Dissecting T cell lineage relationships by cellular barcoding

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

[1]Division of Immunology and [2]Central Microarray Facility, the Netherlands Cancer Institute, 1066 CX Amsterdam, Netherlands
[3]Istituto Superiore di Sanità, 00161 Roma, Italy

T cells, as well as other cell types, are composed of phenotypically and functionally distinct subsets. However, for many of these populations it is unclear whether they develop from common or separate progenitors. To address such issues, we developed a novel approach, termed cellular barcoding, that allows the dissection of lineage relationships. We demonstrate that the labeling of cells with unique identifiers coupled to a microarray-based detection system can be used to analyze family relationships between the progeny of such cells. To exemplify the potential of this technique, we studied migration patterns of families of antigen-specific CD8[+] T cells in vivo. We demonstrate that progeny of individual T cells rapidly seed independent lymph nodes and that antigen-specific CD8[+] T cells present at different effector sites are largely derived from a common pool of precursors. These data show how locally primed T cells disperse and provide a technology for kinship analysis with wider utility.

CORRESPONDENCE
Ton N.M. Schumacher:
t.schumacher@nki.nl

Abbreviations used: Cy3, Cyanine-3; DLN, draining LN; LDLN, lung draining mediastinal LN; TDLN, tumor-draining axillary/inguinal LN.

To ascertain whether different cell types have a common or separate ancestor has generally been difficult. Historically, differentiation assays have been used to study lineage relationships. In most of these assays, the in vitro or in vivo developmental potential of progenitors with a homogeneous phenotype is studied under different conditions. However, the finding that a phenotypically homogeneous cell population can give rise to two types of cellular output does not necessarily imply that individual cells within the population are multipotent. This problem can be remedied by the use of clonal assay systems in which the cellular output of a single (injected) cell is studied, but the difficulties of single cell experiments preclude its use in many settings. An appealing strategy to address this issue is the use of larger pools of progenitor cells in which each individual cell carries a mark that allows one to distinguish its offspring from the offspring of other cells. In previous studies, the site of random retroviral integration has been used as a marker to identify the progeny of individual hematopoietic cells by Southern blot analysis or PCR (1–3). Furthermore, Golden et al. (4) have previously used a sequencing-based detection technique to study lineage relationships between different cell populations in the developing brain by retroviral tagging. In this paper, we extend this general concept to a high-throughput approach in which individual progenitor cells are tagged with a molecular barcode that can be analyzed by microarray hybridization. This technique allows a high-throughput analysis of differentiation and migration patterns of families of cells. We demonstrate the utility of this technology by dissecting the kinship of antigen-specific T cells that accumulate at priming and effector sites during local antigen-driven T cell responses. This technology will be of value in addressing lineage relationships and progenitor potential of T cell subsets and other subsets of hematopoietic cells.

E. Swart and J.W.J. van Heijst contributed equally to this paper.
K. Schepers' present address is Institute for Regeneration Medicine, University of California San Francisco, San Francisco, CA.
R. Arens' present address is La Jolla Institute for Allergy and Immunology, La Jolla, CA.
The online version of this article contains supplemental material.

RESULTS

Dissecting lineage relationships by cellular barcoding: general strategy and proof of principle

To uniquely tag individual cells with molecular barcodes, we generated a retroviral plasmid library, consisting of ~5,000 different plasmids. Each plasmid contains GFP as a marker gene and a so-called barcode, formed by a semirandom stretch of 98 bp of noncoding DNA. To allow for rapid analysis of barcodes present within a given cell population, each barcode present in this library was also amplified individually by PCR and spotted on microarray slides, resulting in a barcode microarray. In a typical experiment, cells are labeled by retroviral transduction with the barcode library and reintroduced into mice. After in vivo expansion and differentiation of the barcode-labeled cells, barcodes present in different subsets of cells are recovered by PCR, fluorescently labeled, and hybridized on the barcode microarray. Determining the overlap of barcodes between different subsets by cellular barcoding (Fig. 1 A) should then allow one to determine the kinship between these cell populations. In addition to the retroviral barcode library used here, a lentiviral barcode library was prepared to allow tagging of quiescent cell types that are refractory to retroviral infection (unpublished data).

Before assessing the feasibility of kinship analysis, we first established whether the barcode microarray shows any cross-hybridization between different barcodes present in the barcode library. When the barcode microarray was hybridized with PCR products of 273 individual *Escherichia coli* clones of the master library that were labeled with Cyanine-3 (Cy3), with Cy5, or with both Cy3 and Cy5 dyes, no detectable crosshybridization between different barcodes was observed (Fig. 1 B). Second, as the effective size of the barcode library will decrease proportionally with the number of barcodes that is introduced per cell, we also established retroviral transduction conditions under which the mean number of barcodes per cell is close to one (Fig. S1, available at http://www.jem.org/cgi/content/full/jem.20072462/DC1).

As a first test of the barcoding technology, we assessed whether cellular barcoding can be used to analyze kinship between cell populations that are either kin or nonkin. For this purpose, OVA-specific MHC class I–restricted T cell receptor transgenic OT-I T cells were transduced with the barcode library. Subsequently, sorted barcode-labeled OT-I T cells were i.v. injected into mice that were then challenged with antigen (i.v. challenge with *Listeria monocytogenes* encoding OVA [LM-OVA]). At day 7 after challenge, the OT-I T cell pool present in the spleen was split into two samples (which are evidently derived from the same pool of precursors), hereafter referred to as half-samples. From the genomic DNA of these half-samples, barcodes were recovered by PCR, labeled with Cy3 and Cy5 dyes, and cohybridized on the barcode microarray. Resulting data are presented in two-dimensional plots, depicting the Cy3 and Cy5 signals of each individual spot on the barcode microarray. In these plots, each dot represents a clone (or family) of cells that are the progeny of a precursor containing a specific barcode. The presence or
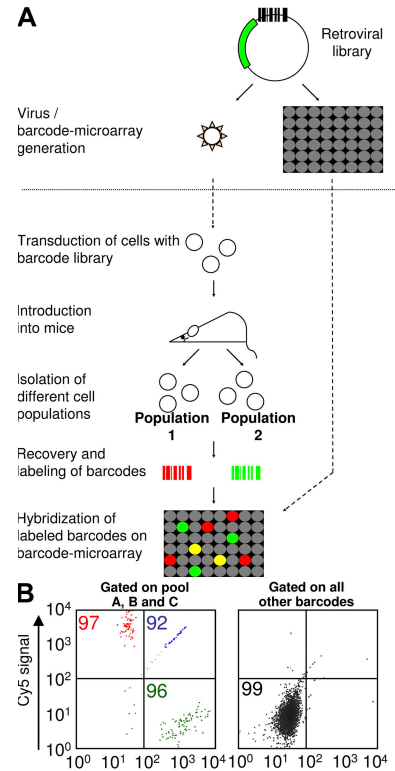

Figure 1. Cellular barcoding strategy. (A) To allow the analysis of lineage relationships between cell populations, a retroviral plasmid library was constructed containing a GFP marker gene and a semirandom stretch of 98 bp of noncoding DNA (the barcode). The barcodes that are present in this library 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 T cells or other cells. Barcode-labeled cells are introduced into mice to allow proliferation and differentiation, and at different time points after infusion, populations of cells can be isolated for kinship analysis by comparing the barcodes present within the different cell populations. Barcode analysis is performed by hybridization on microarray slides containing the individual barcodes. (B) Analysis of a hybridization with PCR products of 273 individual *E. coli* clones of the master library. The PCR products were pooled into three pools (A, B, and C), each containing PCR products of 91 clones. The x-axes show nonnormalized fluorescence intensities of a Cy3-labeled sample that contains PCR products of pools A and B, and the y-axes show nonnormalized fluorescence intensities of a Cy5-labeled sample that contains PCR products of pools B and C. Barcodes that are present in pools A, B, and C are plotted on the left and are labeled green, blue, and red, respectively. Barcodes that are not present in pools A, B or C are plotted on the right. Colored numbers indicate the percentage of barcodes present in the respective gates (black rectangles) as a percentage of the total number of barcodes that is present in the corresponding pool.

absence of a barcode within a cell population is then determined by calculation of the distribution of background signals from averaged Cy3 and Cy5 signals, as described in Materials and methods. When two half-samples from the same spleen are compared for the presence of barcodes, >85% of the barcodes that show detectable signal in one of the spleen samples is also detected in the other half-sample (Fig. 2 A, left and middle).

In contrast, when barcodes recovered from spleen samples of two different mice (representing two cell pools that are non-kin) are compared, the barcodes that are detected are largely unique to either one of the spleen samples (Fig. 2 A, right). The small percentage of the barcodes that is shared between the T cell populations in the two mice is close to the overlap that is expected based on chance (i.e., the introduction and activation of T cells that by chance have the same barcode in the different mice; Fig. 2 A, legend). Together these data indicate that cellular barcoding can distinguish between cells that are kin or nonkin and establish the patterns of barcode overlap that may be expected under conditions of full kinship or lack thereof.

Subsequently, we determined whether cellular barcoding can also detect more subtle differences in kinship, reflecting
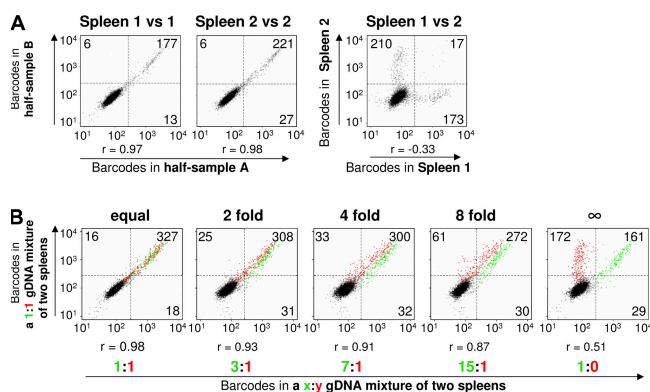
Figure 2. Proof of principle and sensitivity of kinship analysis by cellular barcoding. (A) Barcode analysis of indicated samples (at day 7 after challenge) from mice that received 1,000 barcode-labeled OT-I T cells and that were subsequently challenged by i.v. LM-OVA infection. Plots represent the fluorescence intensities of the barcode microarray spots. Left and middle: Dot plots of barcode analysis of T cells isolated from two half-samples from the same spleen for two individual mice. Right: Dot plot of barcode analysis of T cells isolated from spleens of two different mice. Numbers indicate the number of barcodes present in each quadrant (black rectangles; cutoff used, P < 0.0005). With the assumption that ~3,300 of the 4,743 barcodes can effectively be used for lineage analysis (supplemental Materials and methods, available at http://www.jem.org/cgi/content/full/jem.20072462/DC1) and that each barcode has the same probability to participate in the T cell response, the expected overlap between barcodes present in the two unrelated cell populations in A is 200/3,300 × 200 = 12 (in which 200 is the approximate number of barcodes within a half-sample of a spleen), which is close to the observed overlap between the two unrelated samples (i.e., 17 barcodes). (B) Barcode analysis of genomic DNA mixtures of the two spleen samples shown in A. Plots represent the fluorescence intensities of the barcode microarray spots. The y-axes show fluorescence intensities of a sample that contains a 1:1 mixture of genomic DNA of spleen 1 and 2. The x-axes show fluorescence intensities of samples that contain a 1:1, 3:1, 7:1, 15:1, or 1:0 mixture of genomic DNA of spleen 1 and 2, respectively. Barcodes uniquely present in spleen 1 and spleen 2 are labeled green and red, respectively. Numbers above each plot indicate the level at which the cells of spleen 2 are overrepresented in the sample on the y-axis as compared with the sample on the x-axis. R values represent the correlation between hybridization signals of all barcodes present in either spleen 1 or 2.

situations where only part of the T cells in two effector T cell pools has a common ancestor in the naive T cell compartment or where the progeny of a given naive T cell is enriched but not unique to a certain T cell population. For this purpose, we performed microarray hybridizations of barcodes isolated from genomic DNA 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 is unique to one of the samples and the remainder is shared (i.e., the 1:1 versus 1:0 comparison; Fig. 2 B, right), those barcodes that are common or private (represented by green and red dots, respectively) are readily distinguished. Furthermore, in situations where part of the barcodes is overrepresented 2–8-fold in one of the samples, these barcodes (Fig. 2 B, red dots) become progressively separated from the other barcodes. In the analysis shown in Fig. 2 B, relatedness of cell populations is revealed by grouping barcodes as either shared or private (i.e., present in either the top right quadrant or the bottom right/top left quadrant). As an alternative measure of relatedness, the correlation between the Cy3 and Cy5 hybridization signals for each barcode can be used. Analysis of the data in Fig. 2 B by such correlation analysis demonstrates that an increasing difference in the frequency of half of the barcodes in two samples leads to a progressive loss in correlation (Fig. 2 B, r values). These data indicate that cellular barcoding can be used to reveal cellular progeny that are enriched in a certain cell population and demonstrate that the fluorescence intensity of barcode microarray signals correlates with the frequency at which a specific barcode is present within a certain sample.

Dissecting kinship between antigen–specific T cells at distinct priming and effector sites

Having established the feasibility of lineage analysis by cellular barcoding, we assessed whether T cell families that accumulate at distinct priming and effector sites share common ancestors in the naive 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 (5–9). For instance, it has been shown that T cells that are activated in vitro by dendritic cells from gut draining LN (DLN) or skin DLN preferentially migrate to the corresponding effector site (10–12) upon reintroduction. To determine whether selective migration of T cells also occurs upon antigen encounter in vivo in a case where there is a choice between two antigen–positive effector sites, we analyzed the kinship between antigen–specific T cells in a dual challenge model. For this purpose, mice were challenged by s.c. inoculation of EL4–OVA tumor cells and parallel intranasal infection with WSN–OVA influenza virus to generate two local independent effector sites. We considered the possibility that a potential effect of imprinting could be masked or reduced when T cell activation in the two sets of DLN would not occur simultaneously as, in this case, the progeny of T cells activated within the first set of DLN could seed the other LN bed (Fig. 3) (13)

and thereby dominate the T cell response at that site. In particular, in view of recent data that suggest that programming of T cell functions may not be fixed (10, 12), such a situation could lead to an underestimate of the potential of imprinting. To avoid this issue, we first determined the kinetics of T cell priming at the two different LN sites. Analysis of CFSE dilution of labeled OT-I T cells in tumor and lung DLN of mice that received an s.c. inoculation of EL4–OVA tumor cells or intranasal WSN–OVA influenza infection revealed that detectable T cell proliferation is first observed at days 3 and 4 after antigenic challenge, respectively (unpublished data). To allow simultaneous priming of T cells in the two LN beds, mice were therefore challenged with EL4–OVA tumor inoculation 1 d after intranasal WSN–OVA influenza inoculation in all subsequent experiments (Fig. 3).

The kinship of antigen-specific T cells that accumulate at the two priming and two effector sites was subsequently monitored by barcode analysis. When barcodes are recovered from tumor DLN at day 5 or later time points, the percentage of barcodes recovered from two half-samples of a tumor DLN that is found in both half-samples is markedly higher than the random overlap observed between tumor DLN samples obtained from different mice (Fig. 4). Likewise, when barcodes are recovered from lung DLN, the percentage of barcodes recovered from two half-samples of a lung DLN that is found in both half-samples is markedly higher than the background overlap observed between lung DLN samples from different mice (Fig. 4). Together, these data indicate that barcode analysis of LN–resident T cell populations is feasible, even early after the induction of T cell proliferation.

To reveal whether draining LNs of tumor and lung contain the same sets of T cell families, barcodes obtained from tumor DLN and lung DLN were subsequently compared. Early after antigenic challenge (day 5), the overlap in barcodes present in tumor and lung DLN obtained from the same

mouse (Fig. 4, A [right, top two rows] and C) is small and only slightly higher than the overlap between barcodes obtained from tumor and lung DLN from different mice (Fig. 4, A [right, bottom row] and C). This indicates that early after antigen-induced activation, T cell families mostly remain confined to the original site of priming. Interestingly, from days 6 to 8 after challenge, a marked shift occurs, in which a large fraction of barcodes becomes shared between T cells isolated from tumor and lung DLN, suggesting that within days after priming, the offspring of T cells intermingles within different DLN.

To provide a more quantitative assessment of the similarity of T cell families present at the two sites, data were converted into percentage identity, in which the percentage of shared barcodes, as shown in Fig. 4 C, is corrected for maximal and background overlap (analogous to the correction of data for background and maximal lysis in CTL assays, see Materials and methods for details). Based on this analysis, T cell families within tumor and lung DLN share 8 and 16% identity on days 5 and 6. This abruptly changes to 78 and 75% identity on days 8 and 10. Also, when the data are analyzed by correlation analysis, T cell families at the two sites are clearly distinct early after antigenic challenge but become close to identical from day 8 (Fig. S2, available at http://www.jem.org/cgi/content/full/jem.20072462/DC1). To test whether the increase in barcode overlap between tumor and lung DLN over time is indeed caused by T cell migration, we treated mice with FTY720, a compound which blocks LN egress of T lymphocytes (14, 15). Importantly, when LN egress is blocked during the course of antigenic challenge, the relatedness of T cells that are present in tumor and lung DLN remains low even late during infection (day 8, 29% identity; Fig. S3). These data show that the time-dependent increase in relatedness of T cell families isolated from lung and tumor DLN is a result of T cell migration.
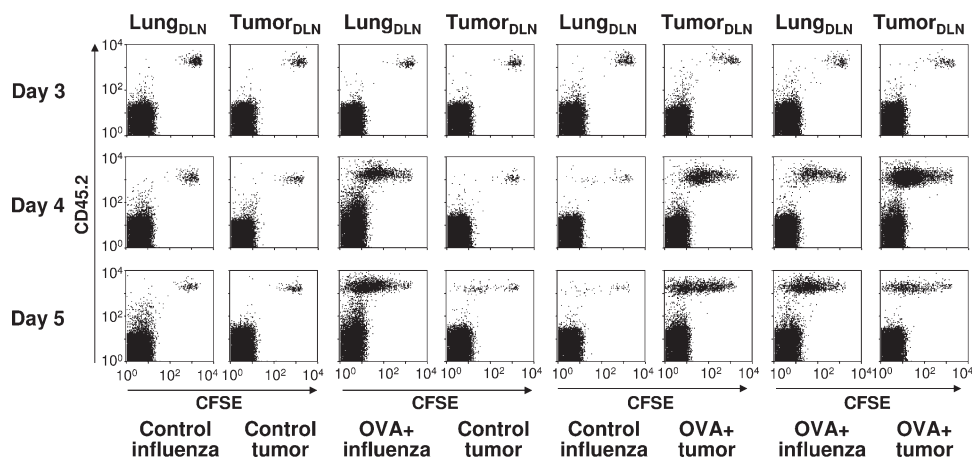
Figure 3. **Simultaneous priming of OT–I T cells in tumor DLN and lung DLN.** Four groups of B6 Ly5.1+ mice that received 3 × 10⁶ CFSE-labeled mock-transduced OT-I T cells (Ly5.2+) on day 0 were challenged intranasally with the indicated influenza virus on day 0 and challenged s.c. with the indicated tumor cells on day 1. Plots show representative flow cytometric analyses of tumor-draining axillary/inguinal LNs (TDLN) and lung-draining mediastinal LNs (LDLN) at different time points after initial challenge. In each plot the same amount of living CD8+ lymphocytes is depicted. Plots are representative for three to four mice per experiment, out of two experiments.
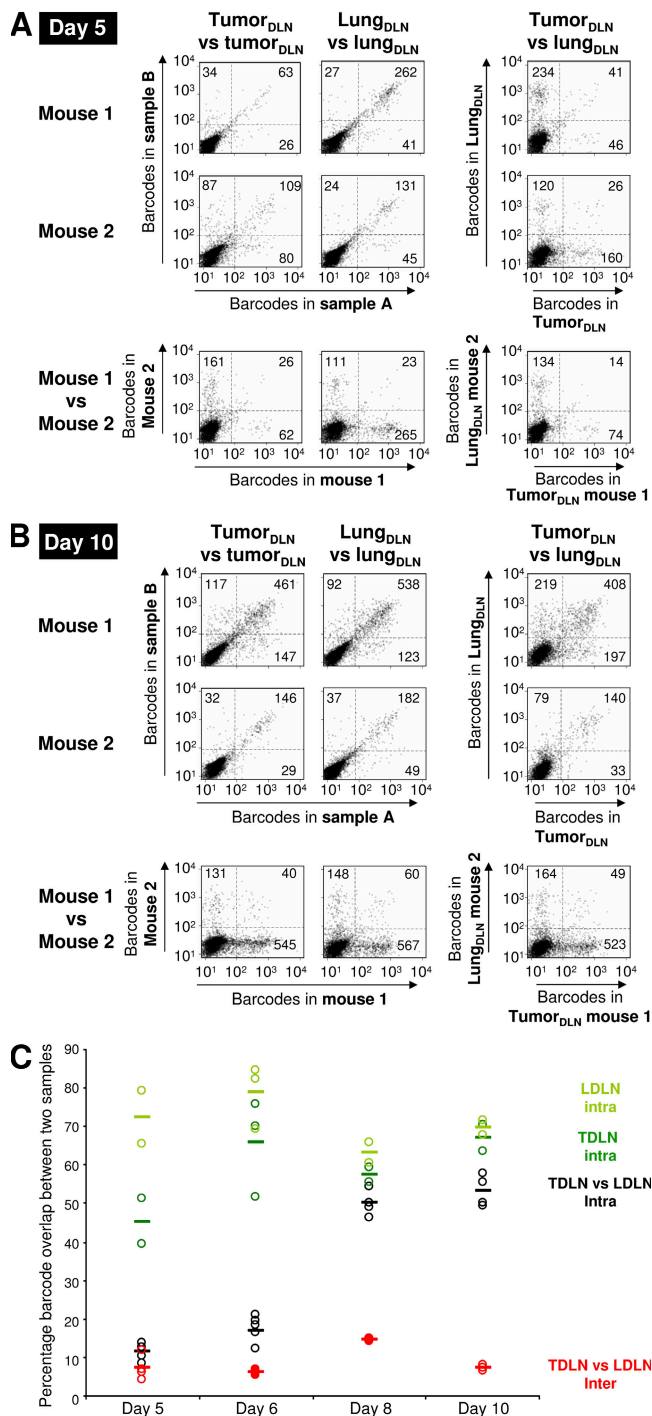
Figure 4. Redistribution of T cell families over LN beds through time. Barcode analysis of T cell populations isolated from draining LNs in mice that received 10,000 barcode-labeled OT-I T cells and that were subsequently challenged intranasally with WSN-OVA influenza virus (day 0) and s.c. with EL4-OVA cells (day 1). At indicated days after challenge, TDLN and LDLN were each split into two half-samples that were separately cultured for 4 d and used for barcode analysis. The results represent data from two to three mice analyzed separately within one experiment. (A and B) Representative dot plots of the fluorescence intensities of the barcode microarray spots at days 5 (A) and 10 (B). Left and middle, top

To test whether barcode analysis of effector site–resident T cells was feasible, barcodes were obtained from tumor and lung tissue at various time points after antigenic challenge. When T cells are isolated from the tumor and lung effector sites during the OVA-specific T cell response, barcodes can be recovered reliably from these samples (Fig. 5, A [left and middle, top two rows] and B). Importantly, of the barcodes that are present at one effector site, the majority is also found at the other effector site within the same mouse (Fig. 5, A [right, top two rows] and B), whereas samples derived from different mice do not show such a shared set of barcodes (Fig. 5, A [bottom row] and B). Furthermore, this high overlap is observed both at day 8, when substantial T cell accumulation at both sites has just begun (Fig. 5), and at the peak of the antigen-specific T cell response (day 10 after challenge; not depicted). Thus, although the T cell populations that develop within tumor and lung DLN initially consist of distinct families, the progeny that accumulates at the lung and tumor effector sites are highly related (81% identity at day 8; Fig. S4 A, available at http://www.jem.org/cgi/content/full/jem.20072462/DC1) Together, these data show that, in this model, the progeny of CD8[+] T cells disperses, essentially indiscriminately, over two effector sites.

To investigate whether this pattern also applies to a third effector site, we analyzed the kinship of T cells migrating to inflamed gut and s.c. tumor effector sites. To this purpose, mice were challenged by oral infection with LM-OVA and s.c. inoculation of EL-4 OVA tumor cells. Again, challenges were timed such that priming occurred simultaneously in gut and TDLNs (unpublished data). Furthermore, by comparing barcodes isolated from different half-samples around the peak of the OVA-specific T cell response, it was established that barcodes from gut tissue can also be recovered reliably (Fig. 6). When barcodes present in T cells that accumulated in the gut and the tumor of the same mouse were compared, no evidence was found for selective migration of T cell families toward the tumor site (Fig. 6, A [right,

two rows: dot plots of barcode analysis of two pools of T cells isolated from the same tissue for two individual mice. Right, top two rows: dot plots of barcode analysis of T cells isolated from the TDLN versus T cells isolated from the LDLN for two individual mice. Bottom row: evaluation of background overlap by comparison of samples from mouse 1 with samples from mouse 2. Numbers indicate the barcodes that are present within each quadrant (black rectangles; cutoff used, $P < 0.005$). (C) Percentage of barcode overlap between different tissue samples. Each sample was compared with a second sample generated from the same tissue (TDLN or LDLN) or from a different tissue (TDLN vs LDLN), either of the same mouse (intra) or of a different mouse (inter). The percentages indicate the number of barcodes that is present within the top right quadrant as a fraction of the total number of barcodes present in the top left, top right, and bottom right quadrant (as indicated in A and B). By normalizing for background and maximal overlap, data can be converted to percentage identity between T cell families at both sites (numbers in text). Percentage identity of T cell families in the control comparisons (TDLN-X vs. LDLN-Y) was 1, 0, −1, and 0% on days 5, 6, 8, and 10, respectively.
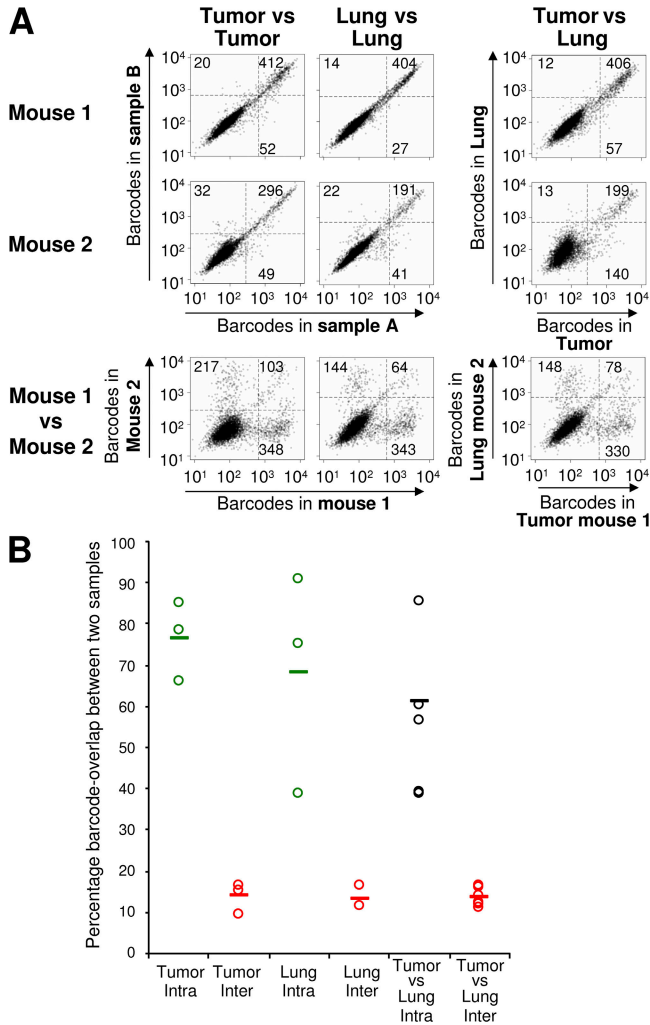
Figure 5. T cells present at lung and tumor effector sites are derived from the same precursors. Barcode analysis of T cell populations from effector sites in mice that received 10,000 barcode-labeled OT-I T cells and that were subsequently challenged intranasally with WSN-OVA influenza virus (day 0) and s.c. with EL4-OVA cells (day 1). At day 8 after challenge, T cells were isolated from tumor and lung tissue for barcode analysis. Results are representative of two independent experiments in which seven mice were analyzed separately in total. (A) Representative dot plots of the fluorescence intensities of the barcode microarray spots. Left and middle, top two rows: dot plots of barcode analysis of two pools of T cells isolated from the same tissue for two individual mice. Right, top two rows: dot plots of barcode analysis of T cells isolated from tumor versus T cells isolated from lung for two individual mice. Bottom row: evaluation of background overlap by comparison of samples from mouse 1 with samples from mouse 2. Numbers indicate the barcodes that are present within each quadrant (black rectangles; cutoff used, P < 0.00005). (B) Percentage of barcode overlap between different tissue samples. Each sample was compared with a second sample generated from the same tissue (tumor or lung) or from a different tissue (tumor vs. lung), either of the same mouse (intra) or of a different mouse (inter). The percentages indicate the number of barcodes that is present within the top right quadrant as a fraction of the total number of barcodes present in the top left, top right, and bottom right quadrant (as indicated in A). By normalizing for background and maximal overlap, data can be converted to percentage

top two rows] and B; and Fig. S4 B). In some mice, a small fraction of barcodes was enriched in gut tissue, suggesting that selective migration toward the gut may occur to some degree. However, the majority of T cell families also appeared to be equally represented at these two effector sites (translating into an identity of 64% between T cell families in gut and tumor of the same mouse). As expected, tissue samples derived from different mice did not show such a shared set of barcodes (Fig. 6, A and B, bottom rows). Together, these data indicate that also in this comparison, the majority of T cell families that locally develop accumulate at both effector sites.

Finally, a possible limitation of the experiments described in the previous paragraphs is that the 3-d in vitro culture used to introduce the Moloney-based retroviral barcode library may conceivably influence T cell migration properties. To address this issue, we have developed a strategy in which naive barcode-labeled T cells are obtained via the transduction of OT-I thymocytes (see Materials and methods for details). Also, when naive barcode-labeled OT-I T cells were used, a high degree of overlap between barcodes present in T cells that accumulated in the lung and the tumor of the same mouse was observed (Fig. S5, A [right, top two rows] and B, available at http://www.jem.org/cgi/content/full/jem.20072462/DC1).

DISCUSSION

In this paper, we developed a novel approach, termed cellular barcoding, to dissect lineage relationships between different cell populations. In this approach, retroviral or lentiviral tagging of individual cells with molecular barcodes is coupled to a microarray-based detection system to allow kinship analysis of the progeny of such cell populations. By analyzing family relationships between different barcode-labeled T cell populations, we demonstrate that cellular barcoding can be used to distinguish cells that are kin or nonkin. In essence, this technology provides an in vivo system for high throughput parallel clonal assays.

To exemplify the potential of cellular barcoding, we analyzed the kinship of CD8[+] T cells that accumulate at different DLN and effector sites. A picture that has emerged from recent studies is that T cells primed within a specific LN, or by antigen-presenting cells obtained from LNs draining a specific tissue site, acquire a phenotype that is consistent with preferential migration (10–13, 16, 17). Furthermore, cells that are induced to display such a phenotype in vitro also show selective homing upon injection (11, 12). Interestingly, one study showed that the in vivo expression of markers that could result in preferential migration is transient, possibly through the action of signals encountered in other LN beds (13). In addition, this reprogramming has also been observed in vitro (10, 12).

identity between T cell families at both sites (numbers in text). Percentage identity of T cell families in the control comparison (tumor-X vs. lung-Y) was 0%.
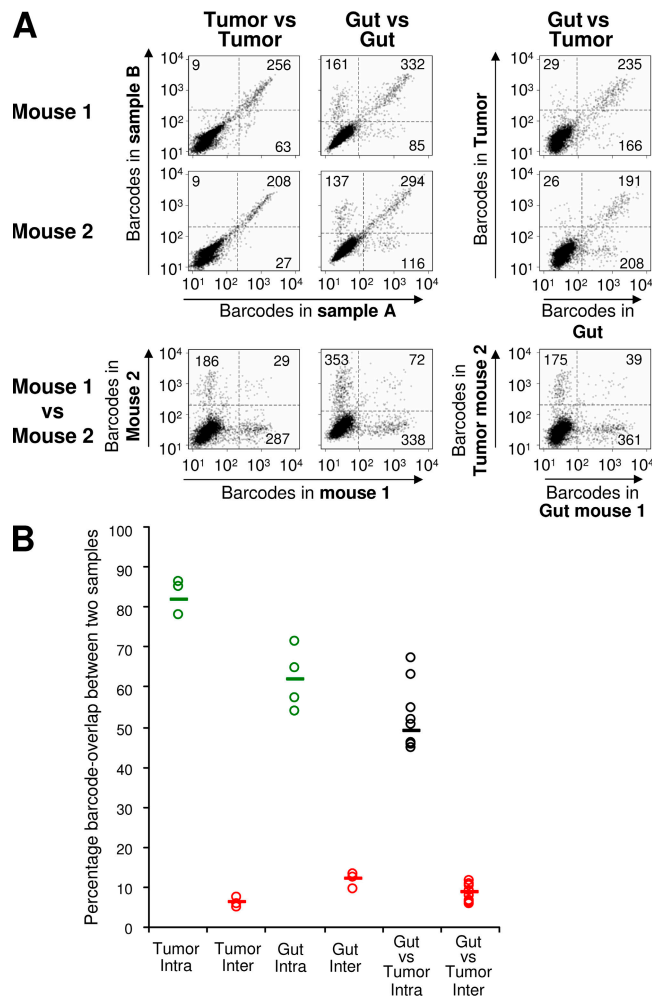
Figure 6. T cells present at gut and tumor effector sites are derived from the same precursors. Barcode analysis of T cell populations in mice that received 10,000 barcode-labeled OT-I T cells and that were subsequently challenged orally with LM-OVA and s.c. with EL4-OVA cells (both on the same day). At day 7 after challenge, T cells were isolated from tumor and gut tissue for barcode analysis. Before analysis, T cells isolated from gut tissue from each mouse were split into two half-samples that were subsequently cultured for 4 d. Results represent data from four mice analyzed separately within one experiment. (A) Representative dot plots of the fluorescence intensities of the barcode-microarray spots. Left and middle, top two rows: dot plots of barcode analysis of two pools of T cells isolated from the same tissue for two individual mice. Right, top two rows: dot plots of barcode analysis of T cells isolated from the gut versus T cells isolated from tumor for two individual mice. Bottom row: evaluation of background overlap by comparison of samples from mouse 1 with samples from mouse 2. Numbers indicate the barcodes that are present within each quadrant (black rectangles; cutoff used, P < 0.0005. (B) Percentage of barcode overlap between different tissue samples. Each sample was compared with a second sample generated from the same tissue (gut or tumor) or from a different tissue (gut vs. tumor), either of the same mouse (intra) or of a different mouse (inter). The percentages indicate the number of barcodes that is present within the top right quadrant as a fraction of the total number of barcodes present in the top left, top right, and bottom right quadrant (as indicated in A). By normalizing for background and maximal overlap, data can be converted to percentage

In the current paper, we have used cellular barcoding to study the kinship between T cells that accumulate at different effector sites in situations where antigen is available at multiple locations. Data obtained in this analysis demonstrate that from the first time point at which substantial amounts of T cells accumulate at the tumor and lung effector sites, the T cell families at both sites are largely identical, even though unique T cell families are found in the LNs draining these tissues shortly after antigen challenge. In addition, a high identity was also observed when comparing T cell families that accumulate in inflamed gut and s.c. tumor tissue. In this case, a small set of barcodes appears enriched in gut tissue in at least some animals. Although these data suggest that a minority of T cell families might display the capacity to selectively migrate to this site, based on signal intensity, the contribution of this set of clones to the total T cell pool in gut is not expected to be large. Collectively, these data indicate that after a brief period of local accumulation, the dominant pattern of T cell accumulation at effector sites is aselective. We speculate that selective T cell migration may only play a significant role at very early time points during infection (13) and perhaps in situations where the effector sites generate limited inflammatory signals. In contrast, we suggest that aselective migration forms the dominant process at the moment two effector sites produce sufficient proinflammatory signals, and aselective migration, in fact, seems a preferred strategy to combat spreading infections.

Interestingly, a strong time-dependent increase in the relatedness of T cell families isolated from lung DLN and tumor DLN of the same mouse was observed. Furthermore, this increase was found to be dependent on lymphocyte egress from the LN, as it was blocked in the presence of FTY720. The mingling of T cell families at two LN sites is consistent with the LN recycling of recently activated T cells observed by Liu et al. (13) but also with the recent description of effector T cell entry into reactive LNs via high endothelial venules (18). As a result of such T cell redistribution over tumor and lung DLN, an indiscriminate accumulation of T cell families at the two effector sites may occur as a secondary effect (Fig. 7 A). As an alternative model, accumulation of effector T cells at both sites could occur independently of LN cross-seeding (Fig. 7 B), and equilibration of LN T cell populations could then occur as a secondary effect, through T cell recycling from effector sites via the afferent lymph (19). In view of the rapid equilibration of T cell families in lung and tumor DLN between days 6 and 8, a time point at which T cell numbers at effector sites are still low, we favor the former possibility, but formal evidence is lacking.

Together, the current data demonstrate that cellular barcoding can be used to dissect the migration patterns of
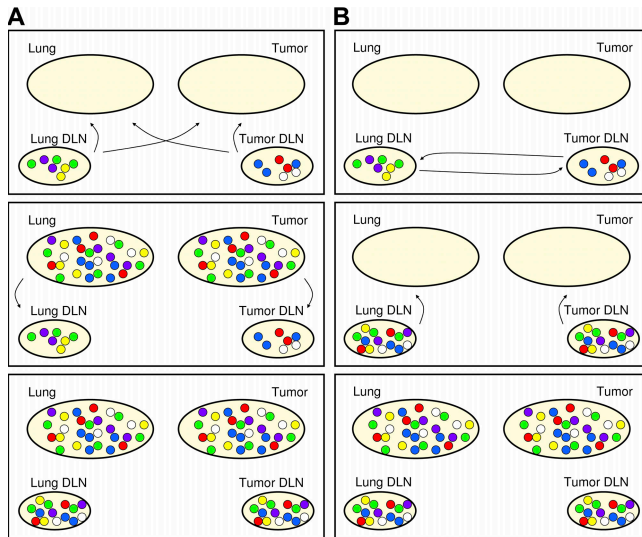
Figure 7. Pathways for intermingling of T cell families. (A) After priming, two local and genetically distinct populations of antigen-specific T cells develop. Upon LN exit, these T cell families accumulate at both effector sites, and mixing of T cell families within the DLN occurs as a secondary phenomenon upon migration or passive transport of T cells via afferent lymph. (B) After priming, two local and genetically distinct populations of antigen-specific T cells develop. Upon LN exit, these T cell families redistribute over the reactive LNs (in which the possibility of reprogramming of migration patterns may exist (11), and both effector sites are subsequently seeded by T cells from all families, regardless of the site of original priming. As evidence for both pathways of effector T cell entry into LNs exists (12, 13), it is possible that both processes operate in parallel.

T cell families in vivo and show that the majority of T cells that are activated during a T cell response acquire the capacity to accumulate at multiple tissue sites. In addition to the potential value of the barcoding technology for the analysis of cellular differentiation and migration within the lymphoid system, we consider it likely that this technology can also be used to examine stem cell, cancer stem cell, and differentiation issues for other hematopoietic and transplantable cell types.

MATERIALS AND METHODS

Mice. C57BL/6 (B6), B6 Ly5.1[+] and C57BL/6.OT-I mice (OT-I) (20) were obtained from the animal department of the Netherlands Cancer Institute. All animal experiments were performed 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. The semirandom stretch $((N)_8-(SW)_5)_5-N_8$ was inserted 3′ of the GFP ORF in the pLentilox3.4 vector (Supplemental Materials and methods, available at http://www.jem.org/cgi/content/full/jem.20072462/DC1). Subsequently, 4,743 individual *E. coli* cell clones containing this vector were picked and grown in 96-well plates. This master library was used to generate a barcode microarray and a Moloney-based retroviral library (supplemental Materials and methods). To generate the barcode microarray, PCR products were generated from the individual *E. coli* cell clones of the master library and were spotted in duplicate onto poly-L-lysine–coated glass slides using the MicroGrid II arrayer (Isogen Life Science). PCR products of 93 additional *E. coli* cell clones that

are not present in the master library were spotted in hexaplicate to function as internal controls for the labeling and hybridization procedure (supplemental Materials and methods).

Retroviral transduction procedure. Retroviral supernatants were generated using Phoenix-E packaging cells, as described previously (21), 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), labeling with anti-PE or anti-APC Microbeads (Miltenyi Biotec), followed by Midi-MACS enrichment (Miltenyi Biotec). Enrichment was ≥90%. Cells were activated in IMDM (Invitrogen) supplemented with 8% heat-inactivated FCS, 100 U/ml penicillin, 100 μg/ml streptomycin (Boehringer Ingelheim), and 0.5×10^{-5} M β-mercaptoethanol (culture medium) in the presence of ConA and IL-7 and transduced with diluted retroviral supernatants by spin infection as previously described (21). Retroviral supernatants were used at dilutions resulting in transduction efficiencies of ∼1–2%. At this transduction efficiency, the mean number of barcodes per cell is close to one (Fig. S1). 24 h after retroviral transduction, transduced T cells were purified using Ficoll Paque PLUS (GE Healthcare), 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 culture medium and twice in HBSS (Invitrogen), resuspended in HBSS, and injected i.v. (1,000–10,000 GFP[+] cells per mouse). To obtain naive barcode-labeled T cells, OT-I thymocytes were transduced with the barcode library in culture medium supplemented with 10 ng/ml IL-7. The next day, transduced cells were sorted for GFP expression and injected in the thymi of B6 recipient mice (∼10[6] cells/mouse). 3 wk after intrathymic injection, CD8[+] T cells containing barcode-labeled OT-I cells were isolated from spleen and LNs. Barcode-modified OT-I T cells generated in this manner are phenotypically naive and respond to antigen encounter with identical kinetics as unmanipulated OT-I T cells (unpublished data).

Influenza A virus, *L. monocytogenes*, and tumor challenge. The recombinant influenza A strain WSN-OVA that expresses the H-2K[b]–restricted OVA$_{257-264}$ epitope (SIINFEKL) has been previously described (22). A recombinant influenza A virus that expresses epitope IV of the SV40 large T antigen was generated as previously described (22). Influenza A strains were grown in and titered on MDCK (Madin Darby canine kidney) cells. Mice were infected with 500 PFU of the influenza A variants by intranasal application. EL4-OVA tumor cells were produced by retroviral transduction of EL4 thymoma cells with a pMX vector that encodes a GFP-OVA$_{257-264}$ fusion protein (23), and EL4-NP cells have been previously described (24). For tumor challenge, tumor cells were washed three times with HBSS, resuspended in HBSS, and injected s.c. into mice (10[7] cells per mouse). The LM-OVA recombinant *L. monocytogenes* strain that expresses OVA (25) was provided by D. Busch (Technical University Munich, Munich, Germany). Mice were infected by gavage with ±10[8] CFU LM-OVA or i.v. with ±2.5 × 10[4] CFU LM-OVA.

Isolation of barcode-labeled T cells, in vitro culture, and flow cytometric analysis. Cell suspensions of spleen, LN, lung, and tumor tissue were generated as previously described (26). Cell suspensions of gut tissue were generated according to a previously published procedure (24). In brief, the small intestine was flushed with 50 ml HBSS containing 2% FCS. Peyers patches were excised and intestine was cut into small pieces longitudinally and laterally. After washing, the tissue was transferred to a 50-ml tube containing 20 ml HBSS/5% FCS/0.1mM EDTA and stirred for 20 min at 37°C. The supernatant was harvested, EDTA treatment was repeated twice, and all supernatants were pooled. Subsequently, the material was digested in 20 ml IMDM containing 200 U/ml collagenase (type VIII; Sigma-Aldrich) and stirred for 1 h at 37°C. The supernatant was harvested and collagenase treatment was repeated once. Supernatants were pooled and the thus-obtained cell suspension was used for subsequent enrichment

of OT–I cells. To enrich OT–I cells from tissue 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. For analysis of barcodes from T cells present in LNs, LN cells were first expanded for 4 d in culture medium supplemented with 20 U/ml of recombinant human IL-2 (Chiron) and 12 ng/ml of recombinant mouse IL-7 (PeproTech). Half-samples were cultured separately to avoid bias introduction during in vitro culture. The mAbs used for flow cytometric analysis were APC-conjugated anti-CD8α (BD Biosciences) and PE-conjugated anti-Ly5.2 (eBioscience). Before analysis, 1 µg/ml propidium iodide (Sigma-Aldrich) was added to enable selection for propidium iodide–negative (living) cells.

Barcode recovery, microarray hybridizations, and data analysis. Genomic DNA was isolated using a DNeasy tissue kit (QIAGEN) and barcodes were amplified by nested PCR (supplemental Materials and methods). PCR products were purified, concentrated, labeled with Cy3 or Cy5 fluorescent groups, and hybridized to the barcode microarray (supplemental Materials and methods). Fluorescence intensities, as quantified using Imagene 6.0 (BioDiscovery, Inc.), were corrected for background noise and normalized according to Yang et al. (27).

After averaging of duplicates, barcodes present above background were selected based on the probability that a signal differed from an artificial background distribution. In barcode microarray hybridizations, the majority of signals reflects a negative biological signal and, consequently, the mode of the signal distribution can be taken as an approximation of the mean of the background distribution. As signals below the mode are the least contaminated by positive signals, an artificial background distribution was constructed from the signals lower than the mode and an extrapolation/mirroring of these signals to the positive side of the slope. For each sample, this artificial background was determined based on the distribution of the averaged Cy3 and Cy5 signals of the hybridization in which two half-samples of the same tissue were hybridized (the cutoffs used [p-values given in the figure legends] are indicated by the horizontal and vertical dividers in the dot plots). The two dimensional plots depict Cy3 and Cy5 signals for all 4,743 barcodes that were present in the master library. For comparison of barcodes in samples analyzed in different hybridizations, we performed in silico hybridizations in which the Cy3 and Cy5 signals of the two samples were plotted against each other. Percentage identity was calculated as:

$$\frac{(A-B^{EXP})-(\sqrt{A-A^{INTER}}\times\sqrt{B-B^{INTER}})}{(\sqrt{A-A^{INTRA}}\times\sqrt{B-B^{INTRA}})-(\sqrt{A-A^{INTER}}\times\sqrt{B-B^{INTER}})}\times100\%,$$

in which A−B is the experimental comparison (EXP) and A−A and B−B comparisons of two samples from a given site, in which the samples are either from the same mouse (INTRA) or from two different mice (INTER). All numbers refer to the fraction of shared barcodes of the total number of detected barcodes in both samples. Conceptually, this calculation is analogous to the calculation of percentage of specific lysis in CTL assays, with the one distinction that the maximal and minimal signals are now defined by the efficiency of barcode recovery from two sites (A−A^{INTRA} + B−B^{INTRA} and A−A^{INTER} + B−B^{INTER}, respectively), assuming that the percentage of shared barcodes within a sample is the product of two samples that have an equal efficiency of barcode recovery.

For each barcode comparison, the Pearson Correlation between Cy3 and Cy5 signals was calculated. To avoid dominance of signals from barcodes not present within a given sample in the correlation analysis, a small fraction of barcodes present in the bottom left quadrant (1/3 of the number of barcodes present in the top left, top right and bottom quadrants) was randomly drawn and incorporated in the correlation analysis. To minimize variations in correlation values introduced by the random drawing of a subset of barcodes from the bottom left, all correlation values were averaged from 100 calculations, based on 100 random drawings from the bottom left.

Online supplemental material. Fig. S1 presents additional data showing that under the conditions used, the majority of transduced cells contains a single transgene. Fig. S2 shows the redistribution of T cell families over LN beds through time as analyzed by correlation analysis. Fig. S3 shows that the redistribution of T cell families over LN beds is dependent on LN egress. Fig. S4 shows the kinship of T cells present at different effector sites as determined by correlation analysis. Fig. S5 shows the kinship of T cells present at lung and tumor effector sites when naive barcode-labeled T cells are used. A detailed description of barcode library generation, microarray generation, barcode recovery, and microarray hybridization is available as supplemental Materials and methods. Online supplemental material is available at http://www.jem.org/cgi/content/full/jem.20072462/DC1.

REFERENCES

1. Williams, D.A., I.R. Lemischka, D.G. Nathan, and R.C. Mulligan. 1984. Introduction of new genetic material into pluripotent haematopoietic stem cells of the mouse. *Nature*. 310:476–480.

2. Dick, J.E., M.C. Magli, D. Huszar, R.A. Phillips, and A. Bernstein. 1985. Introduction of a selectable gene into primitive stem cells capable of long-term reconstitution of the hemopoietic system of W/Wv mice. *Cell*. 42:71–79.

3. Lemischka, I.R., D.H. Raulet, and R.C. Mulligan. 1986. Developmental potential and dynamic behavior of hematopoietic stem cells. *Cell*. 45:917–927.

4. Golden, J.A., S.C. Fields-Berry, and C.L. Cepko. 1995. Construction and characterization of a highly complex retroviral library for lineage analysis. *Proc. Natl. Acad. Sci. USA*. 92:5704–5708.

5. Hamann, A., D.P. Andrew, D. Jablonski-Westrich, B. Holzmann, and E.C. Butcher. 1994. Role of alpha 4-integrins in lymphocyte homing to mucosal tissues in vivo. *J. Immunol*. 152:3282–3293.

6. Johansson-Lindbom, B., M. Svensson, M.A. Wurbel, B. Malissen, G. Marquez, and W. Agace. 2003. 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.

7. Picker, L.J., T.K. Kishimoto, C.W. Smith, R.A. Warnock, and E.C. Butcher. 1991. ELAM-1 is an adhesion molecule for skin-homing T cells. *Nature*. 349:796–799.

8. Svensson, M., J. Marsal, A. Ericsson, L. Carramolino, T. Broden, G. Marquez, and W.W. Agace. 2002. CCL25 mediates the localization of recently activated CD8alphabeta(+) lymphocytes to the small-intestinal mucosa. *J. Clin. Invest*. 110:1113–1121.

9. Tietz, W., Y. Allemand, E. Borges, D. von Laer, R. Hallmann, D. Vestweber, and A. Hamann. 1998. CD4+ T cells migrate into inflamed skin only if they express ligands for E- and P-selectin. *J. Immunol*. 161:963–970.

10. Dudda, J.C., A. Lembo, E. Bachtanian, J. Huehn, C. Siewert, A. Hamann, E. Kremmer, R. Forster, and S.F. Martin. 2005. 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.

11. Mora, J.R., M.R. Bono, N. Manjunath, W. Weninger, L.L. Cavanagh, M. Rosemblatt, and U.H. Von Andrian. 2003. Selective imprinting of gut-homing T cells by Peyer's patch dendritic cells. *Nature*. 424:88–93.

12. Mora, J.R., G. Cheng, D. Picarella, M. Briskin, N. Buchanan, and U.H. Von Andrian. 2005. 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.

13. Liu, L., R.C. Fuhlbrigge, K. Karibian, T. Tian, and T.S. Kupper. 2006. Dynamic programming of CD8+ T cell trafficking after live viral immunization. *Immunity.* 25:511–520.

14. Cyster, J.G. 2005. Chemokines, sphingosine-1-phosphate, and cell migration in secondary lymphoid organs. *Annu. Rev. Immunol.* 23:127–159.

15. Rosen, H., M.G. Sanna, S.M. Cahalan, and P.J. Gonzalez-Cabrera. 2007. Tipping the gatekeeper: S1P regulation of endothelial barrier function. *Trends Immunol.* 28:102–107.

16. Calzascia, T., F. Masson, W. Berardino-Besson, E. Contassot, R. Wilmotte, M. Aurrand-Lions, C. Ruegg, P.Y. Dietrich, and P.R. Walker. 2005. Homing phenotypes of tumor-specific CD8 T cells are predetermined at the tumor site by crosspresenting APCs. *Immunity.* 22:175–184.

17. Dudda, J.C., J.C. Simon, and S. Martin. 2004. Dendritic cell immunization route determines CD8+ T cell trafficking to inflamed skin: role for tissue microenvironment and dendritic cells in establishment of T cell-homing subsets. *J. Immunol.* 172:857–863.

18. Guarda, G., M. Hons, S.F. Soriano, A.Y. Huang, R. Polley, A. Martin-Fontecha, J.V. Stein, R.N. Germain, A. Lanzavecchia, and F. Sallusto. 2007. L-selectin-negative CCR7(-) effector and memory CD8(+) T cells enter reactive lymph nodes and kill dendritic cells. *Nat. Immunol.* 8:743–752.

19. Mackay, C.R., W.L. Marston, and L. Dudler. 1990. Naive and memory T cells show distinct pathways of lymphocyte recirculation. *J. Exp. Med.* 171:801–817.

20. Hogquist, K.A., S.C. Jameson, W.R. Heath, J.L. Howard, M.J. Bevan, and F.R. Carbone. 1994. T cell receptor antagonist peptides induce positive selection. *Cell.* 76:17–27.

21. Kessels, H.W., M.C. Wolkers, M.D. van den Boom, M.A. van der Valk, and T.N. Schumacher. 2001. Immunotherapy through TCR gene transfer. *Nat. Immunol.* 2:957–961.

22. Topham, D.J., M.R. Castrucci, F.S. Wingo, G.T. Belz, and P.C. Doherty. 2001. 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.

23. Kessels, H.W., K. Schepers, M.D. van den Boom, D.J. Topham, and T.N. Schumacher. 2006. Generation of T cell help through a MHC class I-restricted TCR. *J. Immunol.* 177:976–982.

24. Wolkers, M.C., G. Stoetter, F.A. Vyth-Dreese, and T.N. Schumacher. 2001. Redundancy of direct priming and cross-priming in tumor-specific CD8+ T cell responses. *J. Immunol.* 167:3577–3584.

25. Pope, C., S.K. Kim, A. Marzo, D. Masopust, K. Williams, J. Jiang, H. Shen, and L. Lefrancois. 2001. Organ-specific regulation of the CD8 T cell response to Listeria monocytogenes infection. *J. Immunol.* 166:3402–3409.

26. Arens, R., K. Schepers, M.A. Nolte, M.F. van Oosterwijk, R.A. van Lier, T.N. Schumacher, and M.H. van Oers. 2004. Tumor rejection induced by CD70-mediated quantitative and qualitative effects on effector CD8[+] T cell formation. *J. Exp. Med.* 199:1595–1605.

27. Yang, Y.H., S. Dudoit, P. Luu, D.M. Lin, V. Peng, J. Ngai, and T.P. Speed. 2002. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* 30:e15.