



Universiteit
Leiden
The Netherlands

Tracing T cell differentiation by genetic barcoding

Heijst, J.W.J. van

Citation

Heijst, J. W. J. van. (2010, June 24). *Tracing T cell differentiation by genetic barcoding*. Retrieved from <https://hdl.handle.net/1887/15721>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/15721>

Note: To cite this publication please use the final published version (if applicable).

Chapter 7

Tracing cellular origins by inducible DNA diversification in vivo

Jeroen W.J. van Heijst, Jos Urbanus,
Heinz Jacobs* and Ton N.M. Schumacher*

(*These authors contributed equally)

Unpublished

Tracing cellular origins by inducible DNA diversification in vivo

Jeroen W.J. van Heijst¹, Jos Urbanus¹, Heinz Jacobs^{1*} and Ton N.M. Schumacher^{1*#}

¹Division of Immunology, the Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, the Netherlands.

*These authors contributed equally to this work.

#Correspondence: t.schumacher@nki.nl

Mammalian tissues consist of many distinct cell types that each express unique phenotypes and functions. However, our understanding of the cell fate decisions that yield this diversity in cell types is still highly limited. To allow kinship analysis of different cell populations in vivo, we have generated a mouse model in which individual cells can be site specifically marked with unique DNA sequences. This “in vivo barcoding” is based on the V(D)J recombination machinery, normally required to generate antigen receptors in precursor lymphocytes. Upon conditional activation of Cre, a tricistronic V(D)J recombinase cassette is activated by inversion, allowing transient expression of the enzymes RAG1/2 and TdT and recombination of a pseudo-VDJ substrate. Generation of a VDJ product inactivates RAG/TdT, activates GFP and via nucleotide deletion and insertion leaves a heritable genetic signature in each cell. Here we demonstrate the feasibility of in vivo barcoding by inducing RAG/TdT activity in lung and liver cells, as well as in myeloid hematopoietic cells, and demonstrating that many of these cells contain unique DNA barcodes. This technology should prove useful for the analysis of lineage relationships in mammalian development and tissue homeostasis.

Complex organisms are build-up of many distinct cell types that each display different phenotypic and functional characteristics. Humans, for instance, have been estimated to contain 10^{14} cells in total, among which ~ 210 distinct cell types can be discriminated. To determine how these different cell types arise from individual precursor cells has been a long-standing challenge in many biological fields. In theory, two conceptually distinct approaches can be utilized to assess the developmental potential of precursor cells of interest. First, the fate of single precursor cells can be revealed by continuous observation of these cells and their progeny through time-lapse microscopy. In a classical example, direct observation of cell division in the transparent nematode *Caenorhabditis elegans* enabled lineage tracing of all its 959

cells^{1,2}. This experimental approach is however limited to those settings in which differentiation is completed within the time span for which observation is technically feasible and is also limited to those body sites that are accessible to intravital microscopy. As an alternative more broadly applicable approach, technologies have been developed that provide precursor cells of interest with unique cellular tags that are inherited by all progeny upon differentiation. In this case, the branching points that yield different cell lineages are inferred from the overlap in cellular tags recovered from different cell populations that emerge after precursor differentiation in vivo. In early work in this field, radiation-induced chromosome aberrations have been used to follow the developmental potential of cells within the hematopoietic progenitor

compartment, and this approach was successfully used to demonstrate the existence of separate myeloid and lymphoid progenitors³⁻⁵. In subsequent work, replication-deficient retroviruses have been used as a less invasive means of marking precursor cells, utilizing either the viral integration site as a genetic marker⁶⁻⁸, or using a library of retroviruses in which each member contains a unique DNA sequence to provide individual cells with a unique genetic tag^{9,10}. More recently, this strategy of retroviral marking with a library of unique DNA sequences has been extended to a microarray-based detection platform, making it possible to perform high-throughput fate mapping of genetically marked precursor cells¹¹⁻¹³.

While retroviral tagging has proven a powerful approach to study the differentiation potential of specific cell populations, its use is restricted to settings in which cells of interest can be selectively modified by retroviral transduction¹⁴. Because of this, retroviral tagging has primarily been used to study the developmental potential of cell types within the hematopoietic lineage that can be isolated from a mouse, modified *in vitro* and reintroduced *in vivo*. Recently, an alternative approach for cell fate mapping has been described, which makes use of phylogenetic analysis of somatic mutations that accumulate during normal development of every organism^{15,16}. Although this strategy has the advantage that it can reveal a continuum of developmental branches within a single experiment, successful analysis depends on the spontaneous mutation frequency and requires ~100 different markers to be analyzed for each cell to obtain sufficient resolution. Furthermore, this strategy can only be used for retrospective fate mapping, as the precursor cell of interest for lineage analysis cannot be chosen directly.

With the aim to develop a broadly applicable technology to analyze cell fate decisions we set out to generate a mouse model in which individual cells can be marked with unique DNA sequences (barcodes) *in vivo*. Ideally, such strategy should allow one to induce genetic diversity within a cell population of interest at a defined point in development and without having to remove these cells from their physiological niche. This would allow one to map

the fate of individual precursor cells with minimal manipulation. The technology we have developed, termed *in vivo* barcoding, is based on the V(D)J recombination machinery that generates antigen receptor diversity in lymphocytes. Specifically, through Cre recombinase-induced expression of recombination-activating gene (RAG)1 and RAG2 as well as terminal deoxynucleotidyl transferase (TdT), recombination and nucleotide diversification of a pseudo-VDJ substrate is achieved, thereby providing cells with a unique genetic tag. Here we develop the concept of *in vivo* barcoding and provide proof of principle for its value by conditionally inducing genetic diversity in the lung and liver as well as in myeloid hematopoietic cells.

RESULTS

In vivo barcoding: construct design

To generate a mouse model in which cells can be provided with unique genetic tags *in vivo* in a temporally controlled and cell lineage-specific manner, we have borrowed from the V(D)J recombination system that generates genetic diversity in T and B cells. Namely, during lymphocyte development, V(D)J recombination is responsible for the generation of the highly diverse repertoire of antigen receptors, which it accomplishes mainly by the deletion and insertion of nucleotides at recombining gene segments^{17,18}. Nucleotide deletions are the result on exonuclease activity, whereas nucleotide insertions can either be templated P nucleotides¹⁹ or non-templated N nucleotides inserted by TdT^{20,21}. To express such a system in non-lymphoid cells two components are minimally required. The first required component is a V(D)J substrate containing gene segments flanked by recombination signal sequences (RSSs) that act as target sequences for the RAG recombinase. RSSs consist of two types and are comprised of a conserved heptamer and nonamer sequence, separated by a 12 or 23 base pair (bp) spacer. The length of this spacer determines the functionality of the RSS and efficient recombination occurs only between two gene segments flanked by a 12- and a 23-RSS, known as the 12/23 rule²². The second required component is an inducible gene

cassette encoding the RAG recombinase and TdT, which are the only known lymphoid-specific enzymes of the V(D)J recombinase. While RAG1/2 mediate the site-specific recombination between adjacent RSS sites, TdT catalyzes the insertion of N nucleotides at recombination junctions, thereby greatly enhancing junctional diversity.

To test the feasibility of V(D)J recombination-mediated genetic tagging in non-lymphoid cells, we started by cloning a tricistronic TdT-RAG2-RAG1 cDNA cassette into a eukaryote expression vector. The TdT and RAG cDNAs were separated by the self-cleaving 2A peptides E2A and T2A to allow concomitant expression of all three enzymes from a single mRNA molecule²³. This construct was transiently transfected into human embryonic kidney 293T

cells together with a previously described VJ recombination substrate, which contains a V segment flanked by a 23-RSS and a J-segment flanked by a 12-RSS, and which expresses GFP upon recombination^{24,25}. These experiments revealed expression of GFP in ~10% of cells after 48 hours (not shown), indicating the feasibility of inducing V(D)J recombination in non-lymphoid cells. Unfortunately, the total barcode diversity generated in this system was rather limited (estimated to be <100 different DNA sequences; not shown). The main reason for this was that the used VJ recombination substrate gave rise to only one recombination junction (namely V-to-J joining), implying only one event in which nucleotides can be deleted or inserted (analogous to V-to-J recombination in TCR α or IgL chain loci).

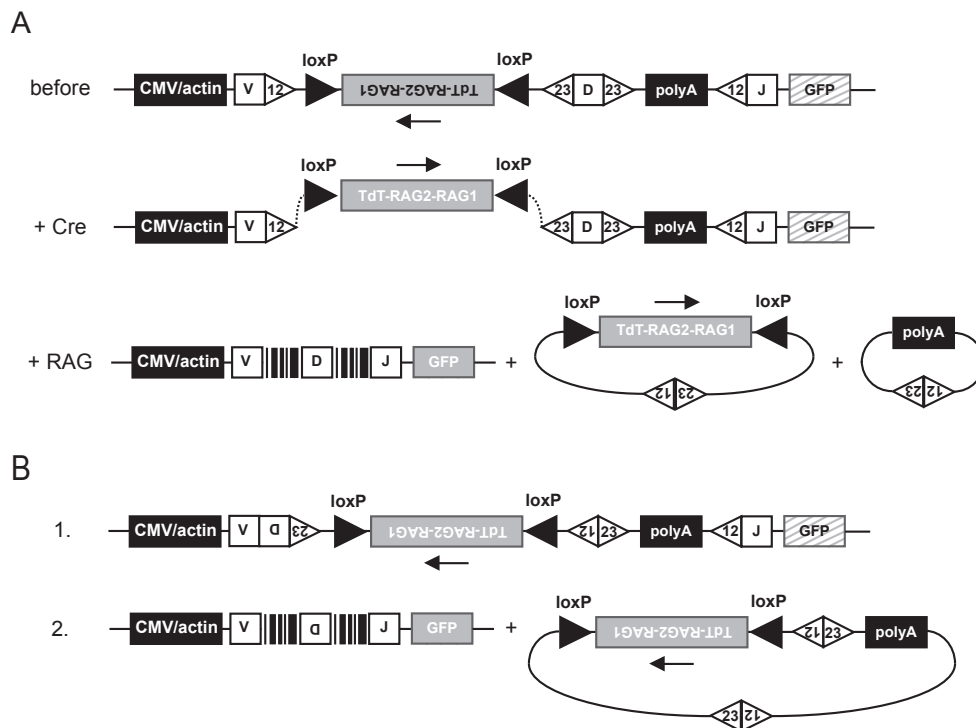


Figure 1. Design and recombination of the in vivo barcoding construct. (A) In the in vivo barcoding construct, an antisense-orientation cassette encoding the V(D)J recombinase enzymes RAG1, RAG2 and TdT is flanked by opposite-orientation loxP sites. Premature GFP expression is prevented by a BGH polyA site in front of the J-segment. Recombination is initiated by expression of Cre recombinase, which leads to inversion of the TdT-RAG2-RAG1 cassette. Subsequently, RAG and TdT are expressed from the ubiquitously-active CMV enhancer and chicken β -actin promoter. RAG binds to the 12- and 23-RSSs and mediates joining of the V-, D- and J-segments. As this joining is inherently imprecise and TdT can mediate non-templated nucleotide insertions at recombination junctions, these V-D-J junctions will consist of highly diverse DNA regions that can be used as genetic tags in kinship experiments. In most cases, recombination will take place between a 12- and 23-RSS in opposite orientation resulting in deletion of intervening sequences. Such recombination results in the inactivation of RAG and TdT as well as in the activation of GFP as a marker. (B) In case recombination occurs between a 12- and 23-RSS in the same orientation, inversion of intervening sequences will take place, leading to V-D-J junctions with inverted D-segments that further increase barcode diversity.

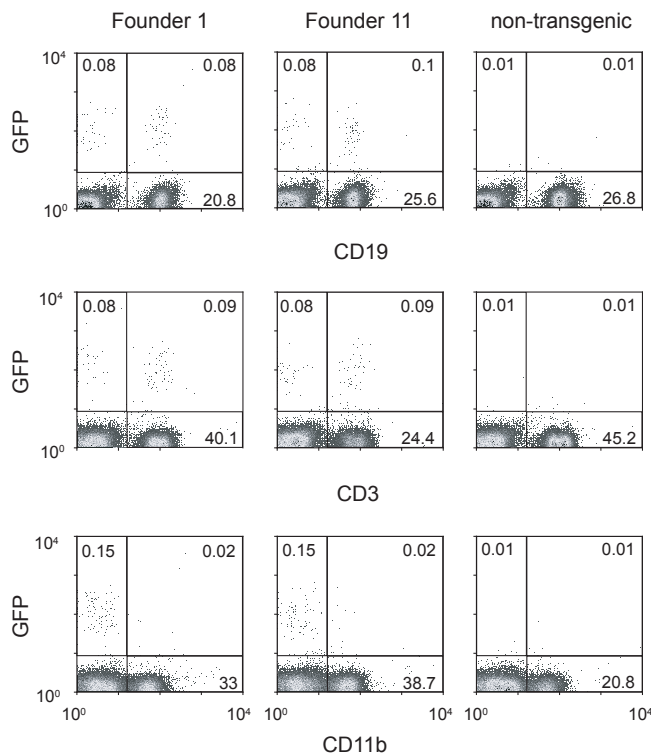


Figure 2. Recombination of the BCM transgene in B and T cells. Peripheral blood was drawn from two BCM mice that were derived from different founders (Founder #1 and #11), as well as from a non-transgenic littermate. Flow cytometry plots indicate the percentage of recombined (GFP⁺) cells among B cells (CD19⁺), T cells (CD3⁺) and myeloid cells (CD11b⁺). Note that precursor B and T cells recombine the BCM transgene, because these cells express RAG during the generation of their antigen receptors.

In an attempt to increase barcode diversity, we modified the VJ recombination substrate by inserting an additional 17-bp D segment flanked by two 23-RSSs and changed the 23-RSS flanking the V-segment into a 12-RSS, resulting in a final V-12-23-D-23-12-J configuration (Fig. 1A). This configuration, which is analogous to the IgH chain locus, makes use of two 12/23-RSS pairs that are placed in opposite orientation and give rise to two recombination junctions (namely V-to-D and D-to-J joining). Therefore, recombination of this substrate leads to two independent DNA diversification events, which due to their close proximity can be read as one barcode sequence (the unmodified recombined V-D-J stretch is 40 bp).

A useful feature of RSSs placed in opposite orientation is that any intervening DNA sequence will be spliced out during recombination. We made use of this property by inserting a bovine growth hormone (BGH) polyadenylation (polyA) stop cassette in between the D- and J-segment and cloning a cDNA coding for eGFP 3' of the J segment. In this way, GFP expression will only start after deletion of the intervening polyA stop cassette, allowing GFP expression to be used as a marker for successfully recombined cells (Fig. 1A). Furthermore, we inserted the TdT-

RAG cassette in between the V- and D-segment, such that during recombination of the VDJ substrate the RAG recombinase will splice out its own coding cDNA, making the cassette self-inactivating. This is an important measure, as transgenic (over)expression of RAG has been shown to be toxic for cells²⁶, so expression of these enzymes should preferably be kept to a minimum. In contrast, overexpression of TdT does not appear deleterious²⁷.

To allow temporal control of the induction of recombination, we cloned the TdT-RAG cassette in an antisense orientation and flanked the cassette by opposite-orientation loxP sites²⁸. Due to these opposite loxP sites, Cre-mediated recombination will lead to an inversion of the TdT-RAG cassette, resulting in expression of these enzymes. Recombination and diversification by RAG and TdT will give rise to a barcode V-D-J junction and expression of GFP, whereas the intervening TdT-RAG and polyA stop cassette will be deleted (Fig. 1A). Notably, in case the V 12-RSS recombines with the D 3' 23-RSS instead of the 5' 23-RSS, recombination will lead to an initial inversion of intervening sequences rather than a deletion (Fig. 1B). As a result, the final V-D-J junction will contain an inverted D-segment, thereby further

increasing the total barcode diversity that can be generated.

Generating the BarCode Mouse

After demonstrating the functionality of the VDJ-recombination substrate in vitro (not shown), the construct was cloned into a backbone containing a CMV enhancer and chicken β -actin promoter (Fig. 1), and targeted into the Rosa26 locus of 129/Ola embryonic stem (ES) cells²⁹. The Rosa26 locus is ubiquitously expressed in mice and therefore presumably allows accessibility of the RSSs (a prerequisite for the induction of V(D)J recombination) in a broad array of tissues. Following injection of two separate ES cell clones into C57Bl/6 blastocysts, 28 transgene-positive chimaeric mice were born. Two of these chimaeras (#1 and #11) were used as founders to establish stable transgenic lines, hereafter referred to as the BarCode Mouse (BCM). Both BCM lines show a comparable efficiency of recombination of the VDJ substrate, as identified by an equal percentage of GFP⁺ cells among peripheral blood B and T cells (\sim 0.4% of B cells and \sim 0.3% of T cells are GFP⁺; Fig. 2). Note that a fraction of B and T cells is GFP⁺ because precursor lymphocytes transiently express the V(D)J recombinase to generate their antigen receptors. Next to B and T cells, some GFP⁺ cells were also found among CD11b⁺ myeloid cells,

albeit at a very low frequency (\sim 0.05% is GFP⁺; Fig. 2), which might represent the progeny of multipotent hematopoietic progenitors that prematurely expressed the V(D)J recombinase²⁵.

To assess the extent of DNA diversification in recombined BCM B and T cells, barcodes from peripheral blood cells were amplified by PCR, individually cloned and sequenced. As depicted in table 1, 15 out of 24 sequences analyzed were unique, all of which contained nucleotide deletions (mean of 9 deletions/barcode) and 87% contained nucleotide insertions (mean of 5 insertions/barcode). This shows that recombination of the BCM transgene in peripheral blood cells leads to the generation of diverse barcode sequences that moreover contain a high degree of "uniqueness", due to the many de novo insertions. This is an important feature, as it decreases the chance that an independent precursor will generate the same barcode sequence during recombination, in which case false-positive kinship interpretations could occur. Interestingly, as theoretically predicted (Fig. 1B), one sequence (#1) showed an inversion of the D-segment, demonstrating that while deletional recombination of the VDJ substrate is the dominant event, inversional recombination can also occur. Furthermore, because the rarity of this event, barcodes

Table 1. Barcodes of peripheral blood B and T cells

Germline	V region AGTCCAGTAG	N or P	D region TCTACTATCGTTACGAC GTCGTAACGATAGTAGA	N or P	J region GTAGCTACTACCG	#
1	AGTCCA----	CTCT	GTCGTAACGATAGTAG-		--AGCTACTACCG	*
2	AGTCCAGTAG		-----CGTTACGAC		---GCTACTACCG	(3)
3	AGTCCAGTA-		-----TCGTTACGAC		GTAGCTACTACCG	(3)
4	AGTCCA----	CTTGGA	TCTACTATCGT-----	CC	GTAGCTACTACCG	
5	AGTCCAG---	CCCA	TCTACTATCGTTACGAC		GTAGCTACTACCG	
8	AGTCCAGTA-	AAGG	-----TATCGTTACGAC		-----TACTACCG	(3)
10	AGTCCAG---	ATT	TCTACTATCGTTA----	G	---GCTACTACCG	
11	AGTCCAGT--	GAGAGA	TCTACTATCGTTACG--		---GCTACTACCG	
12	AGTCCAG---	CCCCCC	-----TTACG--	CC	GTAGCTACTACCG	
13	AGTCCAG---	A	TCTACTATCGTTACGAC	G	-----TACTACCG	
14	AGTCCA----	CCG	---ACTATCGTTACGAC		-TAGCTACTACCG	(2)
15	AGTCCA----	CTTGAC	-CTACTATCGT-----	CC	GTAGCTACTACCG	
16	AGTCCAGT--	CGGGA	TCTACTATCGT-----	CC	GTAGCTACTACCG	
17	AGTCCAGT--	CG	TCTACTATCGTTACGAC	GAG	---GCTACTACCG	
18	AGTCCAGT--	TCTC	TCTACTATCGTTACG--	GAGGT	-----TACCG	(3)

Barcodes present in pooled peripheral blood of two 10-week-old BCM mice were amplified by nested PCR, cloned into pCR2.1-TA, and sequenced. Germline sequence refers to the expected VDJ junction in the absence of nucleotide diversification. N or P refers to inserted nucleotides that can either be non-templated N-nucleotides (resulting from TdT activity) or templated P-nucleotides (resulting from off-center hairpin opening by Artemis). Asterisk (*) indicates a barcode with inverted D region. Numbers in parentheses indicate the frequency of recurrent barcodes.

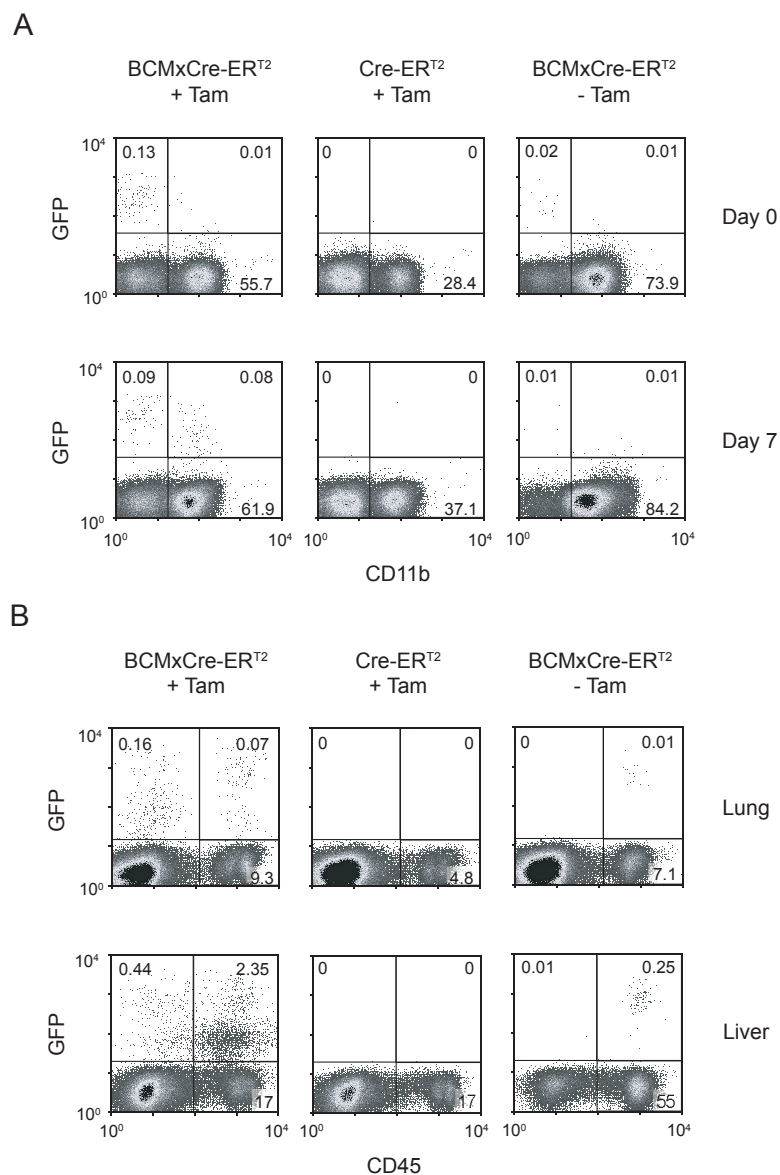


Figure 3. Conditional activation of in vivo barcoding in non-lymphoid cells. (A) BCM mice were crossed to mice that harbor a tamoxifen (Tam)-inducible Cre-ER^{T2} transgene in the Rosa26 locus. Tamoxifen was administered for five consecutive days to both a BCMxCre-ER^{T2} mouse as well as a Cre-ER^{T2} littermate, to induce expression of Cre recombinase and drive recombination of the BCM transgene. At day 0 (before start of treatment) and day 7, blood samples were drawn from both mice, as well as from an untreated BCMxCre-ER^{T2} littermate, to assess the effect of tamoxifen administration. Flow cytometry plots illustrate that the induction of recombination in CD11b⁺ myeloid cells is clearly dependent on tamoxifen. **(B)** At day 8, from the mice described under A, lungs and livers were isolated, digested and cell suspensions were stained with anti-CD45 antibodies to separate cells of hematopoietic and non-hematopoietic origin. Flow cytometry plots indicate that the recombination of non-hematopoietic (CD45⁻) lung and liver cells is controlled by tamoxifen administration.

with inverted D-segments are likely to have high statistical power in kinship experiments. Together, the above data indicate the feasibility of in vivo barcoding in cells of lymphoid origin.

Conditional activation of in vivo barcoding in non-lymphoid cells

To test whether in vivo barcoding can also be applied to non-lymphoid cells, we crossed BCM mice to mice bearing an inducible Cre recombinase construct in the Rosa26 locus (Cre-ER^{T2} mice)³⁰. Cre-ER^{T2} encodes a fusion construct between Cre and a mutated ligand-binding domain of the estrogen receptor (ER^{T2}). In this configuration, Cre is sequestered in the cytoplasm and can only enter the nucleus and induce recombination in the presence of the ER-ligand tamoxifen. Thus,

by administering tamoxifen to the progeny of BCMxCre-ER^{T2} mice, the TdT-RAG cassette can be conditionally activated.

To determine whether this approach allows in vivo barcoding in non-lymphoid cells, a BCMxCre-ER^{T2} mouse received tamoxifen for five consecutive days, after which the induction of GFP⁺ cells was monitored in peripheral blood. As controls, a Cre-ER^{T2}-only mouse also received tamoxifen, while another BCMxCre-ER^{T2} mouse was left untreated. One week after start of treatment, a clear GFP⁺ population was observed among CD11b⁺ myeloid blood cells (0.13%), which was absent from the two control mice (Fig. 3A). This finding indicates that tamoxifen administration can induce recombination in myeloid hematopoietic cells.

To investigate whether tamoxifen can also induce recombination in non-hematopoietic cells, we isolated the lung and liver from each mouse at day 8 and digested these organs to obtain single cell suspensions. Subsequently, these samples were stained by anti-CD45 antibodies to discriminate between cells of hematopoietic and non-hematopoietic origin, and cells were analyzed by flow cytometry. As shown in Fig. 3B, a clear induction of recombination was found in non-hematopoietic cells of the BCMxCre-ER^{T2} mouse upon tamoxifen administration, with 0.18% of CD45⁻ lung cells and 0.54% of CD45⁻ liver cells being GFP⁺. In contrast, no GFP⁺ CD45⁻ cells were found in the untreated BCMxCre-ER^{T2} mouse, whereas a clear GFP⁺ CD45⁺ population was present, as expected. These data directly demonstrate that recombination of the BCM transgene can be induced in non-lymphoid cells in vivo, and moreover that induction of this recombination is strictly dependent on tamoxifen administration, indicating that there is no leakiness in the system.

In order to analyze the extent of DNA

diversification in the recombined lung and liver cells, CD45⁻ GFP⁺ cells of both tissues were sorted and split in two equal halves (A and B). These halves were then used for independent PCR amplification of the barcodes, after which resulting products were subjected to sequencing. As depicted in table 2, 5 out of 8 sequences of lung sample A and 8 out of 13 sequences of lung sample B were unique. All sequences contained nucleotide deletions (mean of 13.5 deletions/barcode) and 38% contained nucleotide insertions (mean of 0.5 insertions/barcode), of which most are potential palindromic P nucleotide insertions. This near-absence of nucleotide insertions in non-hematopoietic lung cells is in striking contrast to the mean of 5 insertions/barcode that was found in BCM peripheral blood B and T cells (Table 1). As we have also observed this disparity in TdT-activity between lymphoid and non-lymphoid cells in a previously generated transgenic mouse that carried a premature version of the BCM construct (not shown), we believe that full functionality of TdT might require additional unknown factors

Table 2. Barcodes of non-hematopoietic lung and liver cells

Germline	V region AGTCCAGTAG	N or P	D region TCTACTATCGTTACGAC GTCGTAACGATAGTAGA	N or P	J region GTAGCTACTACCG	#
Lung A						
1	AGTCCAG---	A	TCTACTATCGTTA----		---GCTACTACCG	
3	AGTCCA----		TCTACTATCGTTACG--		GTAGCTACTACCG	(3)
11	AGTCC-----		-----		-----TACCG	(2)
19	AGTCCAGTA-		-----TCGTTACGA-		---GCTACTACCG	
20	AGTCCAGT--		-----GTTAC---		-TAGCTACTACCG	
Lung B						
25	AGTCCA----		-----TATCGTTACGAC	G	----CTACTACCG	
26	AGTCCA----		-----TATCGTTACG--		-TAGCTACTACCG	(2)
27	AGTCCAGTA-		T-----	C	GTAGCTACTACCG	(3)
30	AGTCCAGTA-	T	--TACTATCG-----		GTAGCTACTACCG	(3)
32	AGTCCAGTAG		-----TATCGTTACGAC		GTAGCTACTACCG	
33	AGTCCA----		-CTACTATCGTTACGA-		---GCTACTACCG	
34	AGTCCAGTA-		-----ACGATA-----		-----G	*
36	AGTCCAGTA-		-----TCGTTACGAC	G	----CTACTACCG	
Liver A						
49	AGTCCA----		-----CGT-----		GTAGCTACTACCG	(15)
Liver B						
73	AGTCCAGTAG	C	TCTACTATCGTTACGAC	G	-----TACTACCG	(14)
75	AGTCCAG---		--ACTATCGTTACGAC		GTAGCTACTACCG	

A BCMxCre-ER^{T2} mouse received 2 mg tamoxifen intraperitoneally for 5 consecutive days. At day 8, lung and liver tissues were digested by collagenase/DNase treatment and GFP⁺ CD45⁻ cells were sorted. From these cells, barcodes were amplified by nested PCR, cloned into pCR2.1-TA, and sequenced. Germline sequence refers to the expected VDJ junction in the absence of nucleotide diversification. N or P refers to inserted nucleotides that can either be non-templated N-nucleotides (resulting from TdT activity) or templated P-nucleotides (resulting from off-center hairpin opening by Artemis). Asterisk (*) indicates a barcode with inverted D region. Numbers in parentheses indicate the frequency of recurrent barcodes.

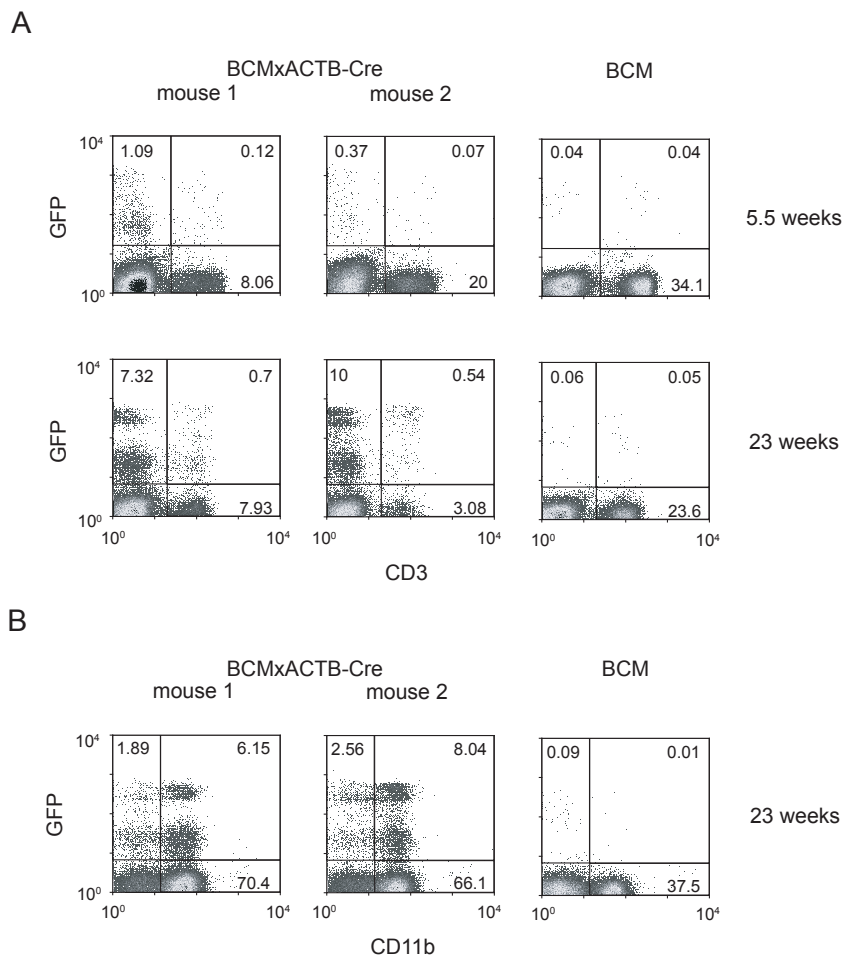


Figure 4. Ubiquitous induction of in vivo barcoding. (A) BCM mice were crossed to mice that ubiquitously express Cre recombinase under control of the human β -actin promoter (ACTB-Cre mice). At 5.5 and 23 weeks of age, blood samples were drawn from two BCMxACTB-Cre mice as well as from one BCM littermate, and samples were stained with anti-CD3 antibodies. Flow cytometry plots depict the time-dependent increase in the percentage of recombined (GFP⁺) T cells. (B) At 23 weeks, the same samples as described under A were stained with anti-CD11b antibodies. Flow cytometry plots indicate that the majority of recombined cells in BCMxACTB-Cre mice are found in the myeloid compartment.

that are absent in non-lymphoid cells.

In contrast to barcodes recovered from the lung, only 1 out of 15 sequences of liver sample A and 2 out of 15 sequences of liver sample B were unique (Table 2). Although this could be a reflection of preferential recombination events, we consider this an unlikely explanation as none of these highly prevalent sequences were present in any of the other independent samples analyzed. Rather, this type of recurrent barcodes is indicative of an inefficient and biased PCR amplification, in which certain sequences preferentially amplified early during the PCR reaction can have a large contribution to the final outcome. Efforts are currently underway to improve the general barcode recovery. Despite this potential inefficient recovery, no shared barcode sequences were found between any of the unrelated samples, suggesting that (albeit based on small numbers) there is a clear experimental window in the current system that allows related cells to be distinguished from unrelated cells. In summary, the above results

demonstrate that in vivo barcoding can be conditionally induced in non-lymphoid cells.

Ubiquitous induction of in vivo barcoding

Although the above data clearly demonstrated the feasibility of in vivo barcoding, successful recombination appeared restricted to less than 1% of target cells. To test whether this low number of recombinants is due to a limit in the number of cells that is amenable to undergo recombination or whether it is due to inefficiency in the recombination process itself, we crossed BCM mice to mice that ubiquitously express Cre recombinase under control of the human β -actin promoter (ACTB-Cre mice). Analysis of peripheral blood T cells of BCMxACTB-Cre mice revealed a clear time-dependent increase in the number of GFP⁺ T cells, with a mean of 0.91% GFP⁺ CD3⁺ cells being present at 5.5 weeks of age and 11.5% GFP⁺ CD3⁺ cells being present at 23 weeks (Fig. 4A). It should be noted that this time-dependent increase in the percentage of recombined cells was accompanied by an

85% reduction in the overall number of CD3⁺ cells in BCMxACTB-Cre mouse #2. Furthermore, we have observed a complete loss in peripheral blood CD19⁺ B cells in several BCMxACTB-Cre mice as early as 8 weeks of age (not shown), suggesting that constitutive activation of the BCM transgene is generally toxic for B and T cells. This observation is reminiscent of the phenotype of RAG transgenic mice, which display severe lymphopenia that is most prominent in the B cell compartment²⁶. It is at present unclear what the mechanism behind this selective loss of B and T cells is. One potential explanation is that continuous rearrangement in hematopoietic progenitor cells results in loss of (part of) the antigen receptor loci, leading to a developmental block in lymphocyte generation²⁶. Alternatively, continuous rearrangement in mature lymphocytes could lead to antigen receptor ablation, which has been shown to lead to a rapid decay in peripheral lymphocyte numbers^{31,32}.

In contrast to what was observed for B and T cells, CD11b⁺ myeloid cells appeared unaffected by overexpression of the transgene and were present at relatively increased frequencies in these mice (Fig. 4B). As a result, CD11b⁺ cells made up the majority of GFP⁺ cells in the periphery of BCMxACTB-Cre mice, with a mean of 10.3% GFP⁺ CD11b⁺ cells being present at 23 weeks. This number is in the same range as the percentage GFP⁺ CD3⁺ cells, which might suggest that a large part of these GFP⁺ cells were descendants of common ancestors that recombined the BCM transgene at the hematopoietic stem or early progenitor cell level. Comparison of barcodes recovered from peripheral blood and progenitor cells should reveal whether this is indeed the case. Notably, three distinct levels of GFP expression can be observed in these BCMxACTB-Cre mice (a dull, bright and very bright GFP⁺ population, Fig. 4B). Discrimination between the bright- and very bright-GFP signal is potentially the result of distinct promoter activity in different immune cell subsets, as CD11b⁺ GFP^{very bright} cells are SSC^{high}, whereas CD11b⁺ GFP^{bright} cells are SSC^{low} (not shown). Additional phenotypic analysis should reveal the precise identity of these different populations. In contrast, the GFP^{dull} population

is potentially the result of an incomplete recombination event, in which only the V- and D-segment but not the D- and J-segment were joined. As a result, the intervening TdT-RAG cassette was spliced out, while the polyA stop cassette in between the D- and J-segment remained, resulting in a ~10-fold lower GFP signal. PCR amplification of the VDJ junction of sorted GFP^{dull} cells should reveal whether these cells indeed have undergone only partial recombination, in which case the amplification product is expected to be ~400 bp longer. Taken together, the above results demonstrate that constitutive activation of the BCM transgene can lead to a time-dependent increase in the recombination frequency, indicating that the general low efficiency of recombination is most likely due to an inefficiency of the recombination process itself.

DISCUSSION

In this study, we have generated a novel mouse model that can be used to trace kinship of different cell populations by the induction of unique genetic tags in vivo. This “in vivo barcoding” approach makes use of Cre/lox-mediated recombination to conditionally activate expression of the V(D)J recombinase enzymes RAG1, RAG2 and TdT. Together, these enzymes drive recombination of a pseudo-VDJ substrate, which via the deletion and insertion of nucleotides leads to a heritable genetic signature in each cell (a DNA barcode), and GFP expression as a marker. We have demonstrated the feasibility of in vivo barcoding by conditionally inducing recombination in non-hematopoietic lung and liver cells and showing that unrelated cells are identified as such by the presence of different barcode sequences. Currently, we are further exploring the potential of this technology, by assessing the total barcode complexity that can be generated by this system with high-throughput sequencing.

What kind of questions might be addressed by in vivo barcoding? In general, cell marking by genetic tags can be useful to investigate two categories of biological issues: I) kinship issues and II) clonal diversity issues. Kinship issues deal with the question whether cell

populations that differ in location of functional activity arise from common or separate precursors. A well-known example of this type of question is how all the different immune cell types that can be found in the blood develop? Although initially all blood cells are derived from hematopoietic stem cells, as soon as these stem cells start to differentiate, they will give rise to different progenitors that can still yield some but no longer all blood cell types. Classically, the developmental potential of hematopoietic progenitors is assessed by transferring selective populations of progenitor cells into irradiated recipients and asking which cell types can still be generated by these cells^{33,34}. A limitation of this approach is that it not only isolates cells out of their normal niche (the bone marrow) and places them in a different compartment (the blood); it also changes the host environment in which these cells normally develop (now an "empty" bone marrow). In such system, the observed cell fate decisions are not necessarily reflective of natural steady-state hematopoiesis. In vivo barcoding has the potential to circumvent such issues, as differentiation of hematopoietic stem and progenitor cells can now be studied without the need to isolate these cells from their normal environment. By crossing BCM mice to mice that express Cre recombinase under control of lineage-specific promoters (such as Runx1-Cre or PU.1-Cre mice), it should be feasible to address whether the current cell fate maps that describe hematopoietic development also hold true under natural steady-state conditions.

As we have demonstrated that in vivo barcoding is not limited to cells of hematopoietic origin, a similar approach could be taken towards the goal of generating a complete cell fate map of mammalian development at single-cell resolution. One way to achieve this would be by treating pregnant mothers of BCMxCre-ER^{T2} embryos with tamoxifen at a given time point during embryonic development (e.g. E8.5). Administering tamoxifen in this way has been shown to induce Cre recombinase in developing embryos and to drive recombination of a reporter gene in derivatives of all three germ layers³⁵. In theory, by inducing in vivo barcoding during early embryonic development one could assess at what embryonic stage different cell types that

make up an organ or different organs that make up a mouse become developmentally separated.

A second category of questions that in vivo barcoding can address are clonal diversity issues, which typically ask how many progenitor cells contribute to the development or homeostasis of a given tissue or cell population. A well known example of this type of question is how clonally diverse different tumor populations are? It is still largely unclear whether all cells in a given tumor mass proliferate and contribute to the next cell generation or whether only a fraction of all cells have this property, the latter often being referred to as cancer stem cells³⁶. Classically, this question is addressed by isolating and digesting existing tumors and transferring small subpopulations to recipient mice, following which it is tested whether all or only some founder cells are able to establish new tumors. Again, this experimental procedure involves a change in tumor cell environment, which could potentially alter subsequent cell behavior. In theory, by combining the BCM transgene with a mouse model that uses Cre recombinase to excise floxed tumor-suppressor genes (e.g. BRCA1 floxed x p53 floxed), this system should allow both enhanced spontaneous tumor development as well as marking of individual transformed cells by barcode sequences. By monitoring barcode diversity in tumor populations of different mice in time, it could then be distinguished whether all tumor cells contribute to tumor development and maintenance (barcode diversity will remain constant) or whether the tumor population becomes dominated by the output of a small number of cancer stem cells (barcode diversity will decline). Along similar lines, it should also be feasible to use in vivo barcoding to investigate stem cell dynamics in non-malignant tissues.

In summary, in vivo barcoding is a promising new technology that has the potential to contribute greatly to our current understanding of how cells differentiate from precursor levels to a functionally mature state, with the added value that differentiation can be traced in the natural cell environment. The increased access of many researchers to high-throughput sequencing technology should warrant the broad applicability of this mouse model. Future studies should reveal whether

the total barcode complexity generated by the current system is sufficient to address most experimental questions or whether the construct design should be extended by additional RSS-flanked gene segments to further enhance the combinatorial barcode diversity.

MATERIALS AND METHODS

DNA constructs

The plasmids pEF-BOS-RAG1 and pEF-BOS-RAG2 were kindly provided by Dr. K. Vanura (University of Vienna, Austria). The codon-optimized sequence for TdT (short isoform) was ordered at GeneArt (Regensburg, Germany). To assemble the recombinase part of the BCM construct, the self-cleaving 2A peptide sequences E2A and T2A²³ were generated by PCR and respectively ligated 5' and 3' of RAG2 by using MscI and PflMI restriction sites. The resulting "E2A-RAG2-T2A" fragment was cloned into a pBluescript vector using SacII and NotI. Then, RAG1 was cloned 3' of T2A using NcoI and NotI. The "E2A-RAG2-T2A-RAG1" fragment was cloned into a pVAX vector using AflII and NotI. Next, TdT was inserted 5' of E2A using AflII and XmaI, to generate pVAX-TdT-E2A-RAG2-T2A-RAG1. Finally, two opposite-orientation loxP sites were respectively introduced 5' of TdT by MfeI and BamHI and 3' of RAG1 by SalI and NotI.

The design of the substrate part of the BCM construct was based on plasmid pSBEX3IB containing a GFP-based VJ recombination substrate²⁵ that was kindly provided by Dr. R. Gerstein (University of Massachusetts Medical School, USA). To assemble the VDJ BCM substrate, a "12 RSS-J segment" fragment was generated by PCR and cloned into pBluescript by SacI and KpnI. The sequence for eGFP was released from a Rosa26 targeting vector and ligated 3' of the J-segment by NcoI and SacII. Next, a "V-segment-12 RSS" fragment was generated by PCR and inserted 5' of the J-segment by SacI and EcoRI. Then, a "23 RSS-D segment-23 RSS-bovine growth hormone (BGH) polyA signal" fragment was generated by PCR and cloned in between the V- and J-segment by BglII and EcoRI. Spacer sequences of the two D-segment 23-RSSs were varied to prevent hairpin formation. D-segment sequence is the naturally occurring IgH DSP2.4, with the exception that the ATG sequence was mutated into ATC, to prevent premature translational initiation. The complete "V-segment-12 RSS-23 RSS-D-segment-23 RSS-BGH polyA-12 RSS-J segment-GFP" fragment was cloned into a Rosa26 targeting vector containing a CMV enhancer and chicken β -actin promoter by AscI. Finally, the "loxP-TdT-E2A-RAG2-T2A-RAG1-loxP" cassette was inserted in antisense orientation 3' of the V-segment by PmlI and MfeI, to give rise to the BCM construct as depicted in Fig. 1A.

Mice

The Rosa26 targeting vector containing the BCM construct was linearized by PvuI and electroporated into IB10 E14 129/ola ES cells. Stable transfectants were selected with puromycin and resistant clones were picked and expanded. Correct integration was determined by Southern blotting using a probe directed against the 5' Rosa26 homology arm. Two independent ES cell clones were injected into C57Bl/6 blastocysts to generate 28 transgene-positive chimaeric mice, which were used to establish two independent BCM transgenic lines. Genotyping was performed by staining peripheral blood samples by anti-CD19-PE (BD, clone 1D3), anti-CD3e-PerCP-Cy5.5 (eBioscience, clone 145-2C11) and anti-CD11b-APC (BD, clone M1/70) and screening for GFP⁺ B and T cells by flow cytometry. BCM mice were crossed with Cre-ER^{T2} mice that harbor a Cre-ER^{T2} construct in the Rosa26 locus³⁰. Offspring received 2 mg tamoxifen (Sigma T5648) in 98% sunflower seed oil (Sigma S5007)/ 2% ethanol by intraperitoneal injection for 5 consecutive days. BCM mice were also crossed to ACTB-Cre mice that express Cre recombinase under control of the human β -actin promoter in all cells of the embryo by the blastocyst stage.

Barcode analysis of peripheral blood

Peripheral blood samples of BCM mice were treated with ammonium chloride to remove erythrocytes. Genomic DNA was isolated using a DNeasy tissue kit (Qiagen) and barcode sequences were amplified by nested PCR using Taq polymerase (Invitrogen) and the following primers (5' to 3'): top -ACTCACTATAGGGAGACGCGTGTACC- and bottom -GACACGCTGAAGTATCAAG- in the first round, with 62°C annealing; top -CCTCGAGGTCATCGAAGTATCAAG- and bottom -CGTCCAGCTCGACCAGGAT- in the second round, with 60°C annealing. The resulting PCR products (~170 bp) were cloned using TOPO TA cloning (Invitrogen) and individual colonies were analyzed by BigDye Terminator sequencing (Applied Biosystems).

Barcode analysis of non-lymphoid organs

BCMxCre-ER^{T2} and control mice were sacrificed at day 8 after start of tamoxifen treatment. Mice were perfused with PBS and lungs and livers were harvested. Organs were minced with razor blades and digested with 3.5 mg/ml collagenase IV (Worthington)/ 3.5 μ g/ml DNaseI (Roche) in PBS for 1 h at 37°C while shaking. Resulting cell suspensions were filtered through a 100 μ m cell strainer (BD Falcon), washed, stained with anti-CD45.2-APC (eBioscience, clone 104) and GFP⁺ CD45⁻ cells were sorted on a FACSAria (BD) with 100 μ m nozzle. Barcodes from sorted cells were analyzed as indicated above.

ACKNOWLEDGMENTS

We thank R. Bin Ali for blastocyst injections; R. Gerstein (University of Massachusetts Medical School, USA) for provision of pSBEX3IB containing a GFP-based VJ recombination substrate; K. Vanura (University of Vienna, Austria) for provision of pEF-BOS-RAG1 and pEF-BOS-RAG2; S. Gilfillan (Washington University School of Medicine, USA) for provision of pBluescript-TdT; I. Nijman and E. Cuppen (Hubrecht Laboratory, the Netherlands) for analysis of sequence data and P. Krimpenfort and J. Jonkers for advice on BCM construct design. This work was financially supported by a start-up grant from the Netherlands Cancer Institute.

REFERENCE LIST

- Sulston, J. E. & Horvitz, H. R. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev. Biol.* **56**, 110-156 (1977).
- Sulston, J. E., Schierenberg, E., White, J. G. & Thomson, J. N. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **100**, 64-119 (1983).
- Wu, A. M., Till, J. E., Siminovitch, L. & McCulloch, E. A. A cytological study of the capacity for differentiation of normal hemopoietic colony-forming cells. *J. Cell Physiol.* **69**, 177-184 (1967).
- Wu, A. M., Till, J. E., Siminovitch, L. & McCulloch, E. A. Cytological evidence for a relationship between normal hemopoietic colony-forming cells and cells of the lymphoid system. *J. Exp. Med.* **127**, 455-464 (1968).
- Abramson, S., Miller, R. G. & Phillips, R. A. The identification in adult bone marrow of pluripotent and restricted stem cells of the myeloid and lymphoid systems. *J. Exp. Med.* **145**, 1567-1579 (1977).
- Dick, J. E., Magli, M. C., Huszar, D., Phillips, R. A. & Bernstein, A. Introduction of a selectable gene into primitive stem cells capable of long-term reconstitution of the hemopoietic system of W/W^v mice. *Cell* **42**, 71-79 (1985) 7 .
Keller, G., Paige, C., Gilboa, E. & Wagner, E. F. Expression of a foreign gene in myeloid and lymphoid cells derived from multipotent haematopoietic precursors. *Nature* **318**, 149-154 (1985).
- Lemischka, I. R., Raulet, D. H. & Mulligan, R. C. Developmental potential and dynamic behavior of hematopoietic stem cells. *Cell* **45**, 917-927 (1986).
- Walsh, C. & Cepko, C. L. Widespread dispersion of neuronal clones across functional regions of the cerebral cortex. *Science* **255**, 434-440 (1992).
- Golden, J. A., Fields-Berry, S. C. & Cepko, C. L. Construction and characterization of a highly complex retroviral library for lineage analysis. *Proc. Natl. Acad. Sci. U S A* **92**, 5704-5708 (1995).
- Schepers, K. et al. Dissecting T cell lineage relationships by cellular barcoding. *J. Exp. Med.* **205**, 2309-2318 (2008).
- van Heijst, J. W. et al. Recruitment of antigen-specific CD8⁺ T cells in response to infection is markedly efficient. *Science* **325**, 1265-1269 (2009).
- Gerlach, C. et al. One naïve T cell, multiple fates in CD8⁺ T cell differentiation. *J. Exp. Med.* In press (2010).
- Roe, T., Reynolds, T. C., Yu, G. & Brown, P. O. Integration of murine leukemia virus DNA depends on mitosis. *EMBO J.* **12**, 2099-2108 (1993).
- Salipante, S. J. & Horwitz, M. S. Phylogenetic fate mapping. *Proc. Natl. Acad. Sci. U S A* **103**, 5448-5453 (2006).
- Wasserstrom, A. et al. Reconstruction of cell lineage trees in mice. *PLoS One* **3**, e1939 (2008).
- Gellert, M. V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu. Rev. Biochem.* **71**, 101-132 (2002).
- Roth, D. B. Restraining the V(D)J recombinase. *Nat. Rev. Immunol.* **3**, 656-666 (2003).
- Lafaille, J. J., DeCloux, A., Bonneville, M., Takagaki, Y. & Tonegawa, S. Junctional sequences of T cell receptor gamma delta genes: implications for gamma delta T cell lineages and for a novel intermediate of V-(D)-J joining. *Cell* **59**, 859-870 (1989).
- Komori, T., Okada, A., Stewart, V. & Alt, F. W. Lack of N regions in antigen receptor variable region genes of TdT-deficient lymphocytes. *Science* **261**, 1171-1175 (1993).
- Gilfillan, S., Dierich, A., Lemeur, M., Benoist, C. & Mathis, D. Mice lacking TdT: mature animals with an immature lymphocyte repertoire. *Science* **261**, 1175-1178 (1993).
- Tonegawa, S. Somatic generation of antibody diversity. *Nature* **302**, 575-581 (1983).
- Szymczak, A. L. et al. Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. *Nat. Biotechnol.* **22**, 589-594 (2004).
- Borghesi, L. & Gerstein, R. M. Developmental separation of V(D)J recombinase expression and initiation of IgH recombination in B lineage progenitors in vivo. *J Exp Med* **199**, 483-489 (2004).
- Borghesi, L. et al. B lineage-specific regulation of V(D)J recombinase activity is established in common lymphoid progenitors. *J. Exp. Med.* **199**, 491-502 (2004).
- Barreto, V., Marques, R. & Demengeot, J. Early death and severe lymphopenia caused by ubiquitous expression of the Rag1 and Rag2 genes in mice. *Eur J Immunol* **31**, 3763-3772 (2001).
- Bentolila, L. A. et al. Constitutive expression of terminal deoxynucleotidyl transferase in transgenic mice is sufficient for N region

- diversity to occur at any Ig locus throughout B cell differentiation. *J Immunol* **158**, 715-723 (1997).
28. Sauer, B. Inducible gene targeting in mice using the Cre/lox system. *Methods* **14**, 381-392 (1998).
 29. Soriano, P. Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nat Genet* **21**, 70-71 (1999).
 30. Hameyer, D. et al. Toxicity of ligand-dependent Cre recombinases and generation of a conditional Cre deleter mouse allowing mosaic recombination in peripheral tissues. *Physiol Genomics* **31**, 32-41 (2007).
 31. Lam, K. P., Kuhn, R. & Rajewsky, K. In vivo ablation of surface immunoglobulin on mature B cells by inducible gene targeting results in rapid cell death. *Cell* **90**, 1073-1083 (1997).
 32. Polic, B., Kunkel, D., Scheffold, A. & Rajewsky, K. How alpha beta T cells deal with induced TCR alpha ablation. *Proc Natl Acad Sci U S A* **98**, 8744-8749 (2001).
 33. Osawa, M., Hanada, K., Hamada, H. & Nakauchi, H. Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science* **273**, 242-245 (1996).
 34. Matsuzaki, Y., Kinjo, K., Mulligan, R. C. & Okano, H. Unexpectedly efficient homing capacity of purified murine hematopoietic stem cells. *Immunity* **20**, 87-93 (2004).
 35. Hayashi, S. & McMahon, A. P. Efficient recombination in diverse tissues by a tamoxifen-inducible form of Cre: a tool for temporally regulated gene activation/inactivation in the mouse. *Dev Biol* **244**, 305-318 (2002).
 36. Frank, N. Y., Schatton, T. & Frank, M. H. The therapeutic promise of the cancer stem cell concept. *J Clin Invest* **120**, 41-50 (2010).