Neural Crest Migration is Driven by a Few Trailblazer Cells with a Unique Molecular Signature Narrowly Confined to the Invasive Front

by Rebecca McLennan[1*], Linus Schumacher[2*], Jason A. Morrison[1], Jessica M. Teddy[1], Dennis Ridenour[1], Andrew Box[1], Craig Semerad[1], Hua Li[1], William McDowell[1], David Kay[2], Philip K. Maini[2], Ruth E. Baker[2], and Paul M. Kulesa[1,3]**

1 **Stowers Institute for Medical Research, 1000 E. 50th St, Kansas City, MO, 64110, USA**

2 **Oxford University, Centre for Mathematical Biology, Mathematical Institute, 24-29 St Giles', Oxford, OX1 3LB, UK**

3 **Department of Anatomy and Cell Biology, University of Kansas School of Medicine, Kansas City, KS, 66160, USA**

***both authors contributed equally to this work, ****corresponding author**

SUMMARY

Neural crest (NC) cell migration is vitally important to the formation of peripheral tissues during vertebrate development. However, how NC cells respond to different microenvironments to maintain persistence of direction and cohesion in multicellular streams remains unclear. To address this, we profiled eight subregions of a typical cranial NC cell migratory stream at two distinct phases of migration. Hierarchical clustering showed a significant difference in the expression profile of the lead three subregions compared to newly emerged cells. Multiplexed imaging of mRNA expression using fluorescent hybridization chain reaction (HCR) quantitatively confirmed the expression profiles of lead cells. Computational modeling next predicted a small fraction of lead cells that detect direction information is optimal for successful stream migration. Single cell profiling then revealed a unique molecular signature consistent and stable over time in a subset of lead cells within the most advanced portion of the migratory front, which we term as trailblazers. Model simulations that forced a lead cell behavior in the trailing subpopulation predicted cell bunching near the migratory domain entrance. Mis-expression of the trailblazer molecular signature by perturbation of two upstream transcription factors agreed with the in silico prediction and showed alterations to NC cell distance migrated and stream shape. These data are the first to characterize the molecular diversity within a NC cell migratory stream and offer insights into how molecular patterns are transduced into cell behaviors.

INTRODUCTION

In examples that range from primitive streak formation to mechanosensory organogenesis, several embryonic cell populations undergo persistent, directed migration in coordinated groups (Tarbashevich and Raz, 2010; Voiculescu et al., 2014; Piotrowksi and Baker, 2014). When migratory cells fail to reach a target or populate an incorrect location, this often leads to improper cell differentiation or uncontrolled cell proliferation. Thus, studies of embryonic cell migration mechanisms are important for better understanding birth defects and tumor formation.

One of the key features of embryonic cell migration is the persistent, directed movement of cells in multicellular streams. During multicellular streaming, a cell autonomously controls its cytoskeleton but moves with neighbors as a population. A long standing question has been what mechanisms regulate the persistence of direction and cohesion of multicellular streams? Large scale genomic analyses of premigratory or migrating embryonic cells, such as the neural crest, have shed light on the genes expressed in migratory versus non-migratory cells during embryonic development (Gammill and Bronner-Fraser, 2002; Gammill and Bronner-Fraser, 2003; Molyneaux et al., 2004; Adams et al., 2008; Gallardo et al., 2010; Simoes-Costa and Bronner, 2013; Simoes-Costa et al., 2014). However, what remains unclear is how gene expression varies as migrating cells respond to different microenvironments and how this transduces into observed cell behaviors. Therefore, investigative efforts that correlate molecular interrogation with in vivo cell behavior analyses will yield important insights into embryonic cell migration events at the level of both individual cells and the population.

The NC is one of the most striking examples of long distance embryonic cell migration that is accessible to manipulation and in vivo observation. NC cells emerge from the dorsal neural tube and are sculpted by intrinsic and extrinsic signals into discrete, multicellular streams throughout the head and trunk (Kulesa and Gammill, 2010). Analysis of cell behaviors from in vivo time-lapse imaging in chick (Kulesa and Fraser, 1998, 2000; Teddy and Kulesa, 2004, Kulesa et al., 2008; McKinney et al., 2011; Ridenour et al., 2014) and in chick and mouse intestine (Young et al., 2004, 2014; Nishiyama et al., 2012) has shown there are regional differences in cell speed, direction, proliferation, calcium activity, and cell morphology depending on cell position within a NC migratory stream. We previously showed that a computational model of NC cell migration, using the chick cranial NC cell behavioral data, predicted that successful cell persistence of direction and stream cohesion would result from the presence of unique lead and trailing subpopulations (McLennan, Dyson et al., 2012). Further, tissue transplantation studies in which lead cells were placed into the trailing stream or vice-versa showed that NC cell behaviors and gene expression profiles are not pre-determined but depend on stream position. What remains unclear is how the size and changes and in the molecular profile of the lead NC cell subpopulation affects persistence of direction and stream cohesion.

To address these questions, we first isolated NC cells from eight discrete subregions of a typical migratory stream in the head using laser capture microdissection (LCM) and analyzed the expression of 77 genes using real time quantitative polymerase chain reaction (RT-qPCR). We examined regional differences in gene expression and used a newly emerged fluorescent hybridization chain reaction (HCR) strategy for

multiplexed imaging of mRNA expression (Choi et al., 2010, 2014). We next isolated and profiled single lead NC cells (within the most distal portion of the invasive front) at two distinct phases of migration. We compared single cell gene expression of leaders to cells isolated from the entire stream. We identified a molecular signature unique to these selected lead NC cells (which we define as 'trailblazers') narrowly confined to the invasive front. We then used gain- and loss-of-function of transcription factors upstream of the trailblazer molecular signature to determine whether genes expressed as part of this signature are critical to NC cell migration. These experiments were performed in parallel with computational modeling that simulated our experimental scenarios. Our results highlight the molecular heterogeneity of cells during cranial NC cell migration and the requirement for only a small subset of trailblazer cells to ensure persistence of direction and stream maintenance.

MATERIALS AND METHODS

Embryos and cell labeling

Fertilized chicken eggs (Centurion Poultry) were incubated at 38 °C in a humidified incubator until the desired stages of development (Hamburger and Hamilton (HH), 1951). Plasmid DNA (2.5-5 ug/ul) or fluorescein tagged morpholinos (0.5 mM) were injected into the neural tube and electroporated at HH Stage 9 (McLennan and Kulesa 2007). We focused the morpholino knockdowns to specific developmental stages of interest with large number n values, and we did not observe any obvious off-target affects. For electroporations of the trailing neural crest, embryos were incubated

until 10-12 somites at which point plasmid DNA was injected and electroporated with CellTracker CM-DiI (C-7001, Life Technologies).

Gene profiling

(Eight segment) Tissue was harvested by LCM (Zeiss), pre-amplified using a modified version of the Cells-to-Ct kit (Ambion) and analyzed by microfluidic RT-qPCR on the BioMark HD (Fluidigm) as previously described (Morrison et al., 2012). The stream lateral to rhombomere 4 (r4) (in cryosections) was divided into eight subregions (the number of subregions selected for our ability to cut reproducibly) and harvested by LCM. Following LCM, RNA from residual cryosections produced RNA integrity numbers of 5.8-6.8 on a Bioanalyzer 2100 (Agilent). 77 transcripts were pre-amplified from cDNA using 14 cycles per the Cells-to-Ct protocol (Life Technologies). Preamplified cDNAs were diluted with sterile 1X TE and the products analyzed on a Fluidigm BioMark HD at the Children's Hospital Boston IDDRC Molecular Genetics Core. Median Absolute Deviation (MAD) was used to eliminate outliers, resulting in 3-6 biological replicates per sample. Different outliers were automatically removed when comparing subregions 1-8 than comparing subregions 1-3. A trio of reference genes, selected from six candidates, was used to calculate normalized relative quantities and differential expression in Biogazelle's qBASEplus software. Hierarchical clustering by Pearson's dissimilarity was performed using Partek's Genomics Suite software package.

(Single cell) For the quartile analysis, four regions of interest (the number of subregions selected for our ability to manually dissect reproducibly) from the cranial r4

migratory stream, electroporated with Gap43-yellow fluorescent protein (YFP), were chilled 0.1% DEPC PBS and mechanically and chemically dissociated. For the trailblazer analysis, the tips of the arches, containing an average of 8 electroporated NC cells were manually removed from the rest of the embryo and then mechanically and chemically dissociated. Single and healthy YFP+ NC cells were isolated by fluorescence-activated cell sorting (FACs) into Cells-to-Ct lysis solution with 1:100 DNAse I and incubated at room temperature for 20 minutes. Bioanalyzer 2100 RINs determined from tissue remaining after LCM measured >6.8 for neural tube cultures and >8.9 for cryosectioned tissue. cDNA was synthesized directly from entire lysates using the High Capacity cDNA kit (Life Technologies). An 18-cycle, Cell-to-Ct pre-amplification protocol was employed to selectively amplify 96 transcripts with Taqman Gene Expression Assays (Life Technologies). Pre-amplification products were diluted with sterile 1X TE before being run on Fluidigm's BioMark HD. Data were analyzed using Fluidigm's Singular2 software package in R. Limit of detection (LoD) was set at 28 for the single cell isolation method comparison and 26 for the trailblazer and quartile single cell profiling. We calculated normalized relative quantities and differential expression in Biogazelle's qBASEplus software and determined hierarchical clustering and intensity plots using PARTEK's Genomics Suite.

Fluorescent multiplex in situ hybridization (HCR)

Transcripts were visualized in whole embryos and tissue sections by HCR. Chick embryos were incubated to specific developmental stages, tissue rapidly collected in chilled 0.1% DEPC PBS and fixed in fresh 4% paraformaldehyde at ambient temperature for 1 hr. Embryos were then dehydrated and rehydrated before HCR was

performed according to the manufacturer's instructions (Molecular Instruments, Caltech, USA). Tissue sections were cut on a vibratome (Leica VT1000S). In some cases, immunohistochemistry was performed on the tissue following HCR for NC cell specificity. All samples were imaged by confocal microscopy (Zeiss, LSM 780) and quantitation of fluorescence signal applied polyline kymograph analysis (Jay Unruh, Stowers) in Image J (FIJI).

Mathematical modeling

To test the logical conclusions of our mechanistic hypotheses, we used a hybrid computational model (McLennan, Dyson et al, 2012), with individually represented cells and a continuous chemoattractant concentration (Fig. 4A-D). In the model, cells undertake a two-dimensional off-lattice random walk on a growing rectangular domain that represents the NC cell migratory environment (Fig. 4A-D). New cells enter the domain from one end throughout the simulation (Fig. 4D). The chemoattractant concentration is modeled by a reaction-diffusion equation, with the cells acting as sinks, representing internalization of chemoattractant. The key model components are illustrated in Fig. 4, Supplementary Fig. 6, and model parameters are listed in Table 1 (further model information is in the Supplementary Material).

We extended our original computational model (McLennan, Dyson et al., 2012) to incorporate new experimental findings and performed simulations in parallel with the experiments presented in this paper (Figs. 4, 8). Key changes include:
(1) a wider stream of cells that allows for greater cell numbers and more adequate representation of multicellular stream migration; (2) limiting the sensing accuracy of

cells (for chemotaxis). Based on Berg and Purcell (1977) we derived (order of magnitude) bounds of how small a local chemoattractant gradient can be relative to the bulk concentration before cells cannot sense it; (3) assuming that when cells are in contact, the length of the filopodium giving rise to the contact is not fixed. Instead, a range of intercellular distances is allowed, with a maximum beyond which cells lose contact and cease to communicate directional information. These distances are based on empirical data (Supplementary Fig. 6) and this change improves stream cohesion and reduces stream break-up in model migration.

RESULTS

Molecular profiling reveals regional diversities in gene expression within a NC stream

We previously used computational modeling, cell morphometrics, and broad molecular profiling to reveal the existence of at least two subpopulations of cells with distinct molecular profiles within a typical cranial NC cell migratory stream (McLennan, Dyson et al., 2012). This suggested there may exist a much richer phenomena whereby NC cells alter cell behaviors and gene expression as cells travel through different microenvironments to precise targets.

To address this question, we subdivided the pre-otic, cranial NC migratory stream into eight subregions (Fig. 1A-D). Hierarchical clustering of 77 genes revealed distinct differences in gene expression between each of the eight subregions (Fig. 1E, F). Subregions 1-3, 4-5, and 6-8, clustered together, respectively (Fig. 1F). When we organized the genes according to their level of linear expression in subregion 1

(migratory front) relative to subregion 8 (newly emerged), we found non-linear transitions of gene expression between the clustered subregions (Fig. 1F). Some of the genes highly expressed within the migratory front (subregion 1) were dramatically diminished in expression towards the proximal subregions (Fig. 1F). Likewise, some genes expressed at low levels within the migratory front (Fig. 1F; bottom half of graph) showed a significant increase in expression when compared throughout the more proximal subregions. The lead three subregions (1-3; corresponding to the invasive front and distal portion of the cranial NC cell migratory stream) cluster very closely with one another and are distinct from the newly emerged NC cells (subregions 6-8) (Fig. 1G). Of the profiled genes, 44% (34/77) are differentially expressed between these lead three subregions and the more proximal trailing cell subpopulations (subregions 6-8) (Fig. 1G). These data reveal a widespread regional diversity in gene expression profiles based upon cell position within the cranial NC cell migratory stream.

Hierarchical clustering shows uniqueness of the distal stream and at least four characteristic patterns of gene expression

To determine whether the NC cells within the most advanced portion of the migratory front display a unique molecular profile, we compared the gene profiles of cells from subregion 1 (lead 12.5% of migratory stream) to those cells immediately proximal in subregions 2-3 (Fig. 1H). We found that the migratory front shows a molecular profile distinct from cells within subregions 2-3 with significant up or down-regulation of 18% (14/77) of the genes analyzed (Fig. 1H). Genes significantly upregulated in subregion 1 compared to subregions 2 and 3 included *heart and NC*

derivatives expressed 2 (HAND2), aquaporin 1 (*AQP1*), BMP and activin membrane-bound inhibitor homology (*BAMBI*), glypican 3 (*GPC3*) and matrix metalloproteinase 2 (*MMP2*) (Fig. 1H). When we compared the molecular profile of lead (subregion 1) NC cells to the rest of the stream (subregions 2-8), we found similar genes were enriched in the migratory front (*HAND2*, *GPC3*, *MMP2*) (Fig. 1I). Thus, there is a unique molecular profile of the most advanced migratory front.

To determine whether the NC molecular profiles along the migratory stream could be categorized, we examined the expression of individual genes (Fig. 2). First, we found patterns where genes were more highly expressed in the migratory front and expression decreased proximally (Fig. 2A). Second, we observed patterns that peaked mid-stream (Fig. 2B). Third, we found that some genes were highly expressed in the newly emerged NC cells in a pattern that diminished towards the migratory front (Fig. 2C). Lastly, we did find that some genes had uniform expression within all subregions of the stream (Fig. 2D). Thus, the regional differences in neural crest gene expression varied and displayed at least four characteristic patterns.

Multiplexed imaging of mRNA expression using HCR confirms the regional diversity of gene expression within the neural crest stream

To more carefully validate regional differences in gene expression within the NC cell migratory stream, we applied a newly developed technology known as fluorescent HCR (Choi et al., 2010, 2014). The HCR method detects and amplifies mRNA signals within cells and provides a fluorescent readout of expression (Choi et al., 2010, 2014).

We applied HCR to simultaneously visualize the expression profiles of two genes highly expressed by cells within the most invasive front (*HAND2* and *BAMBI*) and combined HCR with HNK-1 immunolabeling to distinctly detect NC cells (Fig. 3A,B). Visual observations confirmed that lead NC cells expressed high levels of *BAMBI* and *HAND2* at HH Stage 15 (Fig. 3A, C). Quantitative analysis of the HCR signals further confirmed the patterns of *BAMBI* and *HAND2* expression (Fig. 3D). Thus, HCR analysis allowed us to better visually and quantitatively confirm regional differences in gene expression, including high expression of specific genes within lead cells at the migratory front.

Computational modeling predicts that a small fraction of lead cells is optimal for successful stream migration

To test aspects of our hypothesis that are difficult to probe in vivo, we use our extended computational framework. Specifically, we asked whether the number of lead cells that can detect spatial gradients in chemoattractant is a critical factor for the success of NC cell migration. Should this be the case, then the gene expression patterns detailed in vivo (Figs. 1,2) would be consistent with this constraint. To test the effect of varying the number of leaders, we first restricted the model simulations to only include non-plastic cell behaviors, such that individual cells that begin migration as leaders (or followers) could not switch from being a leader to a follower (or vice versa). To change the fraction of all migrating cells that are leaders, we varied the time, t_LF at which new cells entering the domain are prescribed followers instead of leaders (Fig. 4I): that all cells that entered the migratory domain up to time t_LF were specified as leaders, and cells that entered after that were specified as followers.

Our simulations reveal that the median distance migrated and the stream density both increase with decreasing leader fraction (Fig. 4I; compare Movie S1 with low (0.051) vs high (0.59) mean leader fraction). While the furthest distance migrated does not change noticeably, it is the movement of cells away from the entrance to the migratory domain (corresponding to near the dorsal neural tube) that proves critical for the successful migration pattern (Fig. 4I,J). This prevents jamming near the domain entrance and enables a higher number of cells to distribute more evenly along the migratory pathway (Fig.4I, J). This trend cannot be extrapolated to zero leaders, in which case the followers move in a random, undirected manner and stay close to the entrance of the domain (Movie S2). Thus, our computational model predicts that a small number of lead cells can efficiently guide the migration of the entire NC cell migratory stream.

Single cell analysis of the migratory front identifies a subpopulation of NC cells that we define as trailblazer cells with a stable and consistent set of highly expressed genes

To test our computational model prediction of a small leader fraction within the migratory front, we performed single cell analysis and looked for evidence of a stable and consistent set of highly expressed genes within leaders. We isolated an average of 8 lead NC cells per embryo by manually removing the tissue containing the first few fluorescently labeled NC cells, dissociating the tissue and then cell sorting via FACS at HH Stages 13,15 (Fig. 5). We determined FACS isolation to be the most efficient method (as compared to LCM both in vitro and in vivo) for analyzing single NC cells

while maintaining the native molecular profile (Supplementary Fig. 2). When we profiled 96 genes (composed of the 77 genes described above and additional genes of interest), we discovered a stable and consistent set of genes expressed by a small fraction of the lead NC cells that we define as 'trailblazers' (Fig. 5).

Trailblazers have a high degree of gene expression homogeneity during both phases of migration examined, as can be seen in the profiles of individual cells (Fig. 5A). Principle component analysis (PCA) plots of the genes during each phase of migration analyzed show that over 70% of the genes are stably expressed with little variation (Fig. 5A, blue squares; Supplementary Table 2). Violin plots confirm similar levels of expression for a range of genes, including *BAMBI*, *NOTCH1* and *CXCR1* (Fig. 5C). When we focused our attention on the highest expressed genes with RT-qPCR Ct values of less than 22, we found that 98% (61/62) of the most highly expressed genes within the migratory front at HH Stage 15 are also highly expressed at HH Stage 13 (Supplementary Table 2). Many of these genes are consistently expressed by a large percentage of the profiled trailblazers at both phases of migration, including BAMBI, *CXCR1*, *NOTCH1*, plakophilin 2 (*PKP2*) and *transcription factor AP-2 alpha (TFAP2A)* (Fig. 5B, C; Supplementary Table 2).

Single cell analysis also shows genes whose expression patterns were dramatically different in the trailblazers when comparing the phases of migration analyzed (Fig. 5B, C). For example, *HAND2* is expressed in a small number of lead NC cells (7%) during the first phase of migration (Fig. 5C, Supplementary Table 2). This dramatically increases during the second phase to 59% of the lead cells (Fig. 5C, Supplementary Table 2). In contrast, the percentage of NC cells highly expressing

Sox10 dramatically decreases (from 69% to 21%) over time (Fig. 5C, Supplementary Table 2). To address whether BAMBI and HAND2 were expressed by newly emerging NC cells, we analyzed HCR expression patterns starting at HH St 9 (Supplementary Fig. 1C). Migratory and pre-migratory NC cells were visualized with either FoxD3 or HNK-1 expression. As the first pre-otic NC cells delaminate from the dorsal neural tube, we found that cells express BAMBI (Supplementary Fig. 1C (box); HH Stage 9). There were a small number of lead cells that expressed high levels of BAMBI at all phases of migration examined (Supplementary Fig. 1C (box); Figure 3). In contrast, we found that HAND2 was not expressed by newly emerged NC cells (Supplementary Fig. 1C (box), Fig. 3). Therefore, we restricted our focus to genes that are expressed highly by at least 50% of the cells at both phases of migration (Supplementary Table 2, bolded genes). Together, these data confirmed that a stable and consistent set of genes is expressed within the trailblazers, after cells encounter the neural crest microenvironment.

Trailblazers have a unique molecular signature and are narrowly confined within the migratory front

To determine the molecular signature unique to trailblazers, we compared the stable and consistently expressed genes to those expressed by NC cells within the entire stream at HH St 15 (Fig. 6A). We measured the similarity of molecular profiles across 318 single NC cells by PCA, violin plots, hierarchical clustering, pairwise correlation and Pearson's correlation (Fig. 6B-D, Supplementary Figs. 3, 4). Our analysis revealed four main results. First, the vast majority of cells from the quartile subregions have a poor correlation with the molecular profile of the trailblazers (Fig.

6B). At the single cell level, PCA shows incomplete overlap between genes expressed by trailblazers and quartile 1 (Fig. 6B). Second, NC cells within each quartile have high correlations with one another (Fig. 6B). Third, the gene expression profile of the trailblazers is distinct from all quartiles, but most similar to quartile 1 (Fig. 6B). Fourth, the average expression of NC cells in the quartiles cluster according to their position within the migratory stream as well as within the local microenvironment depending on the type of clustering method (Fig. 6D; Euclidean, Pearson). Violin plots reveal examples of individual genes that are expressed at higher (*BAMBI*, *CXCR1*, *PKP2*, *HAND2*) or lower (*Sox10*, integrin alpha 3 (*ITGA3*)) levels by trailblazers than cells located more proximal within the migratory stream (Fig. 6C). Measurements of the HAND2 and BAMBI HCR expression at higher resolution within the four quartiles confirmed the regional differences in expression (Supplementary Fig. 1A, B). Thus, the unique molecular signature associated with trailblazer cells is distinct from cells in the entire first quartile and is not shared by other migrating NC cells within the rest of the stream.

To refine the unique molecular signature of the trailblazers, we examined the genes that were differentially expressed between trailblazers and quartile 1 at HH Stage 15 (Supplementary Table 3, bolded genes). This resulted in a list of 17 genes highly expressed in trailblazers and differentially expressed between the trailblazers and migrating NC cells in quartile 1. This list of genes did not take into account comparison of the trailblazers with the remainder of the stream. Therefore, a gene expressed highly by cells in quartile 3 and trailblazers would remain in the trailblazer molecular signature. To remove any such genes, we compared the migratory front to single cell profiles from

cells isolated throughout the stream (quartiles 2-4) and found 94% (16/17) of the genes in agreement (Supplementary Table 3, bolded genes). Thus, these 16 genes are representative of the unique molecular signature of the trailblazer cells (Fig. 9).

Disruption of *HAND2* and *TFAP2A* function alters the unique molecular signature associated with the trailblazers

To determine the changes in the unique molecular signature associated with the trailblazers (Fig. 9) to the loss of *HAND2* or *TFAP2A* function, we profiled migrating NC cells 24 hours after transfection with *HAND2* or *TFAP2A* morpholinos (HH Stage 15). We selected *TFAP2A* and *HAND2* based on the fact that they are transcription factors and their expression is enriched within cells in the lead three subregions (Fig. 1G). We discovered that when *HAND2* is knocked down, several trailblazer signature genes are significantly up-regulated including *AQP1*, *CDH11*, *CDH7*, *CXCR4* and *EPHB1* (Supplementary Table 4, bolded genes). Also, *TFAP2A* is down-regulated in *HAND2* morpholino transfected embryos, indicating crosstalk between the two transcription factors (Supplementary Table 4). Loss of *TFAP2A* function results in the up-regulation of integrin beta 5 (*ITGB5*) and *NEDD9*, and down-regulation of *CXCR4* and *EphB1* (Supplementary Table 4, bolded genes). Loss of either *HAND2* or *TFAP2A* results in up-regulation of *HAND2* expression, suggesting either negative regulation or activation of a compensatory pathway. From this, we conclude that loss of either *HAND2* or *TFAP2A* function within migrating NC cells influences the unique molecular signature associated with the trailblazer NC cells.

Knockdown of *HAND2* or *TFAP2A* expression results in alterations to the NC migration pattern

We hypothesize that genes expressed by trailblazers are critical to cell direction persistence and stream migration. To test this hypothesis, we knocked down *TFAP2A* and *HAND2* function in NC cells using morpholinos. We find that loss of *HAND2* function results in a significant reduction in the area invaded by treated versus control NC cells (Fig. 7A, F,G). Upon close examination, the fluorescence associated with individual NC cells within *HAND2*-morpholino embryos was more punctate than in control morpholinos (Fig. 7A). This phenotype did not correlate with increased cell death (Supplementary Fig. 5). We also observed some cells at the distal portion of the branchial arches in *HAND2*-morpholino embryos (Fig. 7A-E), suggesting that HAND2 is required for NC cells to properly distribute along the migratory pathway and colonize the branchial arches.

Loss of *TFAP2A* function results in dramatic alterations to the NC cell migration pattern (Fig. 7A). First, the distribution of NC cells along the migratory pathway is perturbed when *TFAP2A* is knocked down at both phases of migration due to a statistically significant drop in cell number in the lead subregion of the migratory stream (Fig. 7B-C). During the first phase of migration (HH St 13), the distance migrated and area invaded by *TFAP2A* morpholino transfected NC cells are not statistically different to control embryos (Fig. 7D, F). However, during the second phase of migration, NC cells transfected with the *TFAP2A* morpholino stopped and failed to migrate the entire length of the migratory pathway (Fig. 7E,G). From these results, we conclude *TFAP2A* is necessary for NC cell migration into the branchial arches.

Computational model simulations that overexpress trailblazer genes within trailing cells predicts alterations to the NC migration pattern

To model the overexpression of a trailblazer gene within the trailing subpopulation, we modified the model parameters to convert 50% of trailing cells at random into leaders. This mimics, in silico, the transfection of 50% of the NC cells within the trailing subpopulation with overexpression of the transcription factors of the trailblazer signature. When we forced trailing cells in silico to display a lead cell phenotype, we found that cells remain near the entrance of the migratory domain (Fig. 8A, B). This migratory pattern is similar to the model simulations scenario that introduced higher lead cell fractions (Fig. 4I,J). Thus, forcing a lead cell behavior within the trailing subpopulation in silico predicts disruption to NC cell migration that would be observed as cell bunching near the dorsal neural tube exit.

Overexpression of *HAND2* and *TFAP2A* alters the NC migration pattern in a manner consistent with computational model predictions

To experimentally test whether gain-of-function of the lead NC cell behavior would alter the migration pattern, we selectively overexpressed *TFAP2A* or *HAND2* in the trailing subpopulation (Fig. 8). When *HAND2* is overexpressed in trailing NC cells we found fewer migrating cells (Fig. 8C, D). However, these fewer migrating trailing cells distributed along the migratory pathway (Fig. 8C, D). When *TFAP2A* is overexpressed in trailing NC cells, we observed some cells bunching near the neural

tube, but for the most part cells continued to distribute along the migratory pathway (Fig. 8C-E). These results agreed with our model simulations (Fig. 8A, B) but also supported our previous tissue transplantation experiments (McLennan, Dyson et al., 2012). That is, lead cells placed into the trailing stream either stalled or re-initiated their migration to distribute along the migratory pathway, but not overtake the migratory front (McLennan, Dyson, et al., 2012). Thus, the gain-of-function, model simulations, and tissue transplantation experiments all support the hypothesis that distinct lead and trailing cell behaviors and gene expression must be maintained for proper NC cell migration.

DISCUSSION

In this study, we addressed questions in directed cell migration during vertebrate development using the cranial neural crest (NC) model. Using novel methods to isolate and profile single and small numbers of cells, we analyzed and compared NC cell gene expression profiles during distinct phases of migration. This led us to discover regional differences in gene expression within the NC stream and a consistent (Fig. 1-3) and stable molecular signature unique to the cells within the most distal portion of the migratory front (Figs. 5-6), which we termed as trailblazers. Gain- and loss-of-function experiments then revealed insights into the roles of genes associated with the trailblazer molecular signature to ensure the proper pattern of NC cell migration. In parallel, we used a hybrid computational model to simulate and predict experimental outcomes.

Only a few lead NC cells with guidance information are required for persistence of direction and stream maintenance. Our previous computational model predicted at least two separate cell subpopulations (leaders/trailers), which was confirmed by RT-

qPCR analysis that divided the stream into a 30/70 percentage split (McLennan, Dyson et al., 2012). What was unclear and difficult to manipulate experimentally was whether a change in the number of leaders would affect stream dynamics. This is where our extended computational model proved very useful. In fact, we were able to show that the furthest distance migrated was insensitive to the number of lead cells (Fig. 4). Even a few lead cells could migrate as far as the entire multicellular stream (Fig. 4). Not only that, but higher numbers of leaders were less efficient at guiding the entire stream.

Further analysis of lead NC cells provided insights into the molecular characteristics of the migratory front. Our single cell analysis led to the identification of a unique molecular signature associated with a few lead cells we termed, trailblazers (Fig. 5, 6). Trailblazers are narrowly confined to the most invasive front. The molecular signature of the trailblazers (16/96 genes) is distinct from the molecular profiles of other migrating cells within the stream (Figs. 6, 9; Supplementary Figs. 3, 4). It is important to note that the other 80/96 genes did show changes in expression over time in the trailblazer cells, suggesting a majority of the molecular profile of trailblazers is influenced by the microenvironment (Figs. 5, 6; Supplementary Figs. 3, 4). Thus, computational model simulations that predicted only a few cells are required to direct stream migration in the presence of a chemoattractant on a growing domain led to the discovery of a small subpopulation of lead cells narrowly confined to the migratory front with a consistent and stable molecular signature.

Based on our identification of a unique molecular signature associated with trailblazer NC cells, we next asked whether trailblazer gene function is important for NC cell persistence of direction and stream maintenance. Since *HAND2* expression was

high in lead cranial NC cells, we expected that knockdown of *HAND2* would affect the migration pattern. NC cells with *HAND2* knockdown did invade less area, but did continue to reach distal portions of the second branchial arch (Fig. 7). This may be explained though by the differences in *HAND2* expression levels at the two distinct phases of migration analyzed. *HAND2* expression is high at HH Stage 13 but only in 8% of the trailblazer cells. By HH Stage 15, *HAND2* is expressed highly in 59% of trailblazers. Thus, we conclude that knockdown of *HAND2* does not affect early phases of cranial NC cell migration, but may play a role in the distribution of cells throughout the branchial arch and subsequent cell differentiation (Thomas et al., 1998; Howard et al., 1999; Srivastava et al., 1997; Miller et al., 2003; Hendershot et al., 2007; D'Autreaux et al., 2007).

Knockdown of *TFAP2A* resulted in the failure of NC cells to migrate completely into the branchial arches (Fig. 7). TFAP2A has previously been shown to be expressed by premigratory and migratory cranial NC cells and be important in NC induction, proliferation and differentiation in mice, zebrafish and Xenopus, as well as avian facial tissue growth (Mitchell et al., 1991; Chazaud et al., 1996; Schorle et al., 1996; Shen et al., 1997; Pfisterer et al., 2002; Brewer et al., 2002; Luo et al., 2003; Knight et al., 2003; Knight et al., 2004; Barrallo-Gimeno et al., 2004; Li and Cornell, 2007; Wang et al., 2011). We have previously shown that when leaders are prevented from migrating into the target site by a physical barrier, trailing NC cells sense the paused leaders, reroute around the barrier and become the new leaders (Kulesa et al., 2005). Here, because NC cells throughout the stream were transfected with the *TFAP2A* morpholino, no cells were able to take on the role of the trailblazers and migration was hindered (Fig. 7).

Together, these data support our hypothesis that only a small number of trailblazer cells with an expression profile that is influenced by TFAP2A, are necessary to guide the NC stream in a directed manner.

Computational model simulations that tested the gain-of-function of trailblazer genes within the trailing subpopulation predicted alterations to the NC migration pattern (Fig. 8). When similar experimental perturbations were performed in ovo, we observed alterations to the NC migration pattern (Fig. 8). That is, when *HAND2* was overexpressed in the trailing NC cells, fewer cells were observed along the migratory pathway (Fig. 8). We initially interpreted this result as a possible NC cell delamination defect. However, after comparing the phenotype to computational model simulations (the first 50um of the computer model migratory domain corresponds to NC cell migration from the dorsal midline into the paraxial mesoderm), it is more likely that trailing NC cells that overexpress *HAND2* properly delaminate but then fail to persist in directed migration. Therefore, we hypothesize that *HAND2* expression in vivo contributes to the trailblazer phenotype. This is supported by the computational model simulations that show that leaders within the trailing portion of the stream fail to migrate due to a lack of chemoattractant to follow. Trailing NC cells that look for and follow the stalled leaders also become stalled as a result. Although, we do not have direct in vivo evidence that lead cells detect spatial gradients of chemoattractants differently from trailers, our results clearly show that association of lead cell behaviors with genes expressed at the invasive front is valid. In contrast, when *TFAP2A* was overexpressed in the trailing portion of the migratory stream, no obvious defects were observed (Fig. 8). Since *TFAP2A* was more broadly expressed in the lead subregions of the NC cell

migratory stream and not restricted to the most distal migratory front (Fig. 1H), this result was not surprising.

Theoretical testing of the effect of the number of lead cells to drive multicellular stream migration required that we restrict our in silico experiments to the case of non-plastic lead and trailing cell behaviors. We did this even though our previous tissue transplantation experiments showed chick cranial NC cells alter their gene expression profile depending on cell position within a stream. Here, we disabled phenotype switching in our computational model simulations, for otherwise we would have only been able to change the leader fraction transiently, before phenotype switching restored it towards the unperturbed case. Alternatively, we could have set up the computational model simulations with varying initial lead cell fractions and, given appropriate microenvironmental parameters, observe what leader fraction emerges naturally from phenotype switching. However, this would require at least a phenomenologically correct implementation of switching, which in turn has to be verified experimentally. This is outside the scope of the current manuscript, but will be addressed in future work.

In summary, we show that the embryonic NC microenvironment regulates the gene expression profile and pattern of cranial NC cell migration in a manner that is dependent on cell position and phase of migration. This regulation was identified by the presence of a unique molecular signature associated with trailblazer NC cells that are narrowly confined to the distal migratory front. These data support the hypothesis that a few lead NC cells interpret complex microenvironmental signals differently than other migrating NC cells within the multicellular stream. To test our hypothesis and importance of the unique molecular signature, we showed that mis-expression of

transcription factors *TFAP2A* and *HAND2* may result in significant alterations to NC cell migration pattern. We speculate, based on experiment and computational model simulations, that only a few trailblazers drive directed migration, and perturbations in the identity and function of trailblazers lead to migration pattern defects. When a gene associated with the trailblazer signature is over-expressed (experiment) or the trailblazer cell behavior is forced within the trailing subpopulation (simulation), cells are stranded near the migratory domain entrance. We postulate this is due to a lack of guidance information. These findings were made possible by studying NC cellular and molecular dynamics within the embryonic microenvironment using a closely integrated experimental and theoretical approach.

ACKNOWLEDGEMENTS

PMK would like to thank the Stowers Institute for Medical Research for their kind generosity. LJS would like to thank Louise Dyson for helpful discussions and access to the original code and gratefully acknowledges the U.K.'s Engineering and Physical Sciences Research Council (EPSRC) for funding through a studentship at the Life Science Interface programme of the University of Oxford's Doctoral Training Centre. REM and JAM would like to thank the Histology Core at the Stowers Institute for Medical Research for assistance with cryosectioning.

FIGURE LEGENDS

Figure 1. Distinct regional expression profiles exist within the NC migratory stream. (A-E) Schematic of gene profiling approach. (F) Hierarchical clustering of the 8 subregions from the HH St 15 stream by Pearson dissimilarity based upon the 77-gene profile. All samples shown are relative to subregion 8 with genes ordered by their level of linear expression in subregion 1. (G) Heat map of gene expression differences between leading subregions 1-3 relative to trailing subregions 6-8. (H) Heat map gene expression differences between leading subregion 1 relative to subregions 2-3. (I) Heat map gene expression differences between leading subregion 1 relative to the rest of the stream. e, eye; nc, neural crest; nt, neural tube; r, rhombomere; ba, branchial arch; n, notochord.

Figure 2. Gene expression analysis reveals pattern non-linearities within a typical NC stream. (A-D) Patterns of individual genes show expression (A) increased at the migratory front, (B) peaked within the mid-stream, (C) increased in the trailing subpopulation or (D) uniform throughout the stream. Each subregion is represented by 3-6 biological replicates. (A') Key indicating the eight subregions of the stream. (E) Example line graphs of normalized expression within the stream. Error bars depict the SEM.

Figure 3. **HCR and immunolabeling confirm regional differences in gene expression within the NC stream.** (A) Whole embryo staining and expression pattern of HAND2 and BAMBI HCR in the head at HH Stage 15. The arrowheads point to the first and second branchial arches. (B) Schematic of HCR and immunolabeling

approach. (C) Transverse view of HNK-1 staining; BAMBI and HAND2 HCR at the pre-otic level of a HH Stage 15 embryo. (D) Polyline kymograph analysis (in a 20-um wide band) of the BAMBI and HAND2 fluorescent HCR signal intensity along (D') the NC cell migratory pathway (n=5). The scale bars are all 100um. nt, neural tube; ov, otic vesicle; ba, branchial arch.

Figure 4. Computational experiment that varies the leader fraction predicts stream migration is more efficient with fewer leaders. (A) Cells internalize local chemoattractant. (B) Leader cells (orange) randomly sample directions and move up the chemoattractant gradient. (C) Follower cells (purple) randomly search for other cells and, when in contact with another cell, adopt their direction of movement. (D) Cells are inserted into the domain (x = 0um) throughout the simulation, but only if there is space available. (E) The domain lengthens over time to mimic tissue growth, diluting chemoattractant and contributing to cell movement. (F-H) A sample simulation. The grey contour represents where the chemoattractant gradient is too shallow to be sensed. (I) Change to the number of leader cells; n = 20 stochastic simulations with five different average leader fractions <f_L>. Greyscale boxplots of the distance migrated in the x-direction are overlaid. The color of each boxplot shows the relative stream density (J) A histogram of cell counts vs. distance migrated for five different leader fractions.

Figure 5. Trailblazer NC cells have a partially conserved molecular profile across distinct phases of migration. (A) Heat map of single trailblazer NC cells at HH Stage 13 (n=72/356 cells profiled, isolated from m=35 embryos) and Stage 15 (n=76/212 cells profiled, isolated from m=38 embryos). (B) PCA projections of 96 genes analyzed in single HH Stage 13, 15 trailblazer cells. Blue, shaded squares within each plot

represent an arbitrary PC score (PC1-2) of less than 0.15. (B') and (B'') are magnified insets from PCA projections. (C) Violin plots of selected genes with a violin plot reference key.

Figure 6. Single trailblazer NC cells have a unique molecular profile. (A) Isolation of single NC cells (blue circle) from each quartile of the cranial NC stream. (B) PCA of single trailblazer and quartile NC cells. (C) Violin plots of selected genes. (D) Hierarchical clusterings of single trailblazer and quartile NC cells by Euclidean distance or Pearson dissimilarity based upon averages of the 96-gene profiles. n=318 cells total; n=72 HH Stage 13 trailblazers, n=76 HH Stage 15 trailblazers, n=43 1st quartile, n=41 2nd quartile, n=44 3rd quartile and n=42 4th quartile.

Figure 7. Knockdown of leader behavior alters the NC migration pattern. (A) Fluorescein-labeled control morpholino (n=8), HAND2 morpholino (n=13) or TFAP2A morpholino (n=11) (green) embryos and H2B-mCherry (red) at HH St15 (pre-otic NC stream shown). The asterisk marks the migratory front and arrow points to the thinnest portion of the stream. (B-C) Distribution of the percentage of transfected NC cells along the migratory pathway at HH St13,15. (D-G) Box plots of the (D-E) distance migrated and area invaded (F-G) at HH St13,15. Ba, branchial arch.

Figure 8. Upregulation of leader behavior in the trailing cells alters the NC migration pattern. (A) Trailing NC cells transfected with control EGFP (n=20), HAND2 FL (n=13) or TFAP2A FL (n=14) (green) and DiI (red) at HH Stage 15. The asterisks mark the position of the migratory front and the arrows point to the thinnest portion of the stream. (B-C) Distribution of the (B) percentage of transfected NC cells and (C)

average number of NC cells along the migratory pathway. (D-E) Model simulations of the upregulation of a leader-like behavior in the trailing cells. After time, t_LF every other cell inserted into the domain is a leader (orange) rather than a follower (purple). (D) Overlay of 20 repeats of the stochastic simulation for control (WT) and perturbed (lead+) simulations. Greyscale boxplots show distance migrated in the x-direction and their color gives relative stream density (as in Fig. 4). (E) Histogram of cell counts vs. distance for control and perturbed simulations. r, rhombomere; ba, branchial arch.

Figure 9. Trailblazer molecular signature. The molecular signature of trailblazer NC cells is comprised of 16 genes (shaded region). The arrows indicate shortest pathway regulation (direct, indirect, positive or negative) derived from Pathway Studios.

Table 1. Computational model parameters.

Table 2. Highest expressed genes in single trailblazer cells. Bolded genes are highly expressed in at least 50% of NC cells analyzed.

Table 3. Differential expression between trailblazers and quartiles at HH 15. Bolded genes are highly expressed by trailblazers and differentially expressed between trailblazers and the rest of the stream.

Table 4. Molecular profile after HAND2 or TFAP2A knockdown. Bolded genes are either in the trailblazer molecular signature or HAND2 and TFAP2A.

REFERENCES

Adams, M. S., Gammill, L. S. and Bronner-Fraser, M. (2008). Discovery of transcription factors and other candidate regulators of NC development. *Dev Dyn* **237**, 1021-33.

Barrallo-Gimeno, A., Holzschuh, J., Driever, W. and Knapik, E. W. (2004). NC survival and differentiation in zebrafish depends on mont blanc/tfap2a gene function. *Development* **131**, 1463-77.

Berg, H. C. and Purcell, E. M. (1977). Physics of chemoreception. *Biophys J* 20, 193-219.

Brewer, S., Jiang, X., Donaldson, S., Williams, T. and Sucov, H. M. (2002). Requirement for AP-2alpha in cardiac outflow tract morphogenesis. *Mech Dev* **110**, 139-49.

Chazaud, C., Oulad-Abdelghani, M., Bouillet, P., Decimo, D., Chambon, P. and Dolle, P. (1996). AP-2.2, a novel gene related to AP-2, is expressed in the forebrain, limbs and face during mouse embryogenesis. *Mech Dev* **54**, 83-94.

Choi, H. M., Beck, V. A. and Pierce, N. A. (2014). Next-generation in situ hybridization chain reaction: higher gain, lower cost, greater durability. *ACS Nano* **8**, 4284-94.

Choi, H. M., Chang, J. Y., Trinh le, A., Padilla, J. E., Fraser, S. E. and Pierce, N. A. (2010). Programmable in situ amplification for multiplexed imaging of mRNA expression. *Nat Biotechnol* **28**, 1208-12.

D'Autreaux, F., Morikawa, Y., Cserjesi, P. and Gershon, M. D. (2007). Hand2 is necessary for terminal differentiation of enteric neurons from crest-derived precursors but not for their migration into the gut or for formation of glia. *Development* **134**, 2237-49.

Gallardo, V. E., Liang, J., Behra, M., Elkahloun, A., Villablanca, E. J., Russo, V., Allende, M. L. and Burgess, S. M. (2010). Molecular dissection of the migrating posterior lateral line primordium during early development in zebrafish. *BMC Dev Biol* **10**, 120.

Gammill, L. S. and Bronner-Fraser, M. (2002). Genomic analysis of NC induction. *Development* **129**, 5731-41.

Gammill, L. S. and Bronner-Fraser, M. (2003). NC specification: migrating into genomics. *Nat Rev Neurosci* **4**, 795-805.

Hamburger, V. and Hamilton, H. L. (1951). A series of normal stages in the development of the chick embryo. *Journal of Morphology* **88**, 49-92.

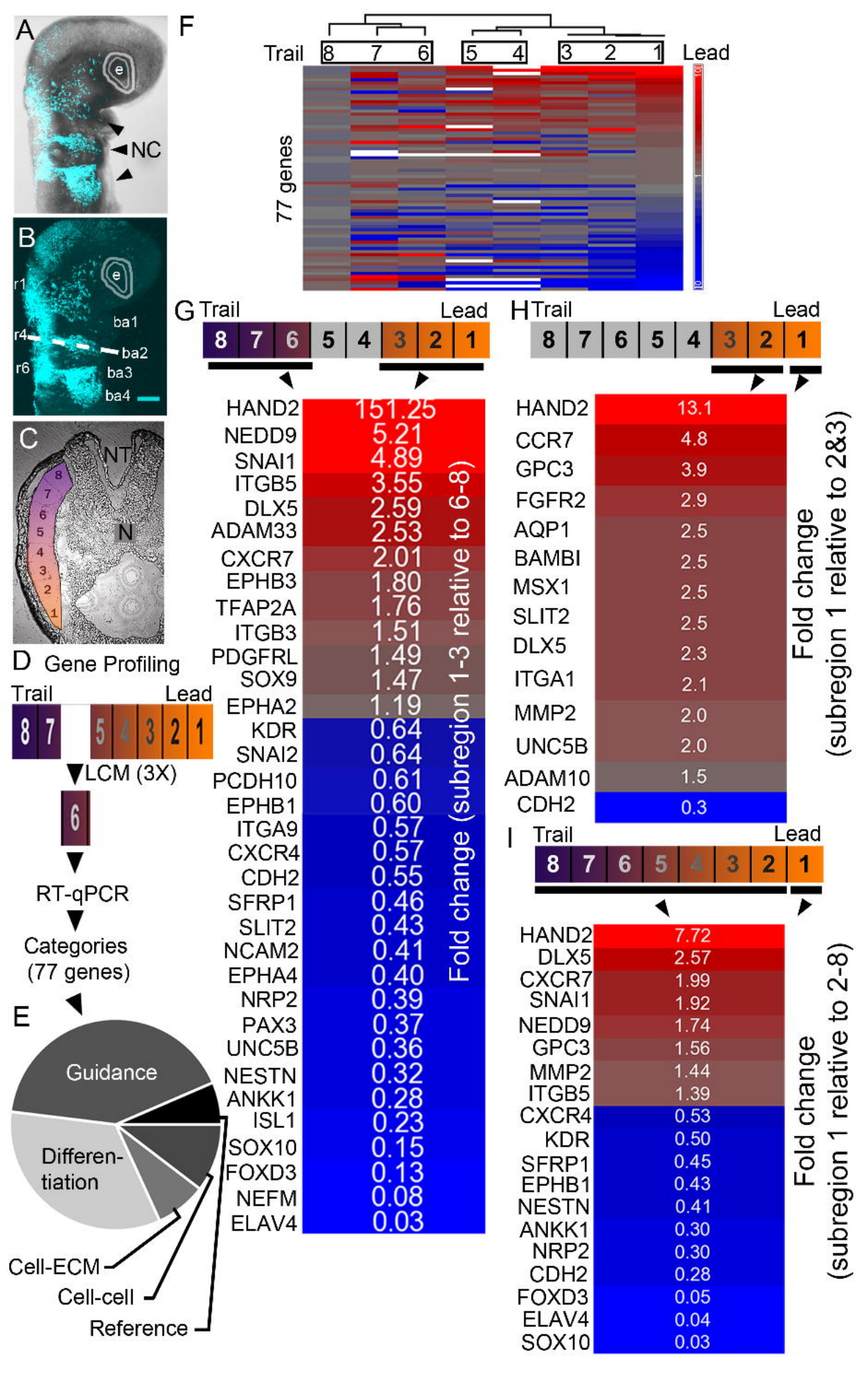
Hendershot, T. J., Liu, H., Sarkar, A. A., Giovannucci, D. R., Clouthier, D. E., Abe, M. and Howard, M. J. (2007). Expression of Hand2 is sufficient for neurogenesis and cell type-specific gene expression in the enteric nervous system. *Dev Dyn* **236**, 93-105.

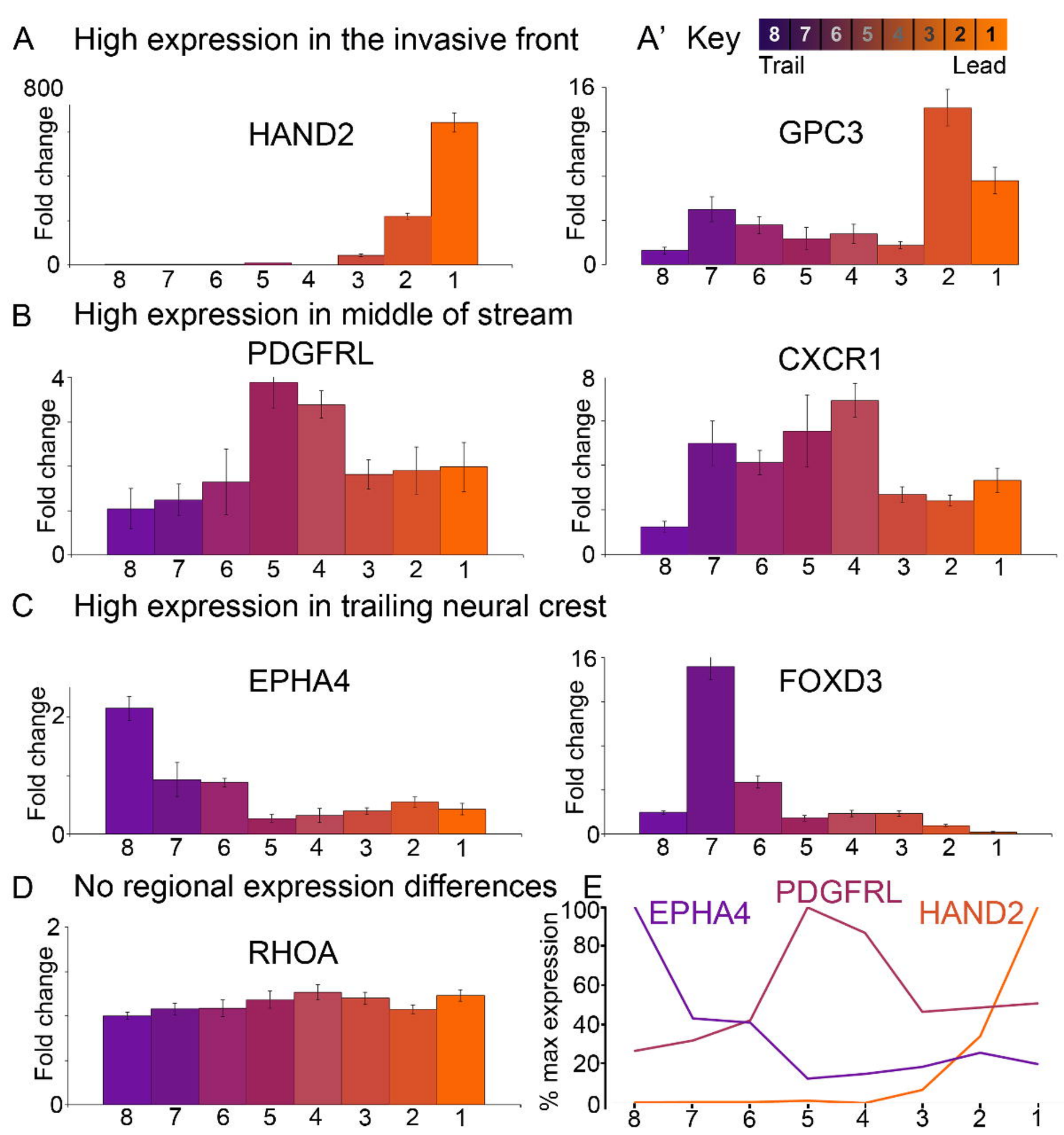
Howard, M., Foster, D. N. and Cserjesi, P. (1999). Expression of HAND gene products may be sufficient for the differentiation of avian neural crest-derived cells into catecholaminergic neurons in culture. *Dev Biol* **215**, 62-77.

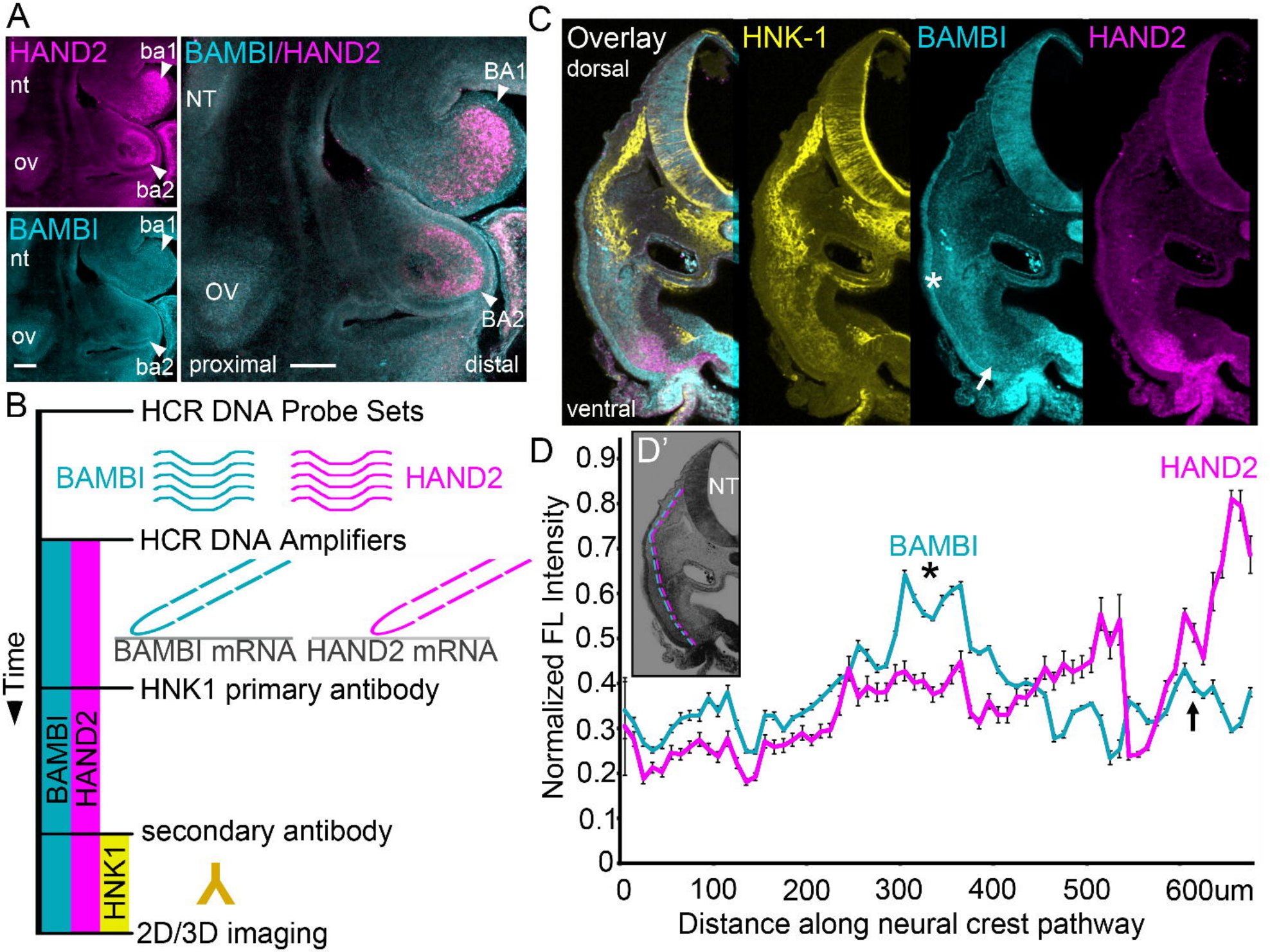
Knight, R. D., Javidan, Y., Nelson, S., Zhang, T. and Schilling, T. (2004). Skeletal and pigment cell defects in the lockjaw mutant reveal multiple roles for zebrafish tfap2a in NC development. *Dev Dyn* **229**, 87-98.

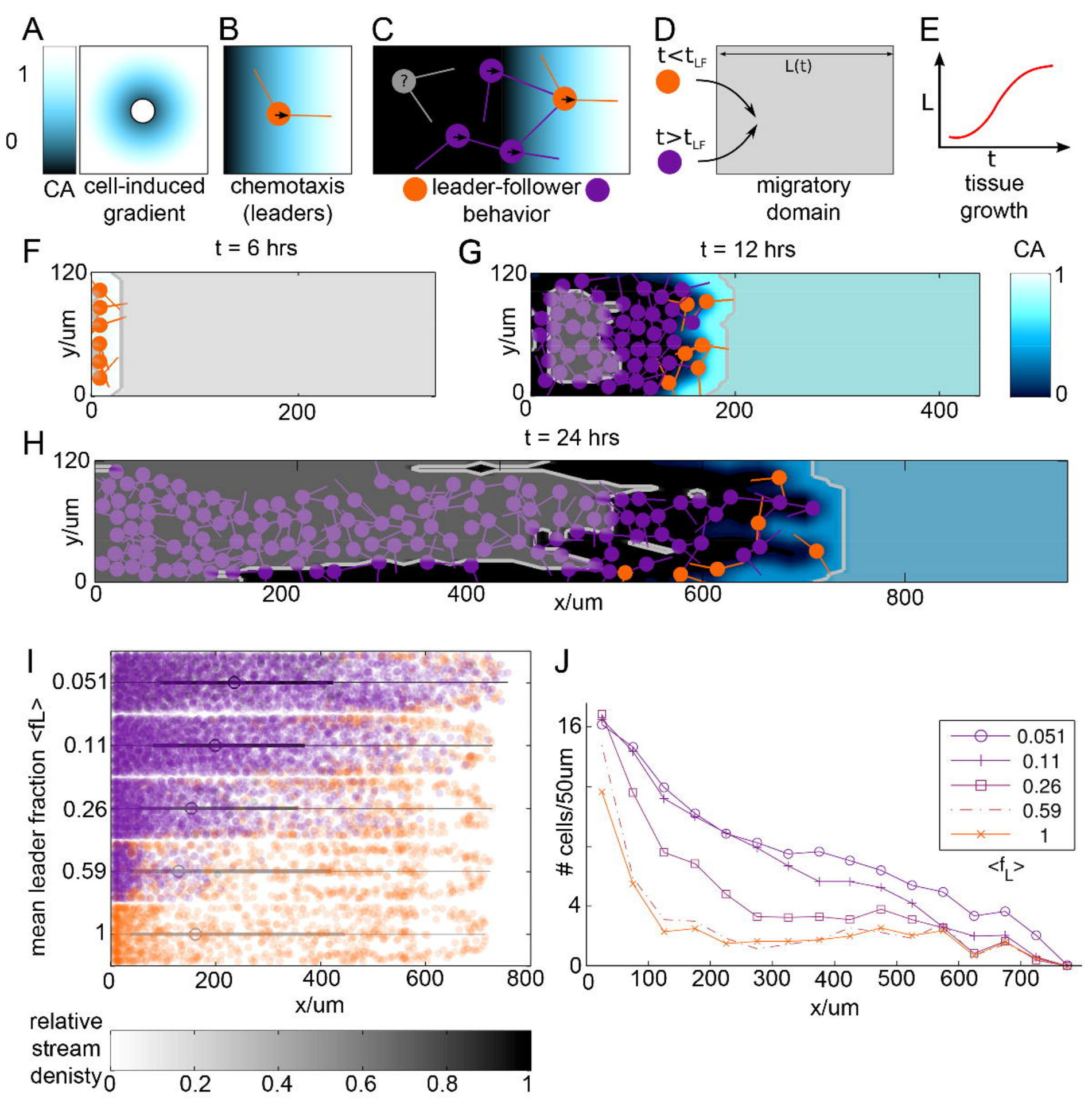
Knight, R. D., Nair, S., Nelson, S. S., Afshar, A., Javidan, Y., Geisler, R., Rauch, G. J. and Schilling, T. F. (2003). lockjaw encodes a zebrafish tfap2a required for early NC development. *Development* **130**, 5755-68.
Kulesa, P. M. and Fraser, S. E. (1998). NC cell dynamics revealed by time-lapse video microscopy of whole embryo chick explant cultures. *Developmental Biology* **204**, 327-344.
Kulesa, P. M. and Fraser, S. E. (2000). In ovo time-lapse analysis of chick hindbrain NC cell migration shows cell interactions during migration to the branchial arches. *Development* **127**, 1161-72.
Kulesa, P. M. and Gammill, L. S. (2010). NC migration: patterns, phases and signals. *Dev Biol* **344**, 566-8.
Kulesa, P. M., Teddy, J. M., Stark, D. A., Smith, S. E. and McLennan, R. (2008). NC invasion is a spatially-ordered progression into the head with higher cell proliferation at the migratory front as revealed by the photoactivatable protein, KikGR. *Dev Biol* **316**, 275-87.
Li, W. and Cornell, R. A. (2007). Redundant activities of Tfap2a and Tfap2c are required for NC induction and development of other non-neural ectoderm derivatives in zebrafish embryos. *Dev Biol* **304**, 338-54.
Luo, T., Lee, Y. H., Saint-Jeannet, J. P. and Sargent, T. D. (2003). Induction of NC in Xenopus by transcription factor AP2alpha. *Proc Natl Acad Sci U S A* **100**, 532-7.
McKinney, M. C. and Kulesa, P. M. (2011). In vivo calcium dynamics during NC cell migration and patterning using GCaMP3. *Dev Biol* **358**, 309-17.
McLennan, R., Dyson, L., Prather, K. W., Morrison, J. A., Baker, R. E., Maini, P. K. and Kulesa, P. M. (2012). Multiscale mechanisms of cell migration during development: theory and experiment. *Development* **139**, 2935-44.
McLennan, R. and Kulesa, P. M. (2007). In vivo analysis reveals a critical role for neuropilin-1 in cranial NC cell migration in chick. *Dev Biol* **301**, 227-39.
McLennan, R., Teddy, J. M., Kasemeier-Kulesa, J. C., Romine, M. H. and Kulesa, P. M. (2010). Vascular endothelial growth factor (VEGF) regulates cranial NC migration in vivo. *Dev Biol* **339**, 114-25.
Miller, C. T., Yelon, D., Stainier, D. Y. and Kimmel, C. B. (2003). Two endothelin 1 effectors, hand2 and bapx1, pattern ventral pharyngeal cartilage and the jaw joint. *Development* **130**, 1353-65.
Mitchell, P. J., Timmons, P. M., Hebert, J. M., Rigby, P. W. and Tjian, R. (1991). Transcription factor AP-2 is expressed in NC cell lineages during mouse embryogenesis. *Genes Dev* **5**, 105-19.
Molyneaux, K. A., Wang, Y., Schaible, K. and Wylie, C. (2004). Transcriptional profiling identifies genes differentially expressed during and after migration in murine primordial germ cells. *Gene Expr Patterns* **4**, 167-81.
Morrison, J. A., Bailey, C. M. and Kulesa, P. M. (2012). Gene profiling in the avian embryo using laser capture microdissection and RT-qPCR. *Cold Spring Harb Protoc* 2012.
Nishiyama, C., Uesaka, T., Manabe, T., Yonekura, Y., Nagasawa, T., Newgreen, D. F., Young, H. M. and Enomoto, H. (2012). Trans-mesenteric NC cells are the principal source of the colonic enteric nervous system. *Nat Neurosci* **15**, 1211-8.

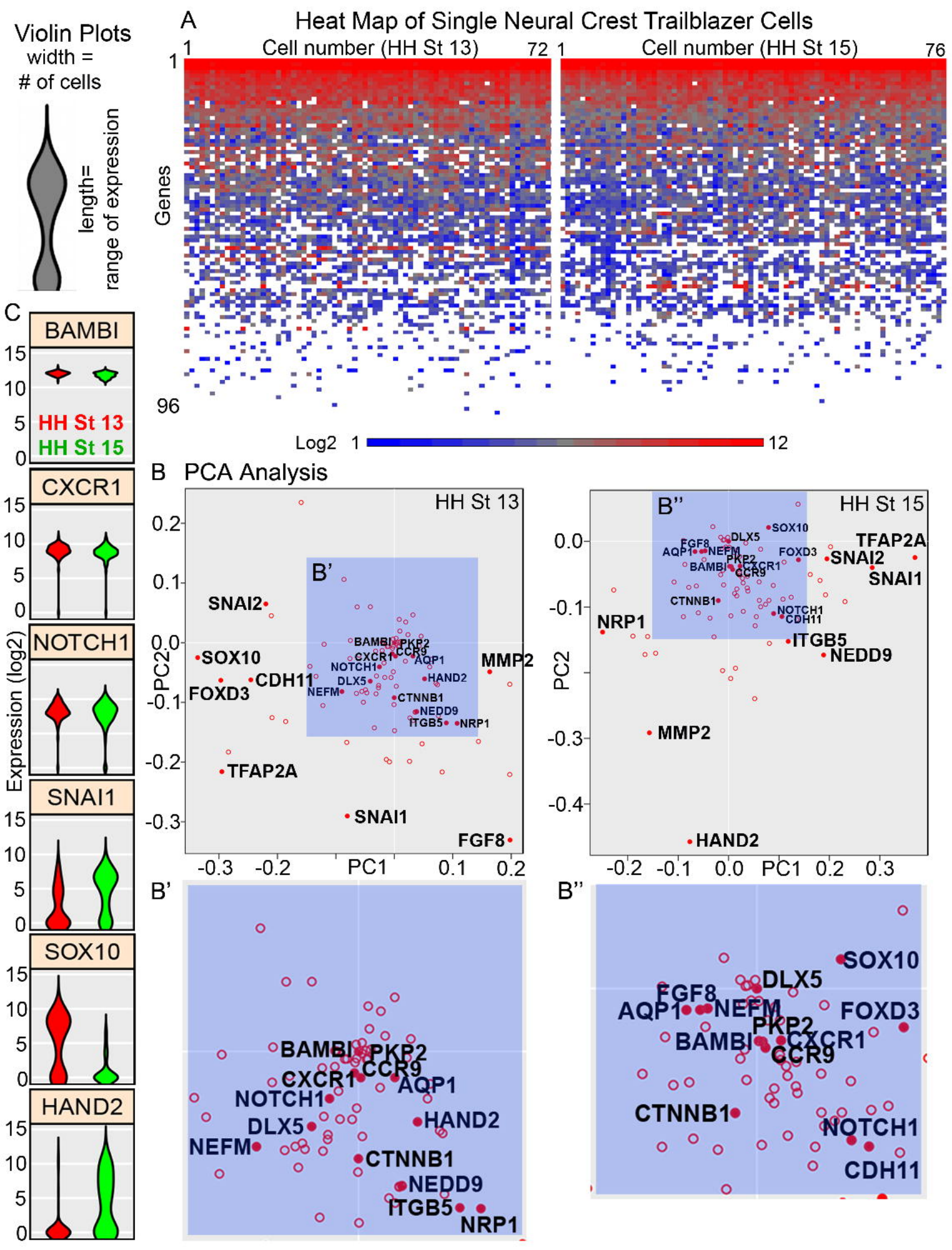
Pfisterer, P., Ehlermann, J., Hegen, M. and Schorle, H. (2002). A subtractive gene expression screen suggests a role of transcription factor AP-2 alpha in control of proliferation and differentiation. *J Biol Chem* **277**, 6637-44.
Piotrowski, T. and Baker, C. V. (2014). The development of lateral line placodes: taking a broader view. *Dev Biol* **389**, 68-81.
Ridenour, D. A., McLennan, R., Teddy, J. M., Semerad, C. L., Haug, J. S. and Kulesa, P. M. (2014). The NC cell cycle is related to phases of migration in the head. *Development* **141**, 1095-103.
Schorle, H., Meier, P., Buchert, M., Jaenisch, R. and Mitchell, P. J. (1996). Transcription factor AP-2 essential for cranial closure and craniofacial development. *Nature* **381**, 235-8.
Shen, H., Wilke, T., Ashique, A. M., Narvey, M., Zerucha, T., Savino, E., Williams, T. and Richman, J. M. (1997). Chicken transcription factor AP-2: cloning, expression and its role in outgrowth of facial prominences and limb buds. *Dev Biol* **188**, 248-66.
Simoes-Costa, M. and Bronner, M. E. (2013). Insights into NC development and evolution from genomic analysis. *Genome Res* **23**, 1069-80.
Simoes-Costa, M., Tan-Cabugao, J., Antoshechkin, I., Sauka-Spengler, T. and Bronner, M. E. (2014). Transcriptome analysis reveals novel players in the cranial NC gene regulatory network. *Genome Res* **24**, 281-90.
Srivastava, D., Thomas, T., Lin, Q., Kirby, M. L., Brown, D. and Olson, E. N. (1997). Regulation of cardiac mesodermal and NC development by the bHLH transcription factor, dHAND. *Nat Genet* **16**, 154-60.
Tarbashevich, K. and Raz, E. (2010). The nuts and bolts of germ-cell migration. *Curr Opin Cell Biol* **22**, 715-21.
Teddy, J. M. and Kulesa, P. M. (2004). In vivo evidence for short- and long-range cell communication in cranial NC cells. *Development* **131**, 6141-51.
Thomas, T., Kurihara, H., Yamagishi, H., Kurihara, Y., Yazaki, Y., Olson, E. N. and Srivastava, D. (1998). A signaling cascade involving endothelin-1, dHAND and msx1 regulates development of neural-crest-derived branchial arch mesenchyme. *Development* **125**, 3005-14.
Voiculescu, O., Bodenstein, L., Lau, I. J. and Stern, C. D. (2014). Local cell interactions and self-amplifying individual cell ingression drive amniote gastrulation. *Elife* **3**, e01817.
Wang, W. D., Melville, D. B., Montero-Balaguer, M., Hatzopoulos, A. K. and Knapik, E. W. (2011). Tfap2a and Foxd3 regulate early steps in the development of the NC progenitor population. *Dev Biol* **360**, 173-85.
Young, H. M., Bergner, A. J., Anderson, R. B., Enomoto, H., Milbrandt, J., Newgreen, D. F. and Whitington, P. M. (2004). Dynamics of neural crest-derived cell migration in the embryonic mouse gut. *Dev Biol* **270**, 455-73.
Young, H. M., Bergner, A. J., Simpson, M. J., McKeown, S. J., Hao, M. M., Anderson, C. R. and Enomoto, H. (2014). Colonizing while migrating: how do individual enteric NC cells behave? *BMC Biol* **12**, 23.

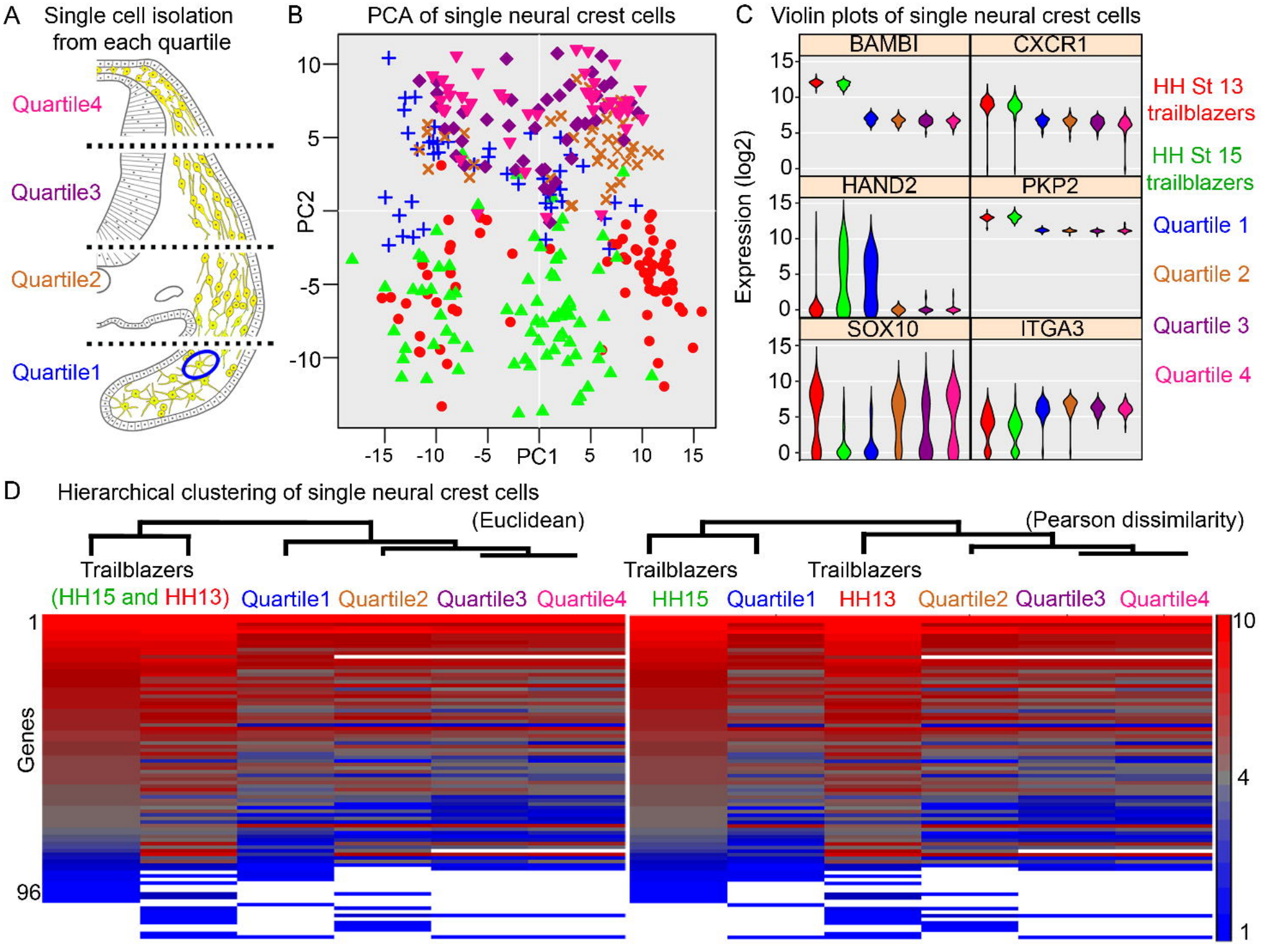

A
e
NC
B
e
r1
r4
r6
ba1
ba2
ba3
ba4
C
NT
N
8
7
6
5
4
3
2
1
D Gene Profiling
Trail
Lead
8 7 5 4 3 2 1
LCM (3X)
6
RT-qPCR
Categories
(77 genes)
E
Guidance
Differen-
tiation
Cell-ECM
Cell-cell
Reference
F
Trail
8 7 6 5 4 3 2 1
Lead
77 genes
G Trail
Lead
8 7 6 5 4 3 2 1
HAND2 151.25
NEDD9 5.21
SNAI1 4.89
ITGB5 3.55
DLX5 2.59
ADAM33 2.53
CXCR7 2.01
EPHB3 1.80
TFAP2A 1.76
ITGB3 1.51
PDGFRL 1.49
SOX9 1.47
EPHA2 1.19
KDR 0.64
SNAI2 0.64
PCDH10 0.61
EPHB1 0.60
ITGA9 0.57
CXCR4 0.57
CDH2 0.55
SFRP1 0.46
SLIT2 0.43
NCAM2 0.41
EPHA4 0.40
NRP2 0.39
PAX3 0.37
UNC5B 0.36
NESTN 0.32
ANKK1 0.28
ISL1 0.23
SOX10 0.15
FOXD3 0.13
NEFM 0.08
ELAV4 0.03
Fold change (subregion 1-3 relative to 6-8)
H Trail
Lead
8 7 6 5 4 3 2 1
HAND2 13.1
CCR7 4.8
GPC3 3.9
FGFR2 2.9
AQP1 2.5
BAMBI 2.5
MSX1 2.5
SLIT2 2.5
DLX5 2.3
ITGA1 2.1
MMP2 2.0
UNC5B 2.0
ADAM10 1.5
CDH2 0.3
Fold change
(subregion 1 relative to 2&3)
I Trail
Lead
8 7 6 5 4 3 2 1
HAND2 7.72
DLX5 2.57
CXCR7 1.99
SNAI1 1.92
NEDD9 1.74
GPC3 1.56
MMP2 1.44
ITGB5 1.39
CXCR4 0.53
KDR 0.50
SFRP1 0.45
EPHB1 0.43
NESTN 0.41
ANKK1 0.30
NRP2 0.30
CDH2 0.28
FOXD3 0.05
ELAV4 0.04
SOX10 0.03
Fold change
(subregion 1 relative to 2-8)

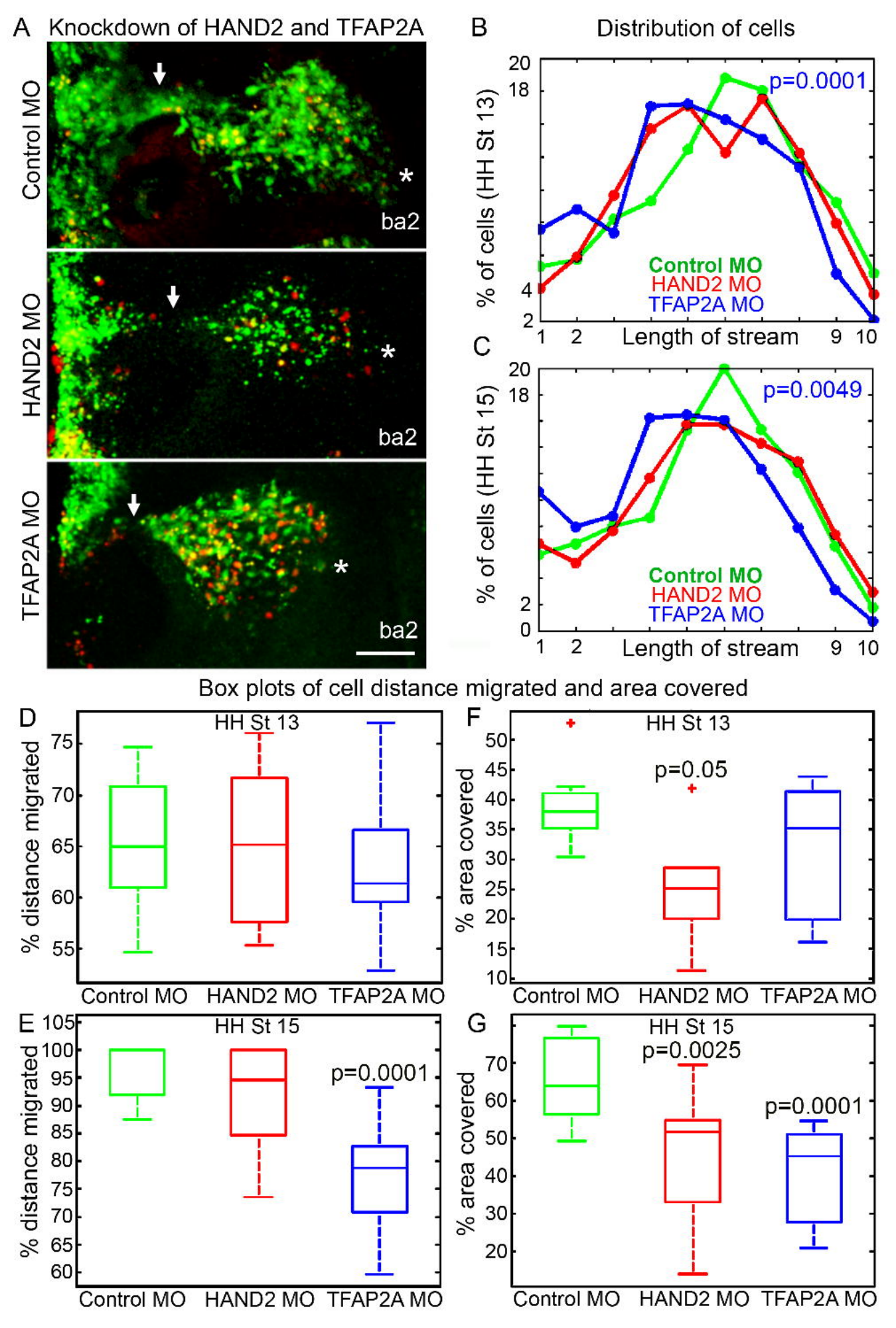

A High expression in the invasive front
A' Key
8 7 6 5 4 3 2 1
Trail
Lead
HAND2
GPC3
Fold change
800
16
0
B High expression in middle of stream
PDGFRL
CXCR1
4
8
C High expression in trailing neural crest
EPHA4
FOXD3
2
D No regional expression differences
RHOA
E
% max expression
100
80
60
40
20

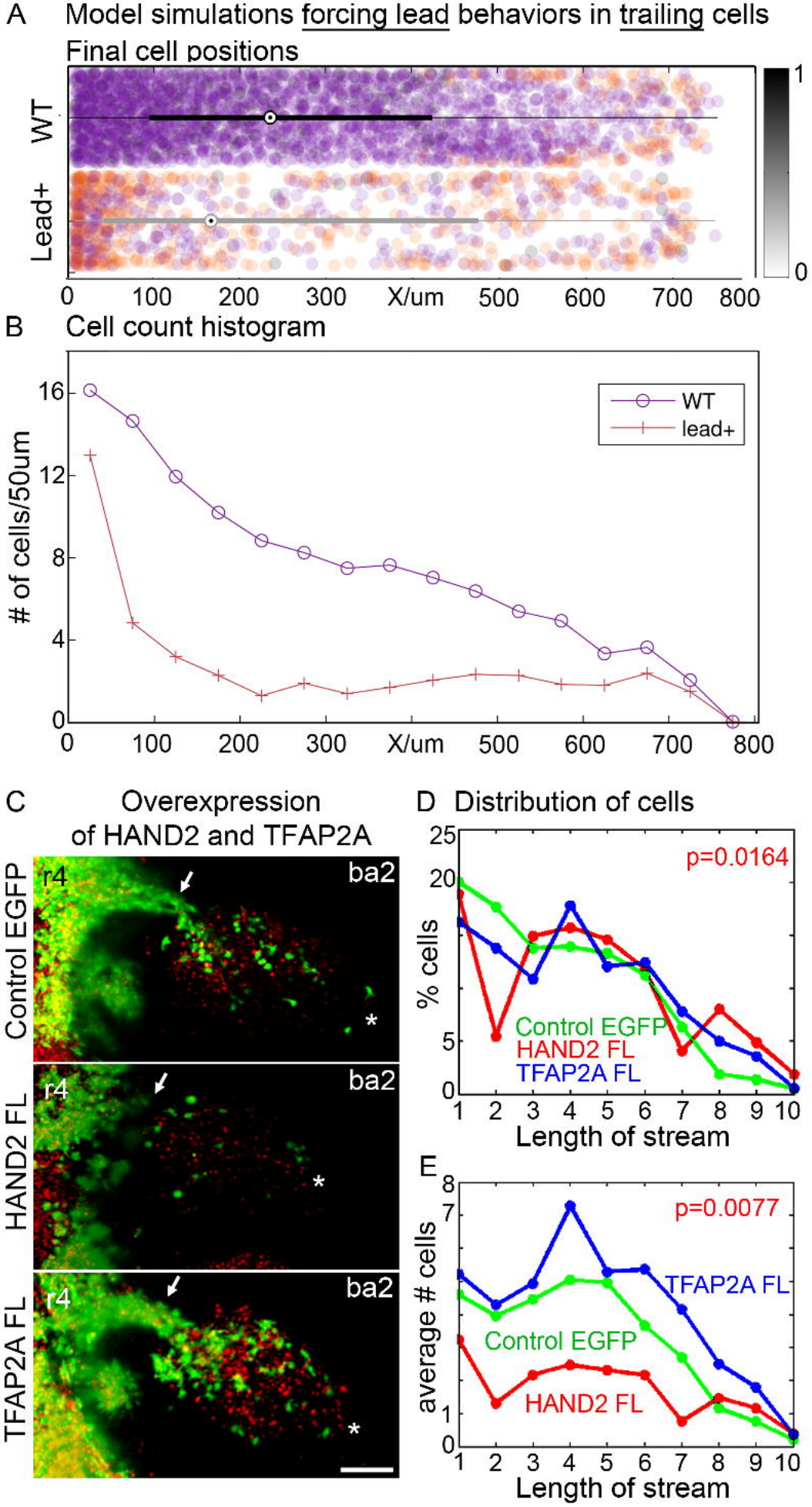

A
HAND2
ba1
nt
ov
ba2
BAMBI
ba1
nt
ov
ba2
BAMBI/HAND2
BA1
NT
OV
BA2
proximal
distal
B
HCR DNA Probe Sets
BAMBI
HAND2
HCR DNA Amplifiers
BAMBI mRNA
HAND2 mRNA
HNK1 primary antibody
Time
BAMBI
HAND2
HNK1
secondary antibody
2D/3D imaging
C
Overlay
dorsal
ventral
HNK-1
BAMBI
HAND2
*
D
D'
NT
HAND2
BAMBI
*
Normalized FL Intensity
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0
100
200
300
400
500
600um
Distance along neural crest pathway


A
1
0
CA
cell-induced gradient
B
chemotaxis (leaders)
C
?
leader-follower behavior
D
t<t_LF
t>t_LF
L(t)
migratory domain
E
L
t
tissue growth
F
t = 6 hrs
120
y/um
0
0
200
G
t = 12 hrs
120
y/um
0
0
200
400
CA
1
0
H
t = 24 hrs
120
y/um
0
0
200
400
x/um
600
800
I
mean leader fraction <fL>
0.051
0.11
0.26
0.59
1
0
200
400
600
800
x/um
relative stream denisty
0
0.2
0.4
0.6
0.8
1
J
cells/50um
16
4
0
0
100
200
300
400
500
600
700
x/um
0.051
0.11
0.26
0.59
1
<f_L>
f_L


Violin Plots
width = # of cells
length= range of expression
A
Heat Map of Single Neural Crest Trailblazer Cells
Cell number (HH St 13)
Cell number (HH St 15)
Genes
Log2
B PCA Analysis
HH St 13
HH St 15
B'
B''
PC1
PC2
C
BAMBI
CXCR1
NOTCH1
SNAI1
SOX10
HAND2
HH St 13
HH St 15
Expression (log2)


A
Single cell isolation from each quartile
Quartile4
Quartile3
Quartile2
Quartile1
B
PCA of single neural crest cells
PC2
PC1
10
5
-5
-10
-15
-10
-5
5
10
15
C
Violin plots of single neural crest cells
BAMBI
CXCR1
HAND2
PKP2
SOX10
ITGA3
Expression (log2)
15
10
5
0
HH St 13 trailblazers
HH St 15 trailblazers
Quartile 1
Quartile 2
Quartile 3
Quartile 4
D
Hierarchical clustering of single neural crest cells
(Euclidean)
(Pearson dissimilarity)
Trailblazers
(HH15 and HH13)
Quartile1
Quartile2
Quartile3
Quartile4
Trailblazers
HH15
Quartile1
Trailblazers
HH13
Quartile2
Quartile3
Quartile4
Genes
1
96
10
4
1


A
Knockdown of HAND2 and TFAP2A
Control MO
HAND2 MO
TFAP2A MO
ba2
B
Distribution of cells
% of cells (HH St 13)
p=0.0001
Control MO
HAND2 MO
TFAP2A MO
Length of stream
C
% of cells (HH St 15)
p=0.0049
Box plots of cell distance migrated and area covered
D
HH St 13
% distance migrated
Control MO
HAND2 MO
TFAP2A MO
F
HH St 13
p=0.05
% area covered
E
HH St 15
p=0.0001
G
HH St 15
p=0.0025
p=0.0001


A
Model simulations forcing lead behaviors in trailing cells
Final cell positions
WT
Lead+
1
0
0
100
200
300
X/um
500
600
700
800
B
Cell count histogram
# of cells/50um
16
12
8
4
0
WT
lead+
0
100
200
300
X/um
500
600
700
800
C
Overexpression
of HAND2 and TFAP2A
Control EGFP
HAND2 FL
TFAP2A FL
r4
ba2
D
Distribution of cells
p=0.0164
% cells
25
20
5
0
Control EGFP
HAND2 FL
TFAP2A FL
1
2
3
4
5
6
7
8
9
10
Length of stream
E
p=0.0077
average # cells
8
7
1
0
TFAP2A FL
Control EGFP
HAND2 FL
1
2
3
4
5
6
7
8
9
10
Length of stream

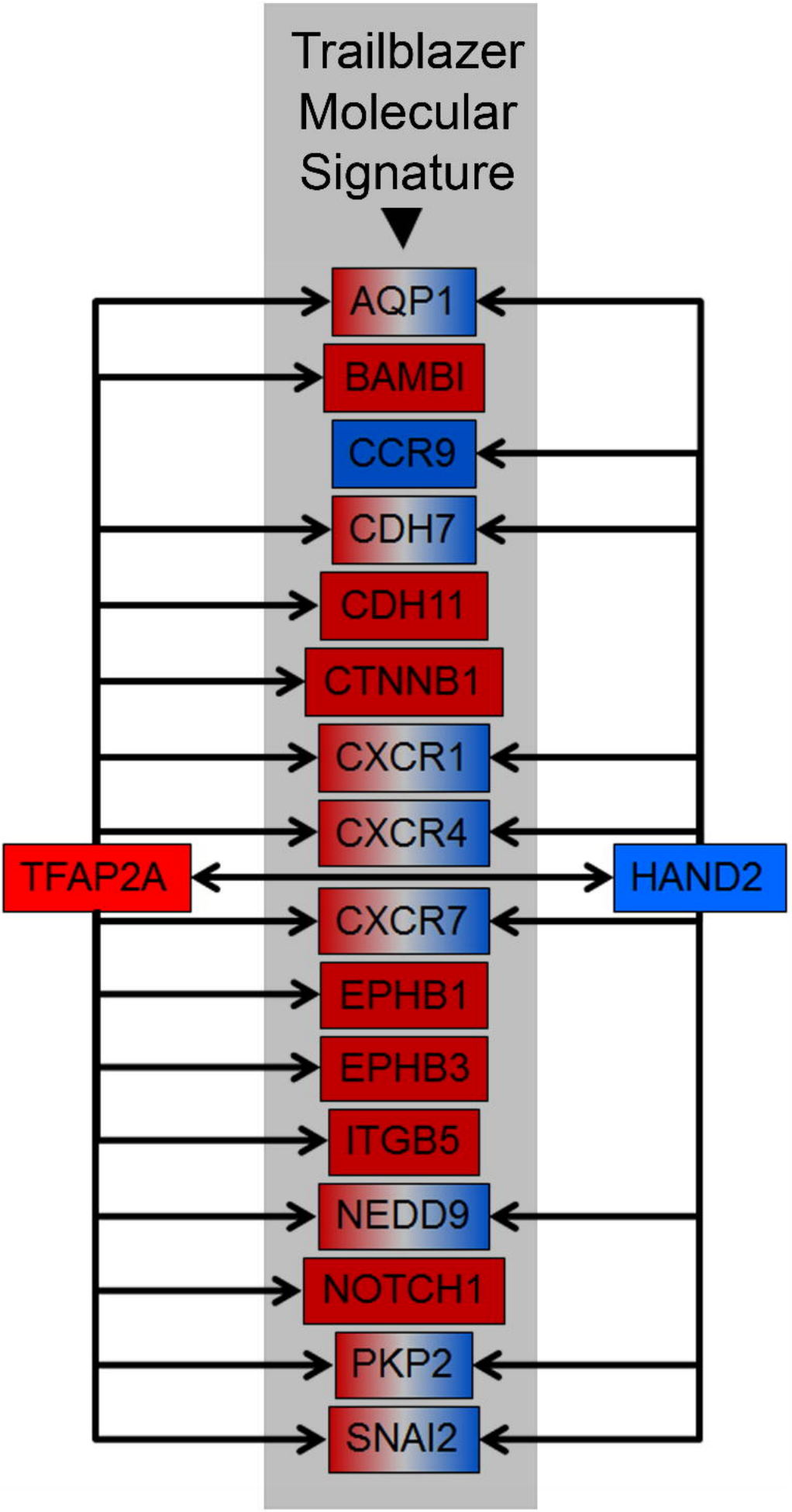

Trailblazer
Molecular
Signature
AQP1
BAMBI
CCR9
CDH7
CDH11
CTNNB1
CXCR1
CXCR4
TFAP2A
HAND2
CXCR7
EPHB1
EPHB3
ITGB5
NEDD9
NOTCH1
PKP2
SNAI2

Table 1: **Model parameters**

Parameter values listed were used as a default, unless otherwise stated. Where a range is given, the model gives qualitatively similar results within that range, and the italicised value is the one used for the figures in this paper.

	Description	Value	Reference
t_{LF}	time after which newly inserted cells are followers	variable	see results section
n_{filo}	directions sampled per timestep	2	n/a, see notes
Δt	simulation time step	1 min	n/a
R	cell radius (nuclear)	7.5μm	McLennan & Kulesa (2010)
v_{lead}	cell speed (leader cells)	41.6μm/h	Kulesa *et al.* (2008)
v_{follow}	cell speed (follower cells)	49.9μm/h	Kulesa *et al.* (2008)
L_y	height of migratory domain	120μm	McLennan *et al.* (2012)
L_x	length of migratory domain (grows logistically, Eq. (1))	300μm to 1100μm	McLennan *et al.* (2012)
l_{filo}	sensing radius	27.5μm	see notes
$l_{\mathrm{filo}}^{\mathrm{max}}$	maximum cell separation before contact is lost	45μm	see notes
$\Delta c/c$	sensing accuracy	0.001 to *0.1*	Section S1.2
D	diffusion coefficient of chemoattractant (effective)	*0.1* to $10^5\mu\mathrm{m}^2/\mathrm{h}$	see notes
χ	production rate of chemoattractant	*0.0001* to 1/h	see notes
λ	chemoattractant internalisation rate	100 to *1000*/h	see notes
k_{in}	rate at which cells enter the domain	10/h	see notes

Notes

Experimental time: Cell migration starts approximately six hours after electroporation ($t = 0$).

Directions sampled per timestep, n_{filo}: This cannot be directly related to the number of filopodia, which are greater in number, but sample at a lower speed (McLennan *et al.* 2012).

Diffusion coefficient of chemoattractant, D: The primary identified chemoattractant in chick cranial neural crest migration is VEGF^{165} (McLennan *et al.* 2010). Its related isoform VEGF^{164} is known to bind to extracellular matrix (ECM), and studies in angiogenesis estimate as little as 1% may be freely diffusing, the rest bound to ECM and cellular receptors (Mac Gabhann *et al.* 2006). Hence, we choose a low effective diffusivity.

Production rate of chemottractant, χ: In other tissues, VEGF production, or estimates thereof, range from 0.01-0.20 molecules/cell/s (Yen *et al.* 2011), 4.39-5.27$\cdot 10^{-5}$molecules/μm$^-$2/s (Vempati *et al.* 2011) to $0.25 \cdot 10^{-17}$pmol/μm^2/s(Mac Gabhann *et al.* 2006). In our system, the rate of VEGF production is unknown and difficult to measure. Howeever, it is outweighed by internalisation through migrating neural crest cells, as VEGF is not seen to be replenished in trailing portions of the stream (McLennan *et al.* 2010). Thus, we assume χ to be low.

Chemoattractant internalisation rate, λ: To our knowledge, no estimates or measurements of VEGF internalisation rate of chick cranial neural crest exists. Angiogenesis studies have used values of $k_{\mathrm{VEGFR2}} = O(10^{-4})$/s (Mac Gabhann & Popel 2005, Yen *et al.* 2011). Berg & Purcell (1977) estimate the number of receptors needed for a near-optimal sensing accuracy as $N_R = R/s$, where R is the cell radius and s the receptor size. With $s = O(\mathrm{nm})$, we can estimate the number of receptors to be $N_R \geq 10^4$. If receptor internalisation rates are comparable to other tissues, a lower bound on λ would be given by $k_{\mathrm{VEGFR2}} N_R > 1$. However, the concentration of VEGF in our system is unknown, and hence the units of c, and therefore λ, in our model are arbitrary. We assume a high λ to ensure quick consumption of chemoattractant by cells.

Rate at which cells enter the domain, k_{in}: This is the rate of attempted cell insertions, in a typical simulation on the order of 10% of insertions are unsuccessful. It should be noted here again that our simulations are a two-dimensional abstraction of the three-dimensional migratory stream, which may contain 4-5 times as many cells *in vivo* in the transverse (z) direction.

Sensing radius, l_{filo}: This was calculated as the sum of the cell radius (7.5μm) and the mean filopodial length (which was directly measured from the cell body (Fig. S4B) to be 9μm and estimated from total cell size (Fig. S4A) to be circa 20μm). Since we have only implemented contact between filopodium and cell body, but not between two filopodia, which does occur *in vivo* (Teddy & Kulesa 2004), we allow for a greater effective length.

Maximum cell separation before contact is lost, $l_{\mathrm{filo}}^{\mathrm{max}}$: The maximum cell size including filopodia was measured to be 86.3μm (Fig. S4A), half of which gives an estimate of maximum cell separation of 43.15μm. Independent measurements of filopodial lengths gave a maximum of 30.4μm (from the cell body, Fig. S4B), which, together with the cell radius $R = 7.5\mu$m and the average filopodial length (allowing for interfilopodial contact) of 9μm gives an estimate of 46.5μm.

Gene	AVE Ct of HH13 Trailblazers expressing	% of HH13 Trailblazers expressing
PKP2	13.00973678	100.00
BAMBI	13.99207538	100.00
CXCR1	16.923582	98.61
CXCR4	17.46313907	100.00
CTNNB1	17.51107514	100.00
CCR9	17.70692509	100.00
NOTCH1	17.73621817	98.61
CDH11	17.78358413	87.50
HAND2	17.88303114	6.94
FGF8	18.05853726	19.44
SOX10	18.42511568	68.06
FGFR1	18.48812662	94.44
SLIT1	18.50881211	88.89
TFAP2A	18.55281934	70.83
GPC3	19.12203206	100.00
ITGB3	19.15008624	73.61
ROBO1	19.20342733	100.00
ADAM10	19.27265348	98.61
EPHA4	19.31785986	86.11
FOXD3	19.37819436	62.50
SNAI2	19.48469388	75.00
CDH7	19.5414792	77.78
EPHB1	19.5593769	100.00
FZD7	19.59303938	98.61
ADAM33	19.77794749	79.17
AQP1	20.11626496	88.89
SFRP1	20.22819026	87.50
JAG1	20.34777223	69.44
CDH2	20.38084649	98.61
NRP2	20.41404755	90.28
KRT19	20.48268839	59.72
MMP2	20.56005797	55.56
NESTN	20.59934615	93.06
CXCL12	20.68492789	84.72
PAX3	20.700437	69.44
ITGB5	20.80221854	83.33
ERBB4	20.81292209	2.78
CXCR7	20.85370877	76.39
SNAI1	20.85531629	44.44
EPHA2	20.86312155	80.56
VEGFA	20.90223852	55.56
FGF4	20.99185856	8.33
TH	21.00074756	2.78
SPON1	21.00781815	76.39
CFC1B	21.10568255	95.83
PDGFRL	21.12772244	25.00
NRP1	21.17234955	38.89
EPHB3	21.17772359	93.06
NEDD9	21.18236349	54.17
SLIT2	21.22252072	52.78
BMPR2	21.29450543	87.50
PHOX2B	21.30238098	1.39
ITGA9	21.31944168	95.83
CCR7	21.33280342	47.22
PCDH10	21.35060574	12.50
ISL1	21.3965638	45.83
ITGA3	21.4559391	81.94
FGFR3	21.51790159	68.06
RUNX2	21.57578757	9.72
MMP9	21.57810155	9.72
ROBO2	21.62961883	79.17
ANGPT2	21.68445917	66.67
KDR	21.71726391	34.72
NEFM	21.76636245	63.89
EPHA3	21.82859112	20.83
CDH6	21.88794688	25.00
WISP1	21.91030979	1.39
EDNRA	21.92058669	76.39
ELAV4	21.94223617	2.78
VCAM1	21.97214993	1.39
NCAM2	22.25038961	6.94

Gene	AVE Ct of HH15 Trailblazers expressing	% of HH15 Trailblazers expressing
PKP2	12.93527783	100.00
BAMBI	14.22838368	100.00
CXCR1	17.20444133	98.68
CTNNB1	17.80330424	100.00
NOTCH1	17.88460284	96.05
CCR9	18.03276478	100.00
CXCR4	18.17937181	100.00
CDH11	18.43253169	96.05
FGF8	18.59746247	3.95
HAND2	18.8133826	59.21
FGFR1	18.84915588	98.68
TFAP2A	19.11212435	71.05
GPC3	19.29216546	100.00
ITGB5	19.62549784	93.42
ROBO1	19.66322555	100.00
CXCL12	19.80433937	94.74
SLIT1	19.80976563	85.53
MMP2	19.84402343	78.95
ADAM10	19.91648057	100.00
SNAI1	19.93185646	75.00
AQP1	20.04369513	82.89
SNAI2	20.18018349	81.58
JAG1	20.18576431	93.42
EPHB1	20.32304678	94.74
PCDH10	20.33257862	2.63
NEDD9	20.33689848	63.16
CDH7	20.46708801	67.11
ADAM33	20.47524681	80.26
ITGB3	20.49427049	88.16
FZD7	20.54024663	97.37
NRP1	20.62063313	57.89
NRP2	20.70717286	81.58
EPHA4	20.73164948	69.74
KRT19	20.79137561	47.37
CXCR7	20.80929894	59.21
EPHA2	20.82868704	85.53
KDR	20.93523081	43.42
CDH2	21.03293914	94.74
PDGFRL	21.12663188	53.95
VEGFA	21.18781963	48.68
ISL1	21.20050602	32.89
EDNRA	21.2195303	82.89
SPON1	21.318706	68.42
EPHB3	21.35982518	81.58
NESTN	21.3977451	84.21
CFC1B	21.41857232	93.42
NEFM	21.43616253	47.37
CCR7	21.44227401	65.79
SFRP1	21.49885016	71.05
BMPR2	21.50252218	81.58
SLIT2	21.55061038	47.37
UNC5B	21.56368985	80.26
ITGA9	21.56943449	89.47
VCAM1	21.66058977	9.21
MMP9	21.66456188	11.84
CDH6	21.77948169	21.05
SOX10	21.80649016	22.37
RUNX2	21.84936929	14.47
EPHA3	21.90526321	26.32
FGFR2	21.92818368	67.11
ANGPT2	21.94015081	68.42
FOXD3	21.98262749	40.79

Higher in trailblazers compared to quartile 1	p value	fold change
BAMBI	0.00E+00	16.82
PCDH10	9.52E-02	7.90
FGF8	1.40E-02	5.60
CCR9	0.00E+00	5.57
ELAV4	5.13E-02	4.89
FGFR2	5.93E-06	3.85
EPHB1	2.92E-06	3.66
NRP1	1.28E-03	3.34
CDH7	5.74E-07	3.01
MMP9	3.11E-02	2.94
PDGFRL	4.46E-06	2.84
AQP1	2.59E-04	2.66
CXCR1	4.35E-13	2.65
CDH6	1.16E-02	2.50
CXCR7	2.46E-04	2.32
SLIT1	1.83E-04	2.17
CXCR4	1.76E-09	2.15
PKP2	2.05E-12	2.13
CDH11	1.53E-05	2.08
CCR7	4.40E-03	2.08
CTNNB1	3.60E-08	1.81
ITGB5	5.98E-05	1.73
EPHB3	2.78E-03	1.72
SLIT2	3.36E-02	1.64
SNAI2	3.55E-02	1.60
NEDD9	3.37E-02	1.57
NOTCH1	3.47E-02	1.49

Higher in trailblazers compared to quartiles 2-4	p value	fold change
BAMBI	0.00E+00	20.60
HAND2	2.21E-05	18.53
PCDH10	3.64E-02	6.02
NRP1	5.28E-07	4.54
CCR9	0.00E+00	4.21
AQP1	7.77E-15	3.24
NEDD9	4.96E-10	3.21
CXCR1	0.00E+00	3.10
CXCL13	7.27E-02	3.02
ITGB5	3.04E-13	3.01
PDGFRL	5.96E-11	2.83
JAG1	6.45E-07	2.83
MMP2	2.67E-07	2.78
EPHA1	3.52E-04	2.76
FGFR2	1.13E-06	2.64
CXCR7	1.26E-06	2.52
SLIT2	2.30E-06	2.42
PKP2	0.00E+00	2.22
ISL1	2.49E-02	2.11
RUNX2	4.30E-02	2.06
NEFM	2.59E-03	2.01
KDR	9.39E-02	1.99
NOTCH1	7.14E-07	1.99
CTNNB1	0.00E+00	1.97
CXCR4	4.29E-11	1.87
EPHB3	2.10E-05	1.83
SOX9	3.86E-02	1.81
EPHB1	1.15E-04	1.81
CCR7	2.83E-03	1.68
CDH11	6.32E-05	1.62
CXCL12	1.57E-03	1.57
CDH6	9.43E-02	1.48
CDH7	1.49E-02	1.41
UNC5B	4.22E-02	1.40
BMPR2	6.78E-02	1.28
SNAI2	3.68E-02	1.28
ROBO1	6.06E-02	1.16

Gene	p	Higher in HAND2 MO	Lower in HAND2 MO
HAND2	0.06899	2.17	
EPHA6	0.08247	2.023	
CDH7	0.07705	0.723	1.383125864
CDH6	0.04665	0.692	1.445086705
CDH11	0.03969	0.529	1.890359168
ITGA9	0.05754	0.696	1.436781609
ITGB3	0.01035	0.542	1.84501845
TFAP2A	0.09821	0.657	1.522070015
SPON1	0.03555	0.574	1.742160279
SOX10	0.02956	0.479	2.087682672
PAX3	0.02137	0.416	2.403846154
FGFR2	0.08873	0.846	1.182033097
FGFR1	0.03253	0.734	1.36239782
EPHB1	0.01072	0.676	1.479289941
NRP2	0.02771	0.593	1.686340641
CXCR4	0.04437	0.292	3.424657534
AQP1	0.07819	0.301	3.322259136
WISP1	0.05466	0.23	4.347826087
KDR	0.0691	0.347	2.88184438

Gene	p	Higher in TFAP2A MO	Lower in TFAP2A MO
HAND2	0.002251	3.764	
EPHA6	0.08972	2.683	
MMP2	0.01251	2.165	
ITGB5	0.02536	1.471	
NEDD9	0.08225	1.321	
FGFR2	0.08381	1.067	
EPHB1	0.05839	0.703	1.422475107
CXCR4	0.06061	0.4	2.5
WISP1	0.07412	0.345	2.898550725