6 is a transcriptional repressor. Its activity is dependent on binding to specific DNA sequences via its C-terminal zinc finger domain and recruiting co-repressor molecules including SMRT, NCoR and BCoR (Wagner et al., 2011). This co-repressor complex then recruits histone deacetylases, leading to transcriptional repression of target genes, including *ATR*, *TP53* and *CDKN1A*(Wagner et al., 2011). Interestingly both *Ncor1* and *Bcor* were also identified as CIS genes in the IM cohorts.

The pathogenic role of deregulated *Bcl6* expression in lymphomagenesis has been demonstrated in a transgenic mouse model in which the full length of the murine *Bcl6* coding sequence was expressed under the control of the immunoglobulin heavy chain ($I\mu$) promoter(Cattoretti et al., 2005). *Bcl6* transgenic mice had an increased number of GCs and at six months of age, showed abnormal polyclonal B cell expansions, with partial effacement of the follicular architecture of lymphoid organs consistent with a benign lymphoproliferative disorder. From 13 months onwards, *Bcl6* transgenic mice showed increased mortality due to clonal B cell lymphomas, which had a mature B cell phenotype (IgM+IgD+CD43-) with histology similar to human DLBCL in most cases (Cattoretti et al., 2005).

Another CIS common to both screens and with a well described role in lymphomagenesis is the mucosa associated lymphoid tissue lymphoma translocation gene 1 (*Malt1*). The detected integrations were in the forward orientation and predominantly in intron 6 of *Malt1* and are likely to be activating integrations. *Malt1* was initially described due to its recurrent translocation, t(11;18)(q21;q21), in MALT lymphoma(Dierlamm et al., 1999). The translocation created a fusion protein and resulted in constitutive activation of the canonical NF- κ B pathway. It was subsequently recognised that *MALT1* binds to and synergises with *BCL10* to promote canonical NF- κ B activation as part of the CARD11-BCL10-MALT1

(CBM) complex (McAllister-Lucas et al., 2011). This activates the inhibitor of kappa B kinase (IKK) complex, leading to degradation of inhibitor of kappa-B α (I κ B α) and release of active NF- κ B dimers into the nucleus. Therefore, as part of the CBM complex *MALT1* activates canonical NF- κ B activation downstream of both T and B cell receptor stimulation, resulting in cytokine production and cellular proliferation in response to antigen stimulation (McAllister-Lucas et al., 2011). In addition to its role in MALT lymphoma, it has recently been recognised that *MALT1* is important in the pathogenesis of DLBCL, even in the absence of described mutations or translocations in this disease. Activated B cell-like DLBCL (ABC) is characterised by constitutive NF- κ B activity and in an shRNA screen, *CARD11*, *BCL10* and *MALT1* were found to be essential to the survival of ABC-DLBCL cells (Ngo et al., 2006). Furthermore, inhibition of *MALT1* protease activity was toxic for ABC-DLBCL (Ferch et al., 2009). In this context it is also noteworthy that other genes involved in the regulation of NF- κ B transcription factor complexes were also identified as CIS genes in this IM screen including *NF κ BIZ* and *NF κ B1*.

Mir17hg (mir17-92 cluster of microRNAs) is another CIS which has previously been implicated in lymphomagenesis. In humans, *Mir17hg* is in a region at 13q31-q32 which is frequently amplified in DLBCL, follicular, mantle and other lymphoma subtypes (He et al., 2005). Overexpression of *Mir17hg* in transgenic mice resulted in increased lymphocyte proliferation and decreased cell death, and lymphoproliferative and autoimmune disease (Xiao et al., 2008). Overexpression of *Mir17hg* with *Myc* in *EuMyc* mice resulted in accelerated development of aggressive lymphoid malignancies (He et al., 2005).

Other CIS identified in our models included well-known transcription factors involved in B cell development including *Foxp1*, *Ets1*, *Pax5* and *Bach2*. *Foxp1* is involved in chromosomal translocations in DLBCL and MALT lymphoma, but it may also be overexpressed by other mechanisms (Goatly et al., 2008). It is reported to be expressed in 40-60% of DLBCL and its overexpression is associated with poor prognosis (Banham et al., 2005; Barrans et al., 2004). Recurrent gains at 11q24.3 were recently described in 23% of DLBCL cases and were associated with significantly higher expression of *ETS1* (Bonetti et al., 2013). *PAX5* is an important regulator of B cell differentiation, which, among other roles, activates *BACH2* and initiates the GC reaction. Translocations involving *PAX5* are described in human B cell lymphomas

(Cobaleda et al., 2007) and loss of heterozygosity of *BACH2* has been described in around 20% of B cell lymphoma (Ichikawa et al., 2014). The overexpression of *BACH2* has been associated with poor prognosis in DLBCL in one study (Ichikawa et al., 2014) although the reverse finding was reported in another (Sakane-Ishikawa et al., 2005).

Along with CIS genes with a well described role in lymphomagenesis, there were a number other CIS that warrant further investigation. The most frequently hit CIS in both cohorts was centred on the gene *Ras* responsive element binding protein 1 (*Rreb1*). *Rreb1* is a paralog of *ZNF821* and encodes a zinc finger transcription factor that binds specifically to the *RAS* responsive elements of gene promoters. It has previously been identified as a putative oncogene in a retroviral and a transposon insertional mutagenesis screen (Starr et al., 2009; Uren et al., 2008) and has been implicated in the pathogenesis of several tumours including childhood ALL (Xiao et al., 2014), thyroid (Thiagalingam et al., 1996), pancreatic (Costello et al., 2012) and colorectal (Kent et al., 2013) malignancies. A three way translocation involving *MLL-ENL* and *RREB1* was also recently described in paediatric AML (Tuborgh et al., 2013).

RREB1 is thought to be activated by *RAS* signalling downstream of the MAPK pathway (Kent et al., 2013; Thiagalingam et al., 1996; Zhang et al., 2003). In a recent study of childhood ALL, *RREB1* was shown to bind to the *PTPRG* promoter and the methylation status of the *PTPRG* locus was found to be a complementary event in oncogenesis and was associated with *RAS* mutation status (Xiao et al., 2014). The differential susceptibility of Balb/c mice to pristane induced plasma cell tumours has been attributed to polymorphisms in the *p16^{INK4a}* gene promoter that effect the *Rreb1* binding site and increase *Rreb1* mediated repression of *p16^{INK}* (Zhang et al., 2003). *RREB1* silencing of *ZIP3* is thought to be an early event in the evolution of pancreatic adenocarcinoma, which results in reduced zinc levels in the ductal and acinar epithelium (Costello and Franklin, 2013). A similar role in the down-regulation of the zinc transporter *hZIP1* has been described for *RREB-1* in prostate cancer (Milon et al., 2010). *RREB1* also represses miR-143/miR-145 promoter activity (Kent et al., 2010) and the loss of miR-145 is seen in *K-RAS* mutated pancreatic cancers (Sureban et al., 2013) and colorectal carcinoma (Kent et al., 2013). *K-RAS* mediated activation of *RREB1* is thought to directly inhibit transcription of the miR-143/145 cluster. This appears to be a feed forward mechanism to potentiate *RAS* signalling in these

tumours as *K-RAS* and *RREB1* are targets of miR-145 and therapeutic restoration of miR-145 abrogates tumorigenesis.

Despite the numerous recent accounts of the oncogenic role for *RREB1*, I could not identify any reports of mutations or dysregulation of *RREB1* in human B cell lymphomas. Nevertheless, in our IM cohorts 23/33 *Vk*hPB* IM and 25/65 *Vk*MYC-TA-hPB* IM tumours had integrations within the *Rreb1* CIS in the top ten hits and these were almost universally in the forward orientation suggesting they are activating mutations. Although *Rreb1* has been described as a CIS in some previous IM screens (Starr et al., 2009; Uren et al., 2008) it was only seen in small numbers of cancers and its specificity to any particular cancer was not determined. By contrast, in my data *Rreb1* was the most commonly hit CIS and this may suggest an important role of this gene as an effector of activated Ras and MAPK signalling. Mutant genes signalling through Ras such as *EGFR*, *FGFR2*, *KRAS* and *BRAF* have been found to be recurrently mutated in mature B cell malignancies (Chapman et al., 2011; Vaqué et al., 2014) and at least in some instances their key oncogenic effects may operate through *RREB1* overexpression. The ability of transposons to activate gene expression in a very different manner to naturally occurring human cancer mutations may be the reason this gene was identified by our studies, but has not been found to be recurrently mutated in human lymphomas.

Notably there were other CIS involving *RAS* pathway genes in the *Vk*MYC-TA-hPB* cohort including *Nras* itself. Although *Csde1* is listed as the central gene in this CIS, the majority of integrations are either in the terminal exon (exon 20) of *Csde1* or intron 1 of *Nras* and these are almost universally in the forward orientation for *Nras* suggesting activating integrations. These *Nras* integrations were amongst the top hits in samples that did not have *Rreb1* integrations as a top 25 hit, suggesting that mutations in the two genes were mutually exclusive, in turn alluding to them having similar and therefore redundant effects. Similarly, two further tumours had top ten integrations in the *Ras* activator *Rasgrp2*, and neither of these tumours had *Rreb1* integrations in significant number. Another CIS in the *MYC-TA-hPB* cohort involved *Rasgrp3*, a member of the *RAS* family of GTPases, and this did co-occur with *Rreb1* as one of the top hits in some, but not all of the tumours in which it was present. Notably, none of these sites were detected as CIS in the *hPB* cohort, which had a much higher rate of *Rreb1* integrations.

Evidence from a conditional *Kras* mutant mouse model is that *Ras* activation is insufficient to transform primary germinal centre B cells. These mice had a conditional *KRas*^{G12D} mutation that was induced by two different *Cre* recombinases thought to be specific to germinal centre B cells (Cy1-Cre and AID-Cre). The mice developed T cell lymphomas, lung adenomas and sarcomas but not B cell lymphomas or plasma cell tumours despite the presence of activated *Kras* in the B lineage cells (Mullins et al., 2013). Even in a tumour-prone *Arf*-null genetic background or following sub-lethal irradiation, the *KRas*^{G12D} mutation failed to induce a mature B cell malignant phenotype, which suggests that *Ras* activation is insufficient to transform primary germinal centre B cells (Mullins et al., 2013). *Rreb1* integrations were also detected in several of the IM mice that died with normal spleen size and no detectable lesion on histopathology at death, which suggests that this is an early and commonly occurring integration in these mice, but is insufficient itself to generate lymphoma.

Another CIS gene of interest is *Cflar* (CASP8 and FADD-like apoptosis regulator), also known as *cFLIP*, which was hit by insertions which were almost universally in the forward orientation in intron 2. This gene was originally identified as a competitive inhibitor of death signalling as it blocks recruitment of caspase-8 to the death inducing signalling complex (DISC) (Budd et al., 2006). Subsequently it has also been found to form a heterodimer with caspase-8 and to activate caspase-8. This heterodimer also links T cell receptor signalling to activation of NF- κ B through the CMP complex (Budd et al., 2006). Notably, lymphoma cell lines with constitutively activated NF- κ B are resistant to induction of apoptosis and *Cflar* is one of the downstream targets whose overexpression is associated with this resistance (Bernal-Mizrachi et al., 2006). Retroviral expression of *Cflar* in B cells was reported to reduce CD95 mediated B cell death and cause retention of activated B cells in the germinal centres, while transduction of B-lymphoma cells with viral *FLIP* resulted in highly aggressive tumours which were resistant to CD95 induced cell death (Budd et al., 2006; Djerbi et al., 1999). The expression of *Cflar* has been associated with poor outcome in Burkitt lymphoma and in DLBCL of both GC and non-GCB sub-types (Harris et al., 2012; Valnet-Rabier et al., 2005).

In conclusion, although the *MYC-TA-hPB* and *hPB* IM mice did not develop plasma cell tumours, expression of the transposase under the control of the *Vk* gene

regulatory elements led to the development of mature B cell lymphomas as the predominant tumours in our mice. The identified CIS genes were similar between the two cohorts, suggesting that the *MYC* transgene was not operative. The identified driver genes included transcription factors, microRNAs and apoptosis regulators, many of which have been found to be de-regulated in human lymphoma. These CIS genes also include some, such as *Rreb1*, which could be the focus of future studies, to better understand the mechanisms by which they contribute to lymphomagenesis and to develop targeted therapies.