Dissection and manipulation of antigen-specific T cell responses

Schepers, K.

Chapter 5

Dissecting lineage relationships between T cells by cellular barcoding

Koen Schepers, Erwin Swart, Jeroen W.J. van Heijst, Carmen Gerlach, Maria Castrucci, Mike Heimerikx, Arno Velds, Ron M. Kerkhoven, Ramon Arens, Ton N.M. Schumacher.

Unpublished

Dissecting lineage relationships between T cells by cellular barcoding

Koen Schepers[1], Erwin Swart[1], Jeroen W.J. van Heijst[1], Carmen Gerlach[1], Maria Castrucci[2], Mike Heimerikx[3], Arno Velds[3], Ron M. Kerkhoven[3], Ramon Arens[1], Ton N.M. Schumacher[1].

1. Division of Immunology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX, Amsterdam, The Netherlands. 2 Istituto Superiore di Sanità, Roma, Italy. 3. Division of Molecular Carcinogenesis and Center for Biomedical Genetics, The Netherlands Cancer Institute, Amsterdam, The Netherlands.

Over the past years it has become increasingly clear that antigen-specific T cell responses are composed of phenotypically and functionally distinct subsets. It is however unclear how these different T cell populations develop from a pool of naïve T cells. To address this issue and other issues related to (T) cell differentiation, we have developed a novel approach for the dissection of lineage relationships between cell populations. In this approach, individual cells are labeled with a molecular barcode consisting of a semi-random stretch of DNA by retroviral transduction. Here we demonstrate that this labeling of individual cells with unique identifiers coupled to a microarray-based detection system can be used to analyze family relationship between the progeny of such cell populations. Furthermore, assessment of kinship by cellular barcoding demonstrated that antigen-specific T cells found at different peripheral effector sites (i.e. the skin and lung) are derived from precursor pools that are largely overlapping.

Cytotoxic (CD8+) and helper (CD4+) T cells detect peptide fragments that are presented on the cell surface by Major Histocompatibility Complex (MHC) class I and II molecules, respectively. Within the antigen-inexperienced T cell repertoire, the frequency of both CD4+ and CD8+ T cells that recognize a given antigen/MHC combination is extremely low (0.01% or less). However, upon encounter of antigen in combination with the appropriate co-stimulatory signals, T cells display a massive proliferative burst and differentiate into a large pool of effector T cells that have acquired the ability to perform effector functions and to migrate into peripheral organs. It is becoming increasingly clear that, rather than being a single homogenous pool of effector T cells, both CD4+ and CD8+ effector pools consist of a number of subsets that can differ in functional and migratory capacities. For instance, CD4+ T cells can develop into so called Th0, Th1 and Th2 cells, which are distinguished based on the patterns of cytokines that they produce[1,2], and both CD4+ and CD8+ T cells have been divided into subsets that differ in their capacity to migrate to specific inflammatory sites[3]. While there is now significant evidence for functional di- (or poly-) chotomy in T cell

responses, in many cases it is unclear at which point in T cell differentiation branching occurs. In a first model, naïve T cells may have a predetermined capacity to produce progeny that differentiates towards a given subset, or such a capacity may be instilled at the time of the first antigen encounter. Evidence for the latter possibility is provided by a series of studies that demonstrate that the quality of the interaction of naïve T cells with antigen-presenting cells can influence the behavior of its progeny with respect to migration and its potential for expansion upon secondary antigen encounter[4-12]. As an alternative to models where the progeny of a given naïve T cell is destined to develop into a specific subset, signals that are received by the progeny itself could influence such differentiation. In this case, where branching occurs during or following T cell expansion, the progeny of each individual naïve T cell is expected to be present in all different subsets.
To study family relationships between T cell subsets we set out to develop a novel technique that allows one to assess whether T cells within two subsets of effector T cells share common ancestors within the naïve T cell pool. We generated a retroviral library of which each clone contains a non-coding semi-random stretch of DNA, allowing one to uniquely tag individual T cells with molecular barcodes by retroviral infection. In this paper we examine the feasibility of analyzing kinship between the progeny of tagged T cell populations and we analyze the kinship between pathogen-specific T cells that migrate to distinct effector sites.

Material and methods

Mice

C57BL/6 (B6) and C57BL/6.OT-I mice (OT-I)[20] were obtained from the animal department of the Netherlands Cancer Institute. All animal experiments were carried out in accordance with institutional and national guidelines and were approved by the Experimental Animal Committee (DEC) of the Netherlands Cancer Institute.

Barcode library and microarray generation

Using the following primers: top, 5'-GGGAAAGTCGACTAAGCCACCATGGTG AGCAAGGGCGAG -3'; and bottom, 5'-CCCGAATTCGATCTCGAATCAGGCG CTTAGGATCCTCCCCTTCCCTCGAGTGTTCTAGTTGACGGCAGCATTACTTGTACA GCTCGTCCATG -3' and the enhanced green fluorescent protein (GFP) gene as template, a PCR product containing GFP and unique cloning sites for library insertion was created and cloned into the *Xho*I and *Eco*RI sites of the retroviral vector pLentiLox3.4 (kind gift from L. van Parijs), to obtain pLentiLox3.4-GFP2. A semi-random stretch of DNA was subsequently cloned 3' of the GFP ORF in the pLentilox3.4-GFP2 vector. For this purpose the oligo: 5'-TGCTGCCGTCAACTAGAACACTCGAGNNNNNNNNSWSWSWSWSWNNNNNNNNS WSWSWSWSWNNNNNNNNSWSWSWSWSWNNNNNNNNSWSWSWSWSWNNNNN NNNSWSWSWSWSWNNNNNNNNNGGATCCTAAGCGCCTGATTCGAGATC-3' was amplified using the following primers: top-LIB primer, 5'-TGCTGCCGTCAACTAGAACA-

3', bottom-LIB primer, 5'-GATCTCGAATC AGGCGCTTA-3'. The resulting library was ligated into the *Xho*I and *Bam*HI sites of pLentiLox3.4-GFP2. Ligation products were introduced into electroporation competent *Escherichia coli* cells (electro Ten-blue, Stratagene) by electroporation and plated on 10 cm Ampicillin-containing LB plates. Subsequently, 4,743 individual *E.coli* cell clones were picked and grown in 96 well plates. This master library was used to generate a barcode-microarray and a Moloney based retroviral library. To generate the barcode-microarray, PCR products were generated from the individual *E.coli* cell clones of the master library using top-LIB and bottom-LIB primers and were spotted onto poly-L-lysine-coated glass slides using the Microgrid II arrayer (Apogent, Cambridge, United Kingdom). Details on the process of preparing the DNA for spotting and preparation of the slides are available on the CMF Web site (http://microarrays.nki.nl/download/protocols.html). For the generation of the Moloney-based pMX-GFP-bc library, DNA was isolated from the pooled *E.coli* cell clones of the master library and subcloned into a variant of the pMX retroviral vector[7] that contains a *Pac*I cloning site, using the restriction enzymes *Pac*I and *EcoR*I. To determine which barcodes can effectively be used for lineage analysis, barcodes recovered in at least one experiment were plotted as a function of the number of experiments performed. In a first set of 7 experiments 3,029 different barcodes were observed and extrapolation of this curve suggest that the number of informative barcodes is approximately 3,300. The lack of detection for the remainder is likely due to loss of barcodes during the different procedures of generating the library for retrovirus production and the microarray. In our experiments only the 3,029 barcodes that were recovered at least once in the first set of experiments were used for lineage analysis.

Retroviral transduction procedure

Retroviral supernatants were obtained by transfection of Phoenix-E packaging cells with the pMX-GFP-bc library in combination with pCLEco[21], using the FuGENE™ 6 transfection reagent (Roche Molecular Biochemicals, Mannheim, Germany) as described previously[22]. Retroviral supernatants were obtained 48 hrs post-transfection and stored at -80°C. Total OT-I splenocytes were isolated and mixed with CD4-enriched B6 splenocytes at a 10:1 ratio. CD4-enriched B6 splenocytes were obtained by staining B6 splenocytes with PE- or APC-conjugated anti-CD4 mAbs (BD Biosciences, MountainView, CA), labeling with anti-PE or anti-APC Microbeads (Miltenyi Biotec, Bergisch Gladbach, Germany) followed by Midi-MACS enrichment according to the manufacturer's protocol (Miltenyi Biotec). Enrichment was ≥ 85%. Cells were activated with concanavalin A (ConA) and IL-7 as described[22]. Retroviral supernatants were thawed and diluted to transduce concanavalin A/IL7-activated cells by spin infection (90 min at 2,000 rpm) resulting in a transduction efficiency of 3-10% (as determined by FACS analysis on a FACSCalibur). Diluted retroviral supernatants were used to lower the possibility of introducing multiple vectors per cell, thereby decreasing the amount of participating barcodes and thus the chance of random overlap between different cell populations. Based on a Poisson distribution of retroviral integration events, less than 12

% of transduced OT-I T cells is expected to harbor more than one barcode at the observed transduction efficiencies[13]. 24 hrs after retroviral transduction, transduced T cells were purified using Ficoll Paque^{TM} PLUS (Amersham Biosciences, Uppsala, Sweden), stained with PE-conjugated anti-Vα2 and APC-conjugated anti-CD8α and sorted on a FACSAria (BD Biosciences) for GFP$^+$ Vα2$^+$ CD8α$^+$ cells. Subsequently, the cells were washed once in IMDM (Life Technologies, Rockville, MD) supplemented with 5% heat-inactivated FCS, 100 U/ml penicillin, 100 μg/ml streptomycin (Boehringer Mannheim), and 0.5 x 10^{-5} M 2-ME (Merck, Hohenbrunn, Germany) and twice in Hank's balanced salt solution (HBSS, Gibco), resuspended in HBSS and injected intravenously (10,000 GFP$^+$ cells per mouse).

Influenza A virus, Listeria monocytogenes, and tumor challenge

The inflova recombinant influenza A strain[19] that expresses the H-2D^b-restricted $OVA_{257-264}$ epitope (SIINFEKL) (kindly provided by D. Topham) was grown in and titered on Madin Darby canine kidney (MDCK) cells. Mice were infected with 500 plaque-forming units (pfu) inflova by intranasal application. EL4-OVA cells were produced by retroviral transduction of EL4 cells with a pMX vector that encodes a GFP-$OVA_{257-264}$ fusion protein[23]. For tumor challenge, tumor cells were washed three times with HBSS, resuspended in HBSS and injected s.c. into mice. The LM-OVA recombinant Listeria monocytogenes strain[24] that contains OVA was kindly provided by D. Busch. Mice were infected by gavage with ± $2x10^9$ colony forming units (CFU) LM-OVA.

Isolation of barcode labeled T cells

Spleen, lung and tumor tissue was isolated and homogenized using a 70 μm nylon cell strainer (BD Biosciences). Red blood cells were removed from the cell suspensions by treatment with erylysis buffer (155 mM NH_4Cl, 10 mM $KHCO_3$, 0.1 mM EDTA (pH 7.4)). To enrich OT-I cells from spleen, lung and tumor samples, cells were stained with PE-conjugated anti-Vα2 mAbs (BD Biosciences), labeled with anti-PE Microbeads (Miltenyi Biotec) followed by AutoMACS (Miltenyi Biotec) enrichment.

Barcode recovery, microarray hybridizations and data analysis

gDNA of cells was isolated using a DNeasy tissue kit (Qiagen, Venlo, The Netherlands) and barcodes were amplified by nested PCR in a DNA Engine PTC-200 PCR machine (Biorad laboratories B.V., Veenendaal, The Netherlands) using the primers:top, 5'-GTCCTGCTGGAGTTCGTGAC -3'; and bottom 5'- GACCAACTGG TAATGGTAGCGAC -3' in the first round, and the top-LIB and bottom-LIB primers in the second round. The first round consisted of a linear PCR of 50 cycles (5'' 94°C, 5'' 55°C, 5'' 72°C) in which only the top primer was present followed by another 30 cycles (5'' 94°C, 5'' 55°C, 5'' 72°C) of exponential amplification in the presence of both primers. The second round consisted of 30 cycles (5'' 94°C, 5'' 57.2°C, 5'' 72°C). Starting with a linear amplification rather than an exponential amplification decreases the probability that the amplification changes the relative

composition of a barcode pool that initially consists of low numbers of barcodes. In all barcode PCRs, reagent controls were performed in parallel to exclude the presence of contaminating barcode sequences. PCR products were purified and concentrated with a MinElute PCR Purification Kit (Qiagen, Inc., Chatsworth, CA), digested with *Bam*HI and *Xho*I, concentrated and purified again using a MinElute PCR Purification Kit, labeled with cyanine-3 or cyanine-5 fluorescent groups using the Universal Linkage System (ULS; Kreatech Biotechnology, Amsterdam, The Netherlands) and purified over a KreaPure spin column (Kreatech Biotechnology). Samples (containing 100 ng of labeled PCR product) were supplemented with 5 µg of each of the following unlabeled oligos: oligo 1, 5'-TCGAGTGTTCTAGTTGACGGCAGCA-3'; oligo 2, 5'-TGCTGCCGTCAACTAGAACAC-3'; oligo 3, 5'-GATCCTAAGCGCCTGATTCGAGATC-3'; and oligo 4, 5'-GATCTCGAA TCAGGCGCTTAG-3'. Subsequently, samples were concentrated by speed vac and diluted in 30 µl of 32.7% formamide, 5x SSC, 0.85% SDS, 2.6% DMSO (final concentrations), heated to 100 °C for 1 min, snap cooled, and applied to the array. Samples were hybridized O/N at 42 °C, washed and scanned using an Agilent DNA Microarray scanner (Agilent Technologies, Palo Alto, CA). Quantification of the resulting fluorescent images was performed with Imagene 6.0 (BioDiscovery Inc., Los Angeles, CA). Fluorescence intensities were corrected for background noise, log transformed and converted into the Flow Cytometry Standard file format 2.0[25] to allow analysis using Summit V 4.2 FACS analysis software (Dako Netherlands B.V., Heverlee, Belgium). The 2D-plots only depict the 3,029 barcodes that were used for lineage analysis.

Results

Dissecting lineage relationships using cellular barcoding: Proof of principle

To study family relationships between different (T) cell subsets we aimed to label cells with a diverse set of barcodes by retroviral transduction. For this purpose, we generated a retroviral plasmid library, consisting of approximately 5,000 different plasmids (of which about 3,000 are used for lineage analysis, see material and methods). Each plasmid contains GFP as a marker gene and a so-called "barcode", which is a semi-random stretch of 98 bp of non-coding DNA. To allow a rapid analysis of barcodes that are present within a given cell population, the barcodes that are present in this library were also amplified by PCR and spotted individually on microarray slides, resulting in a barcode-microarray. In a typical experiment, T cells are labeled by retroviral infection with the barcode library and reintroduced into mice. After induction of antigen-driven differentiation of the barcode-labeled cells into different subsets, barcodes present in these subsets are recovered by PCR, fluorescently labeled and hybridized on the barcode-microarray. Determining the overlap between the barcodes present in the different subsets should then allow one to determine the kinship between these cell populations.

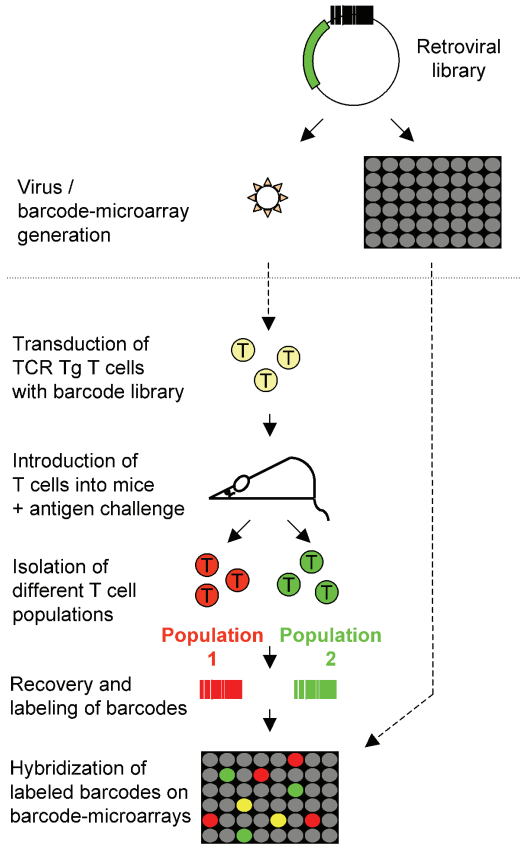


Figure 1. Cellular barcoding strategy. To determine lineage relationships between cell populations, a retroviral plasmid library was constructed containing the GFP gene as a marker and a semi-random stretch of 98 bp of non-coding DNA (the "barcode"). The barcodes that are present in this library were amplified by PCR and individually spotted on microarray slides. In addition, from this plasmid library, a retroviral library was generated that is used to transduce TCR transgenic T cells. Barcode-labeled T cells are introduced into mice, which are subsequently challenged with antigen. At different time points after antigen challenge, populations of T cells can be isolated for kinship analysis by comparing the barcodes isolated from these cells. Barcode analysis is performed by hybridization on microarray slides containing the individual barcodes.

To determine whether this technology (termed cellular barcoding, outlined in Figure 1) can be utilized to analyze kinship between cell populations, we generated effector T cell populations that are either kin or non-kin from a pool of barcode-labeled T cells. For this purpose, we transduced ovalbumin (OVA)-specific T cell receptor transgenic OT-I T cells with the barcode-library. Subsequently, barcode-labeled (GFP+) OT-I T cells were sorted and intravenously injected into mice ($1x10^4$ cells per mouse). On the same day, the mice were challenged with antigen by means of an oral Listeria monocytogenes infection and a subcutaneous injection of EL4 tumor cells that both contain the $OVA_{257-264}$ epitope to ensure strong antigenic stimulation of the barcode-labeled T cells. At day 10 post challenge, barcodes were isolated from the T cells present in the spleen and genomic DNA (gDNA) of each spleen sample was split into two pools. From these pools barcodes were recovered by PCR, labeled with Cy3 and Cy5 dyes and co-hybridized on the barcode-microarray. When two gDNA pools derived from the same spleen are analyzed for the presence of specific barcodes, more than 90% of the barcode-microarray spots that show fluorescent signal above background for one of the samples are also positive in the duplicate sample. In contrast, when barcodes recovered from spleen samples of two different mice (representing two cell pools that are non-kin) are compared (Figure 2A, right panel), the barcodes that are detected are largely unique to one of the spleen samples. In this situation, where two T cell populations are by definition non-kin, the percentage of the barcodes that is shared between the T cell populations in the two mice is

close to the overlap one would expect based on chance (i.e. stochastic activation of a T cell clone with a given barcode in the different mice, see legend Figure 2A). Together these data indicate that cellular barcoding can distinguish between cells that are kin or non-kin and establish the patterns of barcode overlap that may be expected under conditions of full kinship or lack thereof. In addition, the data show that of the approximately 3,000 barcodes that are used for lineage analysis, around 500 are detected above background in a given T cell population (Figure 2A, left and middle panel). Because multiple retroviral integrations (and hence multiple barcodes) per cell are infrequent under the conditions used[13] (see also material and methods), this indicates that approximately 500 of the introduced OT-I T cells participate in the OVA-specific T cell response.

Subsequently, we determined whether cellular barcoding can also detect more subtle differences in kinship, i.e. situations where only part of the T cells in two effector T cell pools has a common ancestor in the naïve T cell compartment, or where the progeny of a given naïve T cell is not unique to but enriched in a certain effector T cell population. To this purpose, we performed microarray hybridizations of barcodes isolated from gDNA of two spleen samples that were mixed at different ratios (1:1, 3:1, 7:1, 15:1 and 1:0). In the most extreme situation, where part of the barcodes are unique to one of the samples and the remainder is shared (Figure 2B. right panel), those barcodes that are common or private (represented by red and green dots, respectively, in Figure 2B. right panel) are readily distinguished. Furthermore, in the situation where part of the barcodes are over-represented 2-8 fold in one of the samples, these barcodes (represented by red dots in figure 2B) become progressively separated from the other barcodes. These data indicate that cellular barcoding can be used to reveal cellular progeny that are enriched in a certain cell population and provides a set of references for barcode-microarray data obtained from biological samples.

Dissecting kinship between antigen-specific T cells at distinct effector sites

Having established the feasibility of lineage analysis by retroviral barcoding, we set out to assess whether T cell populations that accumulate at distinct effector sites share common ancestors in the naïve T cell pool. It has previously been shown that T cell subsets can be distinguished on the basis of differential expression of chemokine receptors and adhesion molecules, and such differential expression is thought to guide the migration of these subsets to distinct sites in the body. For instance, the $\alpha4\beta7$ integrin and chemokine receptor CCR9 guide lymphocytes to the lamina propria of the small intestine and the mucosal epithelium[14-16], whereas E-selectin ligands (E-lig) and P-selectin ligands (P-lig) guide T cells to inflamed skin[17,18]. In addition, it has been shown that T cells that are activated *in vitro* by DCs from gut draining lymph node (DLN) or skin DLN express $\alpha4\beta7$/CCR9 and E-lig/P-lig, respectively, and upon reintroduction preferentially migrate to the corresponding effector site[6,9,10]. These *in vitro* priming data are consistent with the notion that the migration properties of effector T cells may be controlled by imprinting that occurs during the interaction of naïve T cells with antigen-bearing DCs. To determine the impact of imprinting on T cell migration *in vivo*, we

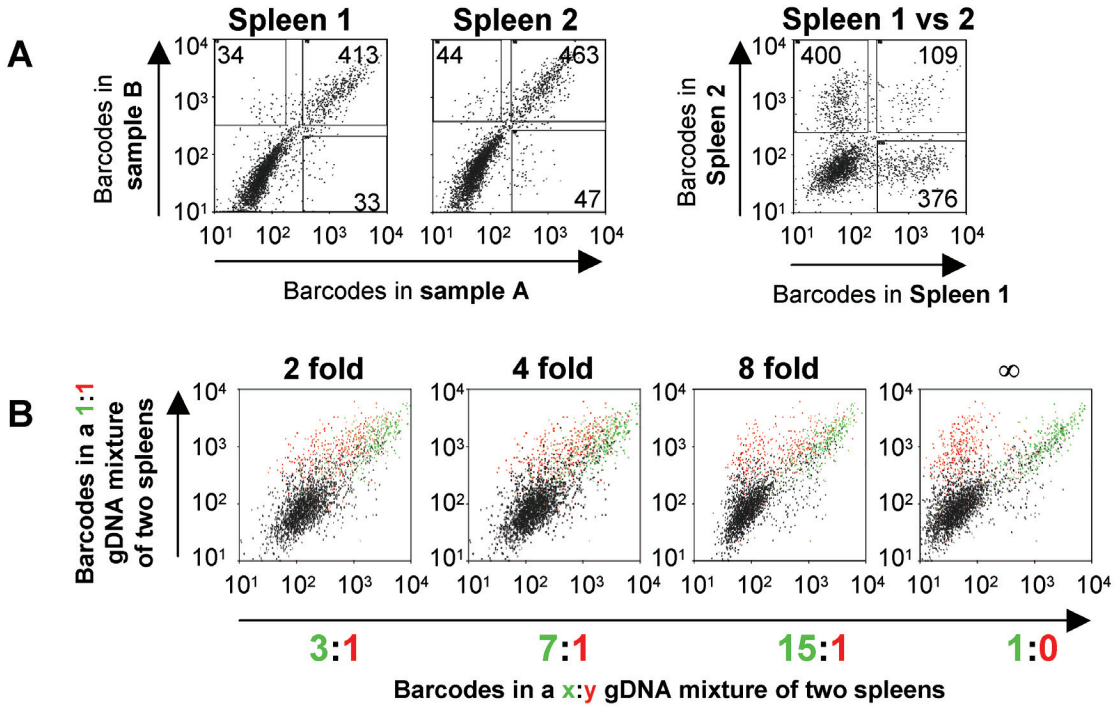


Figure 2. Proof of principle and sensitivity of kinship analysis by cellular barcoding.
A) Barcode analysis of mice that received barcode-labeled OT-I T cells and that were subsequently challenged with LM-OVA and EL4-OVA cells. At day 10 post challenge, T cells were isolated from spleens for barcode analysis. Plots represent the log-transformed fluorescence intensities of the barcode-microarray spots. Left and middle panels: Dot plots of barcode analysis of two gDNA pools of T cells isolated from a spleen for two individual mice. Right panel: Dot plot of barcode analysis of T cells isolated from spleens of two different mice. Black numbers indicate the number of barcodes present in each gate (black rectangles). The amount of barcodes that is shared between the spleen samples of the two different mice (i.e. 109 barcodes) is close to the amount that is expected based on a random overlap in the barcodes present in two unrelated cell populations (i.e. 500 / 3,000 x 500 = 83, is calculated based on the number of barcodes within a gDNA pool of a spleen (i.e. 500) and the amount of spots in the library that is used for lineage analysis (i.e. 3,000)). B) Barcodes were isolated from gDNA mixtures of two spleens that were obtained as described in figure 2A. The y-axes show fluorescent intensities of a sample that contains a 1:1 mixture of gDNA of two spleens (spleen 1 and 2). The x-axes show fluorescent intensities of samples that contain a 3:1, 7:1, 15:1, and 1:0 mixture of gDNA of spleen 1 and 2. Barcodes present in spleen 1 and spleen 2 are labeled green and red, respectively. Black numbers above each plot indicate the level at which gDNA of spleen 2 is over-represented in the sample on the y-axes as compared to the sample on the x-axes.

analyzed the kinship between pathogen-specific T cells present at two different effector sites, i.e. in a subcutaneous tumor and in influenza A-infected lung tissue.

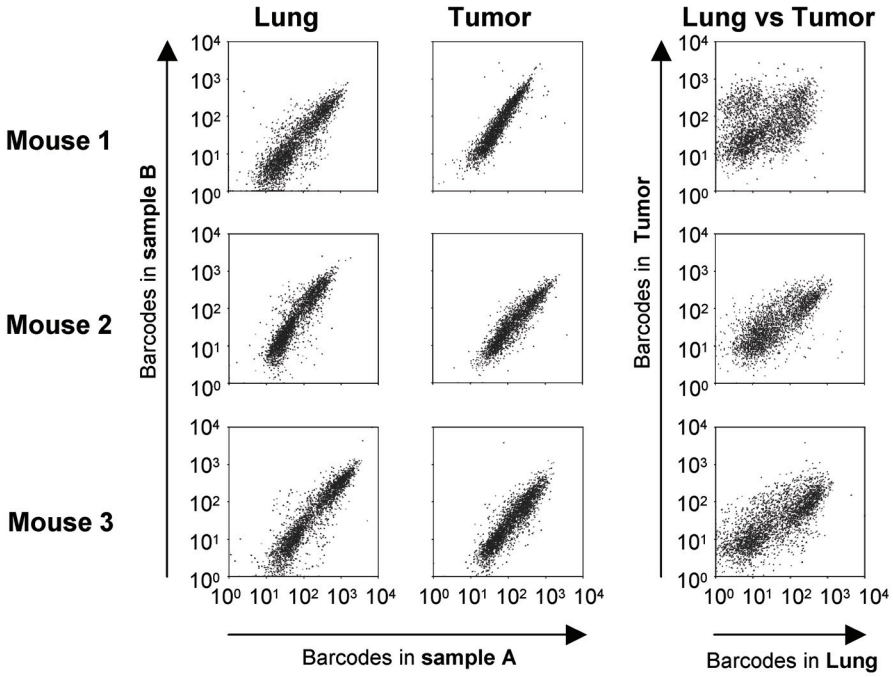


Figure 3. Most T cells present in inflamed lung and tumor are derived from the same precursors. Barcode analysis of T cell populations in mice that received barcode-labeled OT-I T cells and that were subsequently challenged with EL4-OVA cells and inflova. At day 10 post challenge, T cells from the tumor and lung were isolated for barcode analysis. Plots represent the log-transformed fluorescence intensities of the barcode-microarray spots. Left and middle panels: Dot plots of barcode analysis of two gDNA pools of T cells isolated from the same organ (lung and tumor, respectively) for three individual mice. Right panels: Dot plots of barcode analysis of T cells isolated from the tumor versus T cells isolated from the lung for three individual mice. Data are representative for 2 individual experiments.

For this purpose, mice received barcode-labeled OT-I T cells and were then challenged by subcutaneous inoculation of EL4-OVA tumor cells and intranasal infection with an OVA-expressing influenza A recombinant[19] (inflova). At the peak of the OVA-specific T cell response (day 10 post challenge), barcodes were isolated from T cells present at the two effector sites, i.e. tumor and lung. When barcodes isolated from two gDNA pools of the same effector site are co-hybridized, the majority of barcodes that is detected in either sample is also detected in the second sample from this organ and also the intensities of the barcode signals between the two samples are highly correlated (Figure 3, left and middle columns). This indicates that, as is the case for T cells isolated from the spleen, the barcodes present in T

cells of these effector sites can be sampled efficiently. To determine whether the lung-derived and tumor-derived T cells originate from the same pool of precursors, we compared fluorescence intensities of the barcodes isolated from these T cell populations for three individual mice (Figure 3, right column). Notably, of the barcodes that are detected in either lung- or tumor-derived T cells, the majority is also present in the T cell population present at the other site. This indicates that - in this setting - the effect of imprinting on the accumulation of effector T cells at two distinct effector sites is far from absolute. It should be noted that the correlation in signal strength between barcodes recovered from lung versus tumor appears to be lower than that for a self-self analysis either for lung or for tumor resident T cells, and in particular in one of the three mice analyzed, a population of barcodes is highly enriched if not uniquely found at the tumor site, indicating that selective accumulation does occur to some extent.

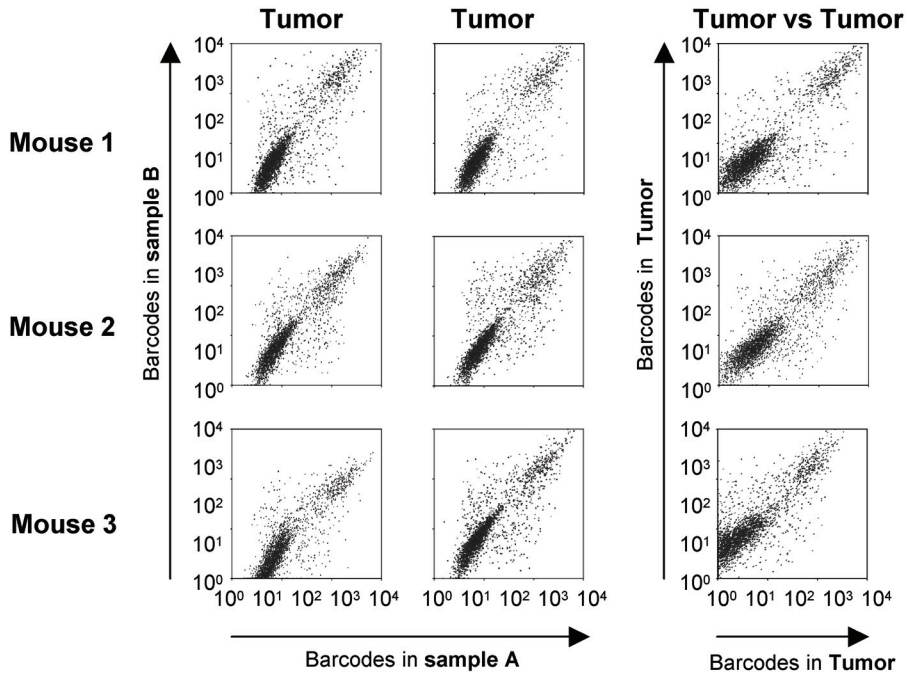


Figure 4. T cells isolated from two different tumor sites are derived from the same precursors. FACS analysis of organs derived from mice that received barcode-labeled OT-I T cells and that were subsequently challenged with EL4-OVA cells at two different subcutaneous sites. At day 10 post challenge, T cells from the tumors were isolated for barcode analysis. Plots represent the log-transformed fluorescence intensities of the barcode-microarray spots. Left and middle panels: Dot plots of barcode analysis of two gDNA pools of T cells isolated from the same organ (tumor from the right and the left flank, respectively) for three individual mice. Right panels: Dot plots of barcode analysis of T cells isolated from the two individual tumors for three individual mice.

Such selective accumulation of T cells derived from the same naïve T cell precursor could occur as a consequence of two processes. a). Selective accumulation may reflect a true difference in tissue tropism shared by progeny of a given T cell b). Selective accumulation may be the result of the stochastic entry of (recently activated) T cells at either site and subsequent local accumulation of its progeny, by proliferation in situ. To distinguish between these two models we compared the barcodes present in tumor-resident T cells in mice that had simultaneously been challenged with tumor cells at two different subcutaneous sites. In this setting tissue-specific migration patterns will not contribute to selective accumulation. However, selective accumulation due to local proliferation can still occur. We found that in this setting, there is very high degree of overlap between the barcodes present in T cells isolated from the two sites (Figure 4). These data indicate that the over-representation of part of the barcodes at the tumor site is due to the preferential migration of progeny derived from a small pool of precursors towards the tumor.

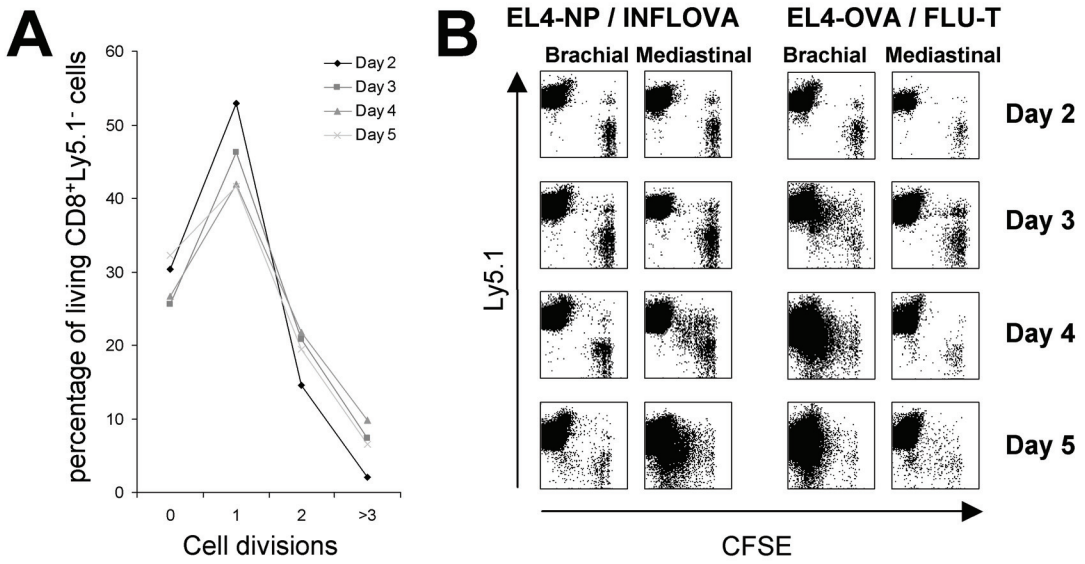
Discussion

Here we used a novel approach, termed cellular barcoding, to dissect lineage relationships between cell populations. In this approach, retroviral tagging of individual cells with molecular barcodes is coupled to a microarray-based detection system to allow lineage relationship analysis of the progeny of such cell populations. By analyzing family relationships between effector T cell populations taken from mice that were challenged with antigen after receiving barcode-labeled TCR transgenic T cells, we demonstrated that cellular barcoding can be used to distinguish between cells that are kin or non-kin. In addition, titration experiments with gDNA of these effector T cells indicate that cellular barcoding can also be used to study whether the progeny of a particular pool of precursor cells is enriched in a certain cell population. We are currently setting up statistical methods to determine to what extent two cell populations are related as based on data obtained by cellular barcoding.

In the experiments described here the number of barcode-labeled T cels that is injected ($1x10^4$) is some 3-fold higher than the diversity of the retroviral library. However, of the injected cells 1200 or less participate in the immune response (figure 2A and data not shown), suggesting that the diversity of our library should be sufficient to detect differences in kinship and formal proof of this notion is provided by the comparison of barcodes present in distinct mice. It is noted that the number of participating clones can be influenced by adjusting conditions of experimental infection, such as pathogen dose (J.v.H., unpublished observations), suggesting that not all barcode-labeled T cells are able to encounter the antigen in the experiments described.

Subsequently, we used cellular barcoding to study the impact of imprinting of homing properties on T cell migration *in vivo*. We show that the majority of T cells present in the inflamed lung and in tumor tissue contain the same barcodes, whereas only a small pool of barcodes was found to be enriched in the tumors. This suggests that the majority of T cells

present in the inflamed lung and in tumor tissue consist of T cells derived from the same precursors, whereas the progeny of only a small pool of precursors is enriched in the tumor, indicating that imprinting of T cell migration has only little impact on the migration of T cells.



Supplementary figure 1. Cell division profiles of ConA-stimulated cells in mice that did or did not receive antigenic challenge. Mice that received CFSE-labeled ConA-stimulated OT-I T cells, were challenged with EL4-NP[26] and inflova, with EL4-OVA and influenza virus containing epitope IV of SV40 large T (Flu-T), or left untreated. At different time points after infection (day 2-5) organs were isolated and stained with anti-CD8 and anti-Ly5.1 antibody. A) For mice that were not challenged with antigen (n=2 per time point) the FACS data of all living CD8+Ly5.1- cells present in the inguinal, brachial and mediastinal lymph nodes and the spleen were accumulated. From CFSE-histograms of these data the percentage of total living CD8+Ly5.1- cells within each successive cycle of cell division were determined. B) Representative FACS profiles of tumor-draining brachial lymph nodes and lung-draining mediastinal lymph nodes of mice challenged with EL4-NT and inflova (left panels) or with EL4-OVA and Flu-T (right panels). Cells shown are live CD8+ lymphocytes. FACS plots are respresentative for 3-4 mice per time point. Note: The small population of Ly5.1+ CFSE+ cells is likely to represent clusters of cells that contain both donor and recipient cells, as additional experiments showed that the Ly5.1+ CFSE+ population is also Ly5.2+.

However, two findings suggest that this analysis could possibly underestimate the impact of imprinting on T cell migration. Firstly, mock-transduced ConA-stimulated T cells were found to undergo several antigen-independent cycles of replication within the first two days after *in vivo* transfer (see supplementary figure 1A). Because of this expansion, two independent samples taken from within one mouse may share a greater number of barcodes than samples obtained from different mice. While the magnitude of this effect is presently unclear, the fact

that T cell expansion post-infusion is observed indicates that the analysis shown in figure 2A may overestimate our ability to detect differences in kinship of T cell populations within an individual mouse. To avoid this complicating factor, we plan to "park" our barcode-labeled cells for some time *in vitro* or *in vivo,* in order to ensure that the barcode-labeled cells are quiescent at the moment of transfer into the mice that receive the antigenic challenge. Secondly, in settings where antigen-induced T cell activation occurs in two distinct lymph node beds, a possible effect of imprinting may be expected to be maximal when T cell priming within the DLN of the two sites occurs simultaneously, and our analysis may underestimate the effect of imprinting should T cell priming at the different sites occur with distinct kinetics. We have observed that T cells are primed approximately 1 day earlier in the tumor DLN than in the lung DLN (supplementary figure 1B) when antigen is offered at both sites simultaneously. In addition, T cells that have been primed in the tumor DLN disseminate throughout the body including the lung DLN by the time of the first T cell priming in the lung DLN (see mediastinal lymph nodes at day 4, supplementary figure 1B). As a consequence, it is possible that the first offspring of T cells primed within the tumor DLN dominate the T cell response that is subsequently initiated in the lung DLN and, upon reprogramming, acquires the capacity to migrate to the lung. To test the possibility that the effect of imprinting of homing properties is masked by differences in the time of priming, we shall perform cellular barcoding experiments in which we ensure that T cells are simultaneously primed at the two different inflammatory sites.

Collectively, our data demonstrate that cellular barcoding is an effective tool for the dissection of lineage relationships between different T cell populations. In addition, because cellular barcoding also provides an estimate of the number of transferred T cell precursors that participate in an antigen-specific T cell response, it should also be possible to determine the effect of parameters such as antigen dose, or duration of antigen exposure, on the recruitment of naïve T cells into the antigen-specific T cell response. Furthermore, we consider it likely that this technology may not only be applied to study cellular differentiation and migration within the immune system, but can also be utilized to examine stem cell/differentiation issues for other cell types.

References

1. Carter, L. L. & Dutton, R. W. Type 1 and type 2: a fundamental dichotomy for all T-cell subsets. *Curr Opin Immunol* **8**, 336-342 (1996).
2. Mosmann, T. R. & Sad, S. The expanding universe of T-cell subsets: Th1, Th2 and more. *Immunol Today* **17**, 138-146 (1996).
3. Schepers, K., Arens, R. & Schumacher, T. N. Dissection of cytotoxic and helper T cell responses. *Cell Mol Life Sci* **62**, 2695-2710 (2005).
4. Bourgeois, C., Rocha, B. & Tanchot, C. A role for CD40 expression on CD8+ T cells in the generation of CD8+ T cell memory. *Science* **297**, 2060-2063 (2002).

5. Calzascia, T. *et al.* Homing phenotypes of tumor-specific CD8 T cells are predetermined at the tumor site by crosspresenting APCs. *Immunity* **22**, 175-184 (2005).
6. Dudda, J. C. *et al.* Dendritic cells govern induction and reprogramming of polarized tissue-selective homing receptor patterns of T cells: important roles for soluble factors and tissue microenvironments. *Eur J Immunol* **35**, 1056-1065 (2005).
7. Iwata, M. *et al.* Retinoic acid imprints gut-homing specificity on T cells. *Immunity* **21**, 527-538 (2004).
8. Janssen, E. M. *et al.* CD4+ T cells are required for secondary expansion and memory in CD8+ T lymphocytes. *Nature* **421**, 852-856 (2003).
9. Mora, J. R. *et al.* Selective imprinting of gut-homing T cells by Peyer's patch dendritic cells. *Nature* **424**, 88-93 (2003).
10. Mora, J. R. *et al.* Reciprocal and dynamic control of CD8 T cell homing by dendritic cells from skin- and gut-associated lymphoid tissues. *J Exp Med* **201**, 303-316 (2005).
11. Shedlock, D. J. & Shen, H. Requirement for CD4 T cell help in generating functional CD8 T cell memory. *Science* **300**, 337-339 (2003).
12. Sun, J. C. & Bevan, M. J. Defective CD8 T cell memory following acute infection without CD4 T cell help. *Science* **300**, 339-342 (2003).
13. Fehse, B., Kustikova, O. S., Bubenheim, M. & Baum, C. Pois(s)on--it's a question of dose.. *Gene Ther* **11**, 879-881 (2004).
14. Hamann, A., Andrew, D. P., Jablonski-Westrich, D., Holzmann, B. & Butcher, E. C. Role of alpha 4-integrins in lymphocyte homing to mucosal tissues in vivo. *J Immunol* **152**, 3282-3293 (1994).
15. Johansson-Lindbom, B. *et al.* Selective generation of gut tropic T cells in gut-associated lymphoid tissue (GALT): requirement for GALT dendritic cells and adjuvant. *J Exp Med* **198**, 963-969 (2003).
16. Svensson, M. *et al.* CCL25 mediates the localization of recently activated CD8alphabeta(+) lymphocytes to the small-intestinal mucosa. *J Clin Invest* **110**, 1113-1121 (2002).
17. Picker, L. J., Kishimoto, T. K., Smith, C. W., Warnock, R. A. & Butcher, E. C. ELAM-1 is an adhesion molecule for skin-homing T cells. *Nature* **349**, 796-799 (1991).
18. Tietz, W. *et al.* CD4+ T cells migrate into inflamed skin only if they express ligands for E- and P-selectin. *J Immunol* **161**, 963-970 (1998).
19. Topham, D. J., Castrucci, M. R., Wingo, F. S., Belz, G. T. & Doherty, P. C. The role of antigen in the localization of naive, acutely activated, and memory CD8(+) T cells to the lung during influenza pneumonia. *J Immunol* **167**, 6983-6990 (2001).
20. Hogquist, K. A. *et al.* T cell receptor antagonist peptides induce positive selection. *Cell* **76**, 17-27 (1994).
21. Naviaux, R. K., Costanzi, E., Haas, M. & Verma, I. M. The pCL vector system: rapid production of helper-free, high-titer, recombinant retroviruses. *J Virol* **70**, 5701-5705 (1996).

22. Kessels, H. W., Wolkers, M. C., van den Boom, M. D., van der Valk, M. A. & Schumacher, T. N. Immunotherapy through TCR gene transfer. *Nat Immunol* **2**, 957-961 (2001).
23. Kessels, H. W., Schepers, K., van den Boom, M. D., Topham, D. J. & Schumacher, T. N. Generation of T cell help through an MHC class I restricted TCR. Submitted. 2006. Ref Type: Generic
24. Pope, C. *et al.* Organ-specific regulation of the CD8 T cell response to Listeria monocytogenes infection. *J Immunol* **166**, 3402-3409 (2001).
25. Dean, P. N., Bagwell, C. B., Lindmo, T., Murphy, R. F. & Salzman, G. C. Introduction to flow cytometry data file standard. *Cytometry.* **11**, 321-322 (1990).
26. Wolkers, M. C., Stoetter, G., Vyth-Dreese, F. A. & Schumacher, T. N. Redundancy of direct priming and cross-priming in tumor-specific CD8+ T cell responses. *J Immunol* **167**, 3577-3584 (2001).