**Neural Crest Invasion is Driven by Trailblazer Cells**

**with a Unique Molecular Signature**

**Date: 8 July, 2014**

**Target Journal: Development (7000 words, including figure legends, excluding references)**

**ABSTRACT (<250):**

Neural crest cell migration is critically important to the formation of peripheral tissues during vertebrate development. However, difficulties in studying cellular and molecular dynamics have limited our understanding of how neural crest cells respond to different microenvironments to form the pattern of discrete, multicellular streams. Here, we address this by examining the gene expression dynamics of cranial neural crest cells during migration. By dividing a cranial neural crest migratory stream into 8 subregions, we determined that neural crest expression profiles are highly dependent on its stream position and local microenvironment. We confirmed the spatial and temporal diversity of neural crest gene expression using multiplexed in-situ hybridization. Both computational model simulations and single cell profiling experiments performed in parallel showed that only a small subregion of the neural crest stream is composed of ‘trailblazers’ with a molecular signature that is consistent between cells and stable over time. Manipulation of two transcription factors identified by high expression within the trailblazers, heart and neural crest derivatives expressed 2 (HAND2) and transcription factor activating enhancer binding protein 2 alpha (TFAP2A), revealed dramatic changes in the cranial neural crest cell migratory pattern. Computational model predictions that simulated these experimental scenarios showed similar disruptions to multicellular streaming. These data on neural crest migration identify a trailblazer subpopulation with a unique molecular signature that includes genes critical to the proper migration pattern.

**INTRODUCTION**

Embryonic cell migration is critically important to the development of peripheral tissues and organs. In examples that range from primitive streak formation to mechanosensory organogenesis, cells travel long distances in a persistent direction as coordinated groups (Tarbashevich and Raz, 2010; Voiculescu et al., 2014; Piotrowksi and Baker, 2014). Mistakes in embryonic cell migration patterns may result in birth defects or tumor formation as cells in errant locations often die or fail to properly differentiate.

One of the fascinating observations of long distance cell migration during development is the movement of cells in multicellular streams. During multicellular streaming, a cell independently controls its cytoskeleton but moves with neighbors as a population. This phenomenon has raised the long standing question of what mechanisms regulate the persistence and cohesion of multicellular streams. Although large scale genomic analyses have shed light on genes expressed in 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), it remains unclear how gene expression dynamically changes as migrating cells travel through different microenvironments. Therefore, investigative efforts that correlate dynamic molecular interrogation with in vivo cell behavior analyses will yield important insights into embryonic cell migration events at the level of both individuals and cell populations.

The neural crest is one of the most striking examples of long distance embryonic cell migration that is accessible to manipulation and in vivo. Neural crest cells emerge from the dorsal neural tube and are sculpted by intrinsic and extrinsic signals into a pattern of discrete streams throughout the head and trunk (Kulesa and Gammill, 2010). In nearly every vertebrate organism, neural crest cells move in multicellular streams, characterized by a cell’s independent control of its cytoskeleton but movements with neighbors as a loosely connected population (McGurk et al., 2014; Kulesa et al., 2013; Young et al., 2014; Epperlein et al., 2007). In chick, cranial neural crest cells respond in a chemotactic manner to vascular endothelial growth factor (VEGF) signaling that is present within the embryonic neural crest microenvironment (McLennan et al., 2010). Gene expression profiling using fluorescence activated cell sorting (FACS) and RT-qPCR to isolate and profile the lead 30% of the chick cranial neural crest cell migratory stream, revealed a distinct pattern of expression (45/84 genes analyzed) in comparison to the 70% trailing subpopulation (McLennan, Dyson et al., 2012). These results supported our previous in vivo cranial neural crest cell behavioral analyses in chick that showed regional differences in cell speed, direction, mitotic and calcium activity, and cell morphology depending on cell position within the migratory stream (Kulesa and Fraser, 1998, 2000; Teddy and Kulesa, 2004, Kulesa et al., 2008; McKinney et al., 2011; Ridenour et al., 2014). What remains to be determined is how embryonic neural crest microenvironmental signals influence gene expression dynamics and whether these signals are relevant to proper neural crest migration.

To address these questions, we first isolated neural crest cells from 8 discrete segments of the cranial migratory stream using LCM and then analysed the expression of 77 genes using RT-qPCR. We confirmed regional differences in gene expression and resolved over-lapping patterns using multiplexed fluorescent in-situ hybridization (HCR). To examine the stability and consistency of gene expression within the invasive migratory front, we isolated and profiled single neural crest cells at two distinct time points: during migration toward, and entry into the branchial arches. Thus we were able to identify a molecular signature unique to the narrow invasive front of the stream. We then used gain- and loss-of-function to determine whether genes expressed as part of this molecular signature of the invasive migratory front are critical to neural crest stream dynamics. This was done in parallel with computational modeling that simulated our experimental scenarios. Our results highlight the heterogeneity of molecular profiles along the cranial neural crest migratory stream, and the requirement for only a small subset of informed leaders we call ‘trailblazers’ to correctly guide proper migration.

**RESULTS**

**Molecular profiling reveals regional diversity in gene expression with a gradual transition in patterning between subregions of the cranial neural crest stream**

We previously used computational modeling and molecular profiling to reveal the existence of at least 2 subpopulations of cells with distinct molecular profiles within a typical cranial neural crest cell migratory stream (McLennan et al., 2012). However, it remained unclear whether these distinct gene profiles varied in a continuous or disjointed manner from the invasive front to the most proximal positioned newly emerged neural crest cells.

To address this question, we subdivided the pre-otic, cranial neural crest migratory stream into 8 distinct subregions and isolated neural crest cells from each subregion (Fig. 1A-D). Hierarchical clustering of 77 genes revealed distinct differences in gene expression between each of the 8 subregions. Subregions 1-3, 4-5, and 6-8, clustered together, respectively (Fig. 1E-F). When we organized the genes according to their level of linear expression in subregion 1 (invasive front) relative to subregion 8 (newly emerged from the neural tube), we found a gradual transition of gene expression changes between the clustered subregions (Fig. 1F). Some of the genes highly expressed within the invasive front (subregion 1) are gradually diminished in expression towards the proximal trailing subregions of the migratory stream (Fig. 1F). Likewise, some genes expressed at low levels within the invasive front (Fig. 1F; bottom half of graph) show a gradual increase in expression when compared throughout the more proximal subregions. These data reveal the widespread regional diversity in gene expression profiles based upon cell position within the cranial neural crest cell migratory stream.

**Hierarchical clustering shows uniqueness of the distal stream and 4 characteristic patterns of expression within the neural crest migratory stream**

The lead three subregions (1-3; corresponding to the invasive front and distal portion of the cranial neural crest cell migratory stream) cluster very closely with one another and are distinct from subregions 6-8 (Fig. 1F). 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. 1H). Of these 34 genes, 38% (13/34) are upregulated and 62% (21/34) are downregulated (Fig. 1H). Upregulated genes include the transcription factors distal-less homeobox 5 (DLX5), HAND2 and TFAP2A, as well as other neural crest-associated genes such as sex determining region Y-box 9 (SOX9), snail homolog 1 (SNAI1), neural precursor cell expressed developmentally down-regulated 9 (NEDD9) and ADAM metallopeptidase domain 33 (ADAM33) (Fig. 1H). These results reveal a more significant difference in the expression profile of the lead three subregions compared to newly emerged neural crest cells than our previous analysis (McLennan et al., 2012).

To determine whether there is a unique molecular profile of the neural crest cells within the most invasive migratory front, we compared the molecular profiles of subregion 1 (lead 12.5%) to those cells immediately proximal in subregions 2-3 (Fig. 1G). The invasive migratory front shows a molecular profile distinct from cells within subregions 2-3 (Fig. 1G). This distinct molecular profile shows significant up or down-regulation of 18% (14/77) of the genes analyzed (Fig. 1I). Genes significantly upregulated in subregion 1 compared to subregions 2 and 3 included HAND2, aquaporin 1 (AQP1), BMP and activin membrane-bound inhibitor homology (BAMBI), glypican 3 (GPC3) and matrix metalloproteinase 2 (MMP2) (Fig. 1I). When we compared the molecular profile of lead (subregion 1) neural crest cells to all other migrating neural crest cells (subregions 2-8), we found similar genes were enriched in the invasive front (HAND2, GPC3, MMP2) (Fig. 1J). Thus, there is a unique molecular profile of the most invasive migratory front.

To determine whether the neural crest molecular profiles could be categorized into distinct patterns, we examined the expression of individual genes throughout the migratory stream (Fig. 2E). First, we found a molecular profile where a gene was expressed higher in the invasive front and expression decreased towards the proximal direction to the most recently delaminated neural crest cells (Fig. 2A). Second, we observed a molecular profile that peaked in expression in the subpopulation of migrating cells within the middle portion of the migratory stream, corresponding to neural crest cells that were actively following the lead neural crest cells (Fig. 2B). Third, there are genes expressed higher in the most recently delaminated neural crest cells in a molecular profile that diminished towards the invasive migratory front (Fig. 2C). Lastly, some genes show a uniform expression pattern across all subregions of the cranial neural crest cell migratory stream (Fig. 2D). Thus, the regional differences in gene expression in the cranial neural crest cell migratory stream vary with 4 distinct patterns.

**Multiplexed fluorescent in-situ hybridization (HCR) confirm the regional expression patterns of genes expressed within the cranial neural crest cell migratory stream**

To confirm regional differences in gene expression within the neural crest cell migratory stream determined by our LCM/RT-qPCR 8-segment analysis, we applied a newly emerging technology known as multiplexed fluorescence in-situ hybridization (HCR) (Choi et al., 2010, 2014). HCR takes advantage of initiator probes and hairpins to amplify mRNA signal within cells as a fluorescent readout of gene expression that allows more accurate quantitative analysis to resolve gene expression patterns and borders (Choi et al., 2010, 2014). We applied HCR to simultaneously visualize the gene expression profiles of 2 genes highly expressed by cells within the most invasive front (HAND2 and BAMBI) and included immunolabeling to distinctly observe neural crest cells (using HNK-1). (Fig. 3). Visual analysis revealed the complex, distinct patterns of BAMBI and HAND2 expression that confirm restricted expression in the lead neural crest cells and mid-stream subregions (Fig. 3). Polyline kymograph analysis that samples the fluorescence intensity in a 20-um wide band along the cranial neural crest migratory pathway shows the localization of gene expression of BAMBI and HAND2 in neural crest cells, confirming the RT-qPCR data (Fig. 3). Thus, the HCR technique allowed us to quantitatively confirm the molecular profile of multiple genes simultaneously from our LCM/RT-qPCR data and support regional differences in gene expression within the cranial neural crest cell migratory stream.

**Computational modeling predicts that a small fraction of lead cells could more successfully drive entire stream migration**

To test the logical conclusions of our mechanistic hypotheses, we used a hybrid mathematical model (McLennan et al, 2012), with individually represented cells and a continuous chemoattractant concentration (Fig. 4A-D). In this model, cells undertake a two-dimensional off-lattice jump-process on a growing rectangular domain (Fig. 4A-D). New cells enter the domain from one end throughout the simulation to represent neural crest cell delamination from the dorsal neural tube (Fig. 4D). The chemoattractant concentration is described by a reaction-diffusion equation, with the cells acting as sinks, representing internalization of chemoattractant. The key model components are illustrated in Fig. 4, and model parameters are listed in Table 1.

We extended our computational model to include aspects based on new experimental data and performed computational modeling simulations in parallel with the experiments presented in this paper (Fig. 4E). Characteristic changes to the previous version of the model (McLennan et al., 2012) are:

* A wider stream of cells allows for greater cell numbers and more adequate representation of multicellular stream migration.
* Cells have a finite sensing accuracy (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.
* A range of intercellular distances is now allowed before cells can no longer stay in contact, cease to communicate directional information, and the stream breaks up. These are based on microscopy measurements of cells sizes and filopodial lengths, and improve stream cohesion, reducing stream break-up in model migration.

We used our updated computational framework to test aspects of our hypotheses that are difficult to probe in vivo. Specifically, we wanted to know whether the number of lead cells is a critical factor for the success of neural crest cell invasion. Should this be the case, then the gene expression patterns detailed in vivo (Figs. 1,2) would be consistent with this constraint. To test this, we first restricted the model simulations to only include non-plastic cell behaviors, such that individual cells that start migration as leaders (or followers) could not switch from a leader to a follower (or vice versa). To change the fraction of total migrating cells that are leaders, we varied the time of transition between lead and follower cells (Fig. 5A), *tLF,* so that all cells that entered the migratory domain up to time *tLF* were leaders, and cells that entered after that were followers.

Our simulations reveal that the median distance migrated and the stream density both increase with decreasing leader fraction (Fig. 5B). 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, since it prevents jamming and enables a higher number of cells to distribute more evenly along the migratory route (Fig.5B, C).

We discovered that smaller leader fractions improve multicellular stream migration (Fig. 5B). Model simulations show that a small lead fraction of neural crest cells is successful in correctly distributing more cells throughout the migratory domain (Fig. 5C). Thus, our computational model predicts that a small number of lead cells can efficiently guide the entire neural crest migratory stream migration.

**Single cell analysis of the invasive front identifies a subpopulation of trailblazer neural crest cells with a stable and consistent set of highly expressed genes**

To determine whether there exists a consistent and stable set of highly expressed genes within the invasive front of the neural crest stream, we performed single cell analysis. That is, we examined at the single cell level whether gene expression profiles were consistent amongst neighboring cells within the invasive front. To do this, we isolated (via manual dissection followed by FACS) of neural crest cells directly within the migratory front (no other transfected neural crest cells in front of them, 1-3 cells per embryo), at two phases of cranial neural crest cell migration: HH St 13 corresponding to active migration entering the target area (branchial arch) and Stage 15 corresponding to migration throughout the target area (Hamburger and Hamilton, 1951) (Fig. 6A). FACS isolation was the most efficient method for analyzing single neural crest cells while maintaining the native molecular profile (Suppl Fig.). We profiled 96 genes, composed of the 77 genes previously used for profiling plus additional genes of interest (Fig. 6B).

Single lead neural crest cells within the invasive front have a high degree of gene expression homogeneity at both phases of migration, as can be seen in the profiles of individual cells (Fig. 6C). Principle component analysis (PCA) plots of the genes at each phase of migration show that over 50% of the genes are stably expressed with little variation (Fig. 6D, blue squares; Suppl. Table 1). Violin plot analysis reveals similar levels of expression in a range of genes, including BAMBI, NOTCH1 and interleukin 8 receptor (CXCR1) (Fig. 6E). In contrast, many genes are dynamically expressed when comparing the two developmental stages, such as HAND2, SNAI1, and SOX10 (Fig. 6E). We focused our attention on the highest expressed genes with RT-qPCR Ct values of less than 22 (Table 2). Strikingly, of the 62 most highly expressed genes within the invasive migratory front at HH St 15, 98% of the genes (61/62) are also highly expressed at HH St 13 (Table 2). Many of the highly expressed genes are consistently expressed by a large percentage of the profiled invasive migratory front at both phases of migration, including BAMBI, CXCR1, NOTCH1, plakophilin 2 (PKP2) and TFAP2A (Fig. 6E, Table 2). However, there are a few examples of the highest expressing genes being expressed by dramatically different numbers of cells within the invasive migratory front. For example, HAND2 is expressed in a small number of lead neural crest cells during the first phase of migration (8%) and this dramatically increases during the second phase to 60% of the lead cells during colonization of the target tissue (Fig. 6E). In contrast, the percentage of neural crest cells expressing Sox10 dramatically decreases (from 69% to 21% ) over time (Fig. 6E). Therefore, we restricted our focus to genes that are expressed highly by at least 50% of the cells at both time points (Table 2). Together, these data reveal a stable and consistent set of genes expressed within the invasive migratory front.

**Trailblazers have a molecular signature that is restricted to the invasive front of the migratory neural crest cell stream**

To determine whether the molecular signature unique to the cells within the neural crest invasive front was shared elsewhere, we compared the molecular profile of trailblazers to single cells isolated from 4 distinct quartiles of the cranial neural crest cell migratory stream (Fig. 7A). We measured the similarity of molecular profiles across 327 single neural crest cells by Pearson correlation (Fig. 7B) and hierarchical clustering (Fig. 7C, D). Our analysis revealed 3 main results. First, the vast majority of cells from the other proximal quartiles have a poor correlation with the molecular profile of the trailblazer neural crest cells, but neural crest cells within each quartile have high correlations with one another (Fig. 7B). Second, the average expression of neural crest cells in the quartiles cluster according to their position within the migratory stream or within the local microenvironment depending on the type of clustering method (Fig. 7C, Euclidean; Fig. 7D, Pearson). Third, the gene expression profile of the invasive front at HH St. 15 is distinct from all quartiles, but most similar to quartile 1 (Fig. 7C, D). At the single cell level, the principle component analysis shows incomplete overlap between genes expressed within the invasive front and quartile 1 (Fig. 7E). Violin plots, which show the distribution of expression of a particular gene reveal examples of individual genes that were expressed at higher (BAMBI, CXCR1, PKP2, HAND2) or lower (Sox10, integrin alpha 3 (ITGA3)) levels by the invasive front than cells located more proximal within the neural crest cell migratory stream (Fig. 7F). Thus, the unique molecular signature associated with cells at the invasive front comprises a much smaller, distinct subpopulation of cells than the entire first quartile (Fig. 7). This gene profile is not shared by other migrating neural crest cells within the the rest of the stream.

To refine the unique molecular signature of the invasive front, we examined the genes that were differentially expressed between the invasive front compared to the genes expressed by other neighboring cells in quartile 1 at HH St 15 (Table 3). This resulted in a list of 17 genes that are highly expressed in trailblazers and differentially expressed between the trailblazers and quartile 1. This list however does not take into account the comparison with the remainder of the stream (subregions 2-4). When we compared the invasive front profile to single cell profiles from cells isolated throughout the stream, we found 94% (16/17) of the genes in agreement (Table 3). Thus, 16 genes are more representative of the unique molecular signature of the invasive front of the cranial neural crest cell stream.

**The transcription factors HAND2 & TFAP2A were identified as likely to influence the molecular signature of trailblazers and therefore their migration**

To test the criticality of the genes enriched in the invasive front during neural crest cell migration, we examined the gain- and loss-of-function of two candidates, TFAP2A and HAND2. We selected the transcription factors TFAP2A and HAND2 based on their upstream signaling position and their enrichment within cells in the lead 3 subregions (Fig. 1H). HAND2 and TFAP2A were also expressed at high levels by the invasive front at both of the phases of neural crest migration (Table 2). Lastly, all of the genes in this unique molecular signature of the invasive front are regulated by either HAND2, TFAP2A or by both (Fig. 11, compiled from pathway studios software).

**Knockdown of HAND2 & TFAP2A expression results in alterations to the neural crest cell migratory pattern**

We hypothesized that cells within the invasive front have a unique molecular signature that is critical to the proper migration of the cranial neural crest migratory stream. To test this hypothesis, we knocked down TFAP2A and HAND2 using morpholinos (Fig. 8A-C). We measured distance migrated and the area covered by the morpholino transfected neural crest cells. Loss of HAND2 function in migrating cranial neural crest cells results in significant reduction in the area covered by treated cells when compared to control cells (Fig. 8A-B, H-I). Upon close examination, the fluorescence associated with individual neural crest cells within HAND2-morpholino embryos was more punctate than in control morpholinos, suggesting cells had stopped and collapsed protrusions (Fig. 8A, B). The consistent punctate labeling in HAND2-morpholino embryos was not due to cell death, since several morpholino transfected neural crest cells did correctly migrate into distal portions of the branchial arches (Fig. 8B, D-G). This suggests that loss of HAND2 function does not restrict cells from migrating to the end of the migratory pathway.

Loss of TFAP2A function results in more dramatic alterations to the cranial neural crest cell migratory pattern (Fig. 8A, C). First, the distribution of the transfected neural crest cells along the migratory pathway is perturbed when compared to neural crest cells transfected with control morpholino measured at both phases of migration (Fig. 8D, E). The trend of the distributions of TFAP2A morpholino transfected neural crest cells is significantly different from control morpholino embryos due to a significant drop in cell number in the lead subregion of the migratory stream (Fig. 8D, E). During the first phase of migration (HH St13), the distance migrated and area covered by TFAP2A morpholino transfected neural crest cells are not statistically different to controls (Fig. 8F, H). However, during the second phase of migration (HH St 15), neural crest cells transfected with the TFAP2A morpholino stopped and failed to migrate the entire distance of the migratory route (Fig. 8G). From these results, we conclude TFAP2A is necessary for the proper migration of the cranial neural crest cells into the branchial arches.

**HAND2 and TFAP2A influence the molecular signature of the neural crest cell invasive front**

To determine the response of the network of genes that were affected by the loss of HAND2 or TFAP2A function, we profiled neural crest cells 24 hours after transfection with HAND2 or TFAP2A morpholinos. We focused on genes we identified within the unique molecular signature of the invasive front (Fig. 11). We discovered that when HAND2 is knocked down, AQP1, CDH11, CDH7, CXCR4 and EPHB1 are all significantly downregulated (Table 4). Interestingly, we also observed that TFAP2A is downregulated in neural crest cells in the invasive front in HAND2 morpholino transfected embryos, indicating crosstalk between the two transcription factors (Table 4). Loss of TFAP2A function by morpholino transfection results in the down-regulation of CXCR4 and EphB1, and the upregulation of genes within the invasive front, integrin beta 5 (ITGB5) and NEDD9 (Table 5). Loss of HAND2 or TFAP2A by morpholino transfection resulted in upregulation 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 neural crest cells influences the expression of other genes within the molecular signature of the invasive front.

**Computational model simulations that mimick a forced leader behavior within the trailing subpopulation predict alterations to the neural crest migratory pattern**

In anattempt to model the perturbation of the trailblazer molecular signature, we modified parameters to convert every other trailing cell into a leader (Fig. 9A). This mimicks, in silico, the transfection of 50% of the neural crest cells within the trailing subpopulation of the migratory stream with the transcription factors identified in the invasive front molecular signature. When we forced trailing cells in silico to display a lead phenotype, we found that cells remain near the entrance of the migratory domain (Fig. 9B, C). This migratory pattern is similar to the scenario when we introduced higher lead cell fractions in the simulation model (Fig. 5C). Thus, forcing a lead cell behavior within the trailing subpopulation in silico predicts severe disruption to the neural crest cell migratory pattern that would be visible as cell bunching near the dorsal neural tube exit.

**Overexpression of HAND2 & TFAP2A alters the neural crest cell migratory pattern in a similar manner to the computational model predictions**

To test whether gain-of-function of the invasive front molecular signature transcription factors, TFAP2A or HAND2, in the trailing subpopulation of neural crest cells would affect the migratory pattern, we overexpressed these genes and measured changes to the stream shape and cell distance migrated (Fig. 10). To restrict the overexpression to trailing neural crest cells, we electroporated pre-migratory neural crest cells at later stages of development (HH St 10 vs HH St 8+), allowing lead neural crest cells to be unaffected. Overexpression of TFAP2A does not appear to change the distribution of migrating neural crest cells compared with control EGFP transfected embryos (Fig. 10C, D-F). However, when HAND2 is overexpressed in trailing cells, the migrating neural crest cells distribute more evenly along the migratory route when compared to the distribution of control EGFP transfected neural crest (Fig. 10A, B, D). Furthermore, when HAND2 is overexpressed in trailing cells, fewer neural crest cells are observed along the migratory route (Fig. 10E). Thus, forcing the expression of the invasive front molecular signature transcription factor, HAND2, into the trailing neural crest has a distinct and significant effect on the neural crest migratory pattern.

**DISCUSSION**

In this study, we used the embryonic neural crest to examine the molecular signals that underlie long distance migration. Dynamic in vivo time-lapse imaging had previously revealed differences in neural crest cell behaviors. These differences appeared to depend on cell position within multicellular streams at the phase of migration. Using novel methods to isolate and profile single and small numbers of migrating chick cranial neural crest cells, we analyzed and compared gene expression patterns within subregions of the migratory streams during migration towards, invasion and colonization of the branchial arches. We performed gain and loss of function experiments to test whether the regional gene expression differences are critical for migration. In parallel, we applied our understanding of neural crest migration in an hybrid computational model to simulate and predict experimental outcomes.

Our work approaches discerning just how embryonic migratory cells respond to microenvironmental cues and whether an understanding of those cues might help predict invasive behaviors. The early embryonic hindbrain region uses a complex set of signaling mechanisms to control and guide the invasion of the neural crest (ref recent reviews). Extrinsic signals within the various microenvironments through which migrating neural crest cells travel appear to regulate the gene expression profile of cells and create distinct subregions within the neural crest multicellular stream. Subregions of cranial neural crest cells isolated and analyzed during migration toward and colonization of the branchial arches have distinct gene expression profiles that cluster in similarity depending on stream position and have 4 characteristic patterns (Fig. 1,2). These measurements demonstrate that neural crest cells are dynamically responsive to different microenvironmental cues and suggest that an understanding of these signals may provide insight into the invasive abilities of multicellular stream migration.

One key question was: What is the size of the lead cell population required to guide the proper migration of the entire multicellular stream? Our previous computational model and qPCR analysis showed proper neural crest migration required at least two separate cell populations (leaders/trailers) with distinct cell phenotype and genotype (McLennan, Dyson et al., 2012). If the number of lead cells was reduced, we would expect to observe a reduction in the migratory efficiency of a multicellular stream. Likewise, if the lead cells were restricted to the invasive front, we would expect alterations to the stream since guidance information transfer to trailing cells would be less efficient. However, we were able to show that the furthest distance migrated was insensitive to the number of lead cells. Even a few lead cells could migrate as far as the entire multicellular stream. This suggests that only a few lead cells with guidance information are required for proper stream migration.

Single cell analysis of lead neural crest cells provided further insights into the molecular characteristics of the lead population. Our single cell analysis led to the identification of neural crest cell trailblazers (Fig. 6). Trailblazers are defined as highly migratory neural crest cells at the very front of the stream with no other cells in front of them. The trailblazers have a molecular signature that is distinct from the rest of the migratory stream (Fig. 11). This signature includes genes involved in cell-cell adhesion (CDH7, CDH11, CTNNB1, and PKP2), cell-extracellular matix (ECM) adhesion (ITGB5), differentiation (NEDD9, NOTCH1, SNAI2), guidance (CCR9, CXCR1, CXCR4, CXCR7. EphB1, EphB3), proliferation (BAMBI) as well as cell shape dynamics (AQP1). Importantly, neural crest cells with this trailblazer molecular signature were exclusively localized to the front of the migratory stream.

The role of the neural crest trailblazer is critical to our understanding of neural crest migration. Our working hypothesis for cranial neural crest directed migration is that of a cell-induced gradient. Computational model simulations that tested the gain-of-function of lead cell genes within the trailing subpopulation predicted alterations to the neural crest cell migration pattern. Specifically, when these experiments were performed in ovo, we observed alterations to the neural crest migratory pattern. Knockdown of TFAP2A in lead neural crest cells resulted in the cells failure to properly invade the branchial arches (Fig. 8). This suggests that only a small number of trailblazers are necessary to guide the entire stream to the correct destination. That is, trailblazers read out surrounding microenvironmental cues and decide what direction to migrate. Cells within the proximal stream remain cohesive with each another as well as with the trailblazers and follow the trailblazers to the correct destination.

Whether the following cells follow because of physical connections to the trailblazers and/or pathways/cues laid down by the trailblazers remains to be seen. This phenomenon is not unlike collective migration seen in animals (Couzin et al., 2005). Large herds, flocks or schools of animals depend on social interactions for information transfer and only need a small proportion of them to be leaders for accurate migration (Couzin et al., 2005; Guttal and Couzin, 2010; Leonard et al., 2012; Miller et al., 2013). Just like in animal models, neural crest cells gain information from their local microenvironments. Animals sense information about food, predators and weather from their surroundings as well as population density and health from each other. Neural crest cells sense information about positive cues, negative cues, and permissiveness from their surroundings as well as population density and health from each other.

Importantly, testing the effect of number of leader cells on migration required that we restrict our in silico experiments to the case of fixed subpopulations, even though in ovo transplantation experiments suggest neural crest cell behavior is influenced by the local microenvironment. We disabled phenotype switching in our computational model, for otherwise we would have only been able to change the leader fraction transiently, before phenotype switching restored it towards the unperturbed case. Alternatively, one could set up the computational experiments with varying initial leader fractions and, given appropriate microenvironmental parameters, observe what leader fractional emerges naturally from phenotype switching. However, this requires 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.

The results of our simulations depend on its parameters, and with arbitrary fine-tuning we could probably find a regime in which migration works well with only leader cells. But, our working hypothesis for cranial neural crest directed migration is that of a cell-induced gradient. Hence, our in silico experiments operate in this parameter regime. This is motivated by a) the lack of evidence for a pre-existing morphogen gradient to guide chick cranial neural crest migration, b) a uniform production of known neural crest cell attractant vascular endothelial growth factor (VEGF) (McLennan et al., 2010), c) as well as evidence for migration in a cell-induced gradient in other systems, such as the zebrafish lateral line primordium (Streichan et al., 2011). The assumption of a finite sensing accuracy also strikes us a biologically realistic one, and our results are qualitatively robust over several orders of magnitude of sensing accuracies.

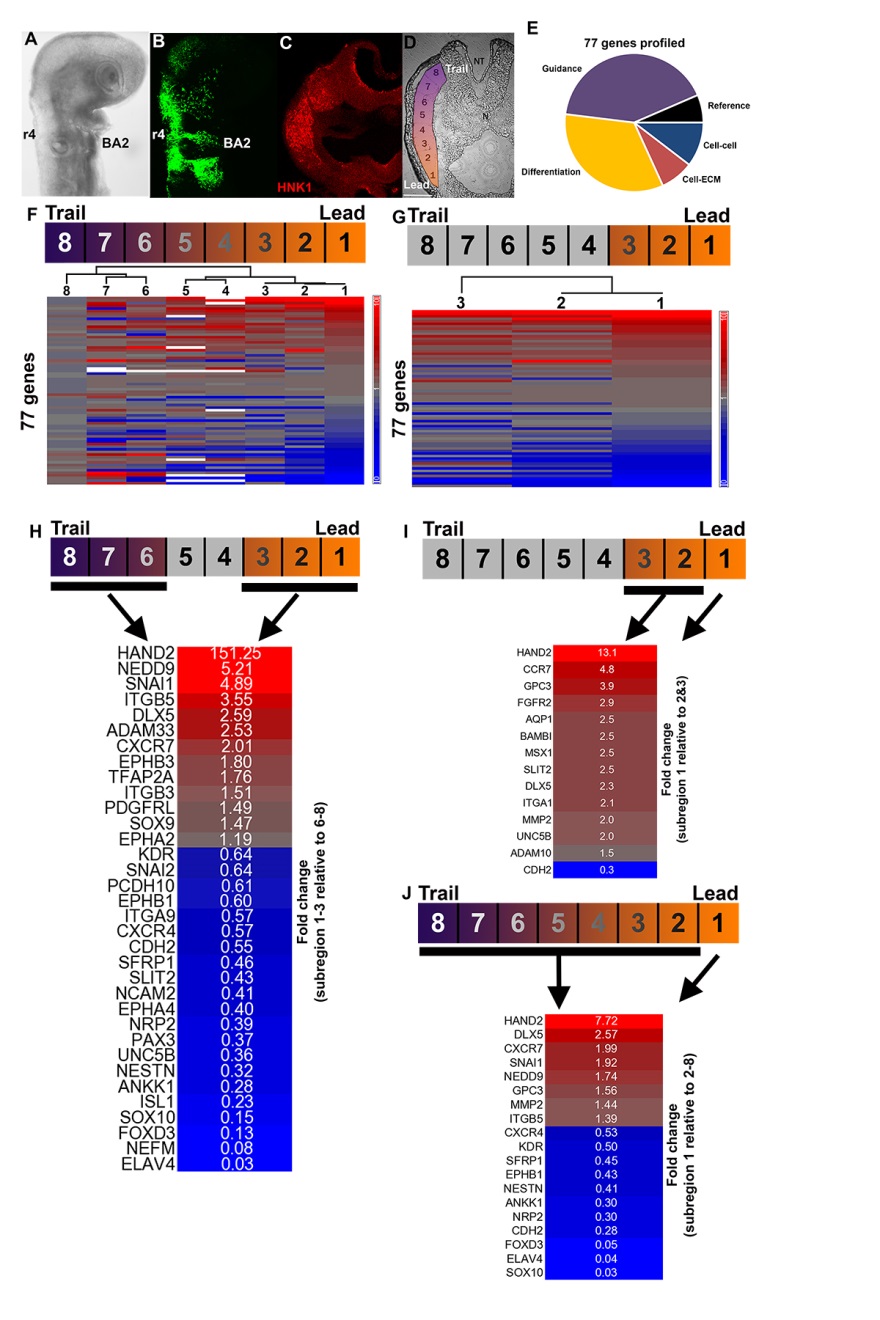
We chose to focus our perturbations on upstream transcription factors, HAND2 and TFAP2A. Neural crest cells with HAND2 knockdown do not show a migratory defect (Fig. 9), however this might be because at HH St 13, HAND2 expression is high but only in 8% of the trailblazers whereas by HH St 15, it is expressed highly in 59% of them. Therefore, HAND2 expression may not be critical during migration but more during branchial arch colonization and cell differentiation (Thomas et al., 1998). When the molecular signature of the trailblazers is perturbed via TFAP2A morpholino, both their migration and the migration of the following neural crest cells is affected. We have previously shown when leaders are prevented from migrating into the target site by a physical barrier, trailing neural crest cells sense the paused leaders, reroute around the barrier and become the new leaders (Kulesa et al., 2005). Here, because neural crest cells throughout the stream were transfected with the HAND2 or TFAP2A morpholinos, no cells were able to take on the role of the trailblazers and migration was hindered (Fig. 8).

The over-expression of these genes in neural crest cells further highlights their function in neural crest migration dynamics. When TFAP2A was overexpressed in the trailing portion of the migratory stream, no obvious defects were seen (Fig. 10). This was surprisingly as the TFAP2A knockdown phenotype was so dramatic. However, TFAP2A expression was a lot broader in the front half of the stream than HAND2 expression and therefore overexpressing this gene may not dramatically alter critical gene expression of downstream factors. When HAND2 was overexpressed, fewer neural crest cells were found in the migratory stream (Fig. 10). The neural crest cells that were found in the stream were evenly distributed along the route as opposed to being denser in the trailing portion as seen in the control EGFP and TFAP2A FL transfected neural crest streams (Fig. 10). Initially this was viewed as a possible delamination defect. However, after comparing this phenotype to the model simulations, it seems more likely that neural crest cells overexpressing HAND2 delaminate from the neural tube but then fail to migrate. The first 50um of migration in the model would correspond to migration from the midline to the edge of the neural tube in vivo. Our in vivo measurements were taken from the edge of the neural tube. Therefore, we hypothesize that HAND2 overexpression in vivo confers a trailblazer phenotype. Similar to the in silico modeling of trailer-to-leader switiching, the ectopically placed leaders fail to migrate due to a lack of VEGF in the trailing portion of the stream. Trailing neural crest cells follow the leaders and because the leaders fail to migrate, so do they.

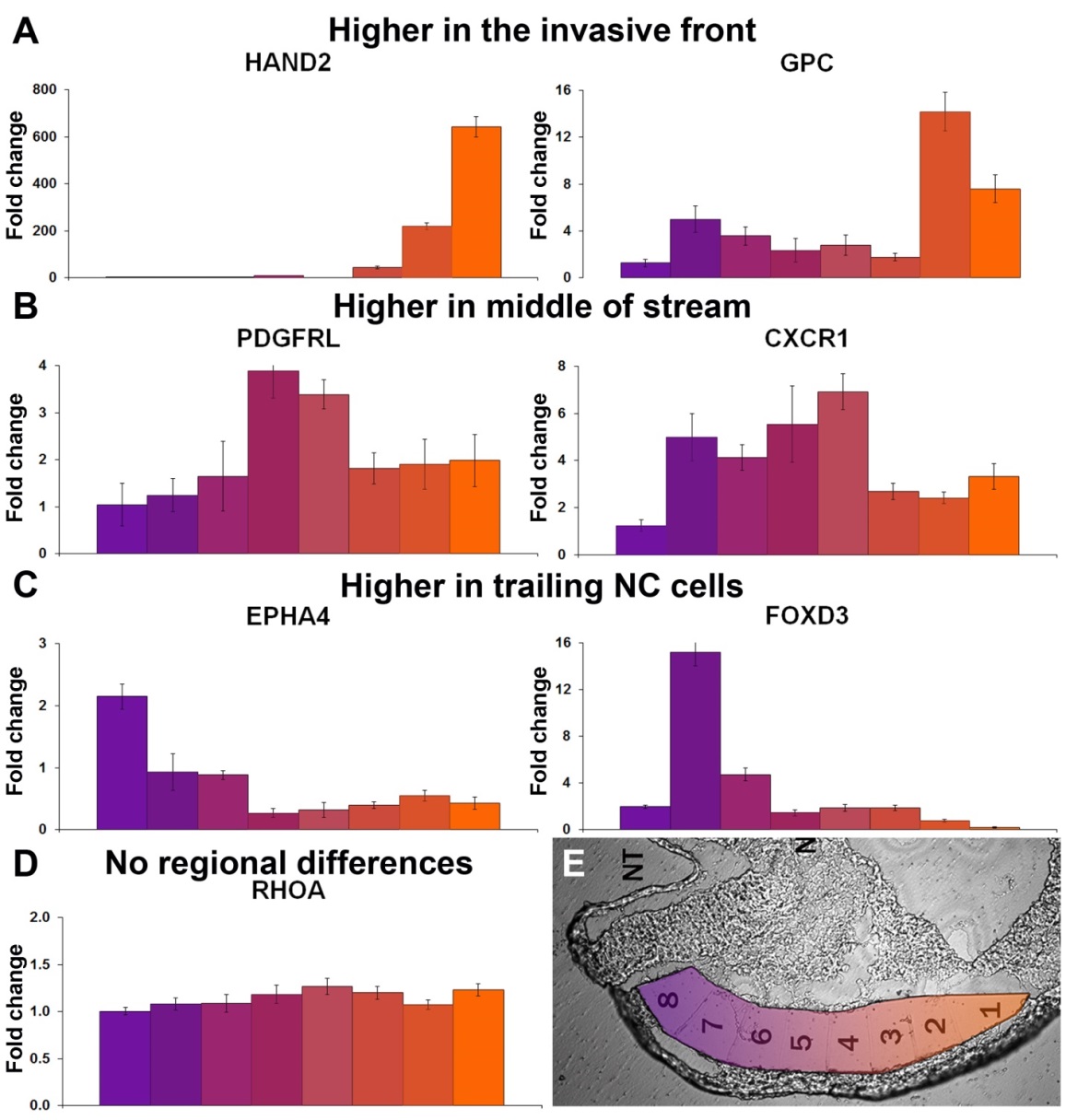
Importantly, we recognize the limitations of our current model. As an example, we have previously shown that knockdown of Np-1 causes neural crest cells in vivo to migrate up to, but not invade, the branchial arches. This phenomenon remains without a conclusive mechanistic explanation and is not captured by our computational model representation. This implies that a cell-induced VEGF gradient is not the only microenvironmental cue and demonstrates the complexity of this biological system and the importance of a systems-based in vivo approach toward understanding the multicellular streaming process.

In summary, the focus of this study was to understand how invasive cells respond to and interact with different microenvironments during embryonic development at the level of a single cell within a discrete multicellular stream. We show that the embryonic neural crest microenvironment exerts a powerful, controlling influence on the gene expression of migratory cranial neural crest cells. This regulation is enacted both spatially and temporally by the embryo and was most evident by the induction of a unique molecular signature within trailblazer cells at the neural crest invasive front. These data highlight the idea that successful invasion of cranial neural crest cells requires that a cell be able to interpret complex microenvironmental signals encountered along the migratory pathway. As an example, we show that TFAP2A and HAND2 act specifically to maintain the proper cranial neural crest migration pattern. Because loss of TFAP2A function disrupts neural crest stream invasiveness, we postulate that the failure to invade occurs by reducing the number of lead TFAP2A+ trailblazer cells within guidance information. Because gain of HAND2 function increases the number of stranded neural crest cells near the neural tube exit, we speculate that this occurs by forcing the lead cells to search for guidance signals already depleted by the trailblazer cells. These findings were made possible by studying neural crest cellular and molecular dynamics within the embryonic microenvironment.

**ACKNOWLEDGEMENTS**

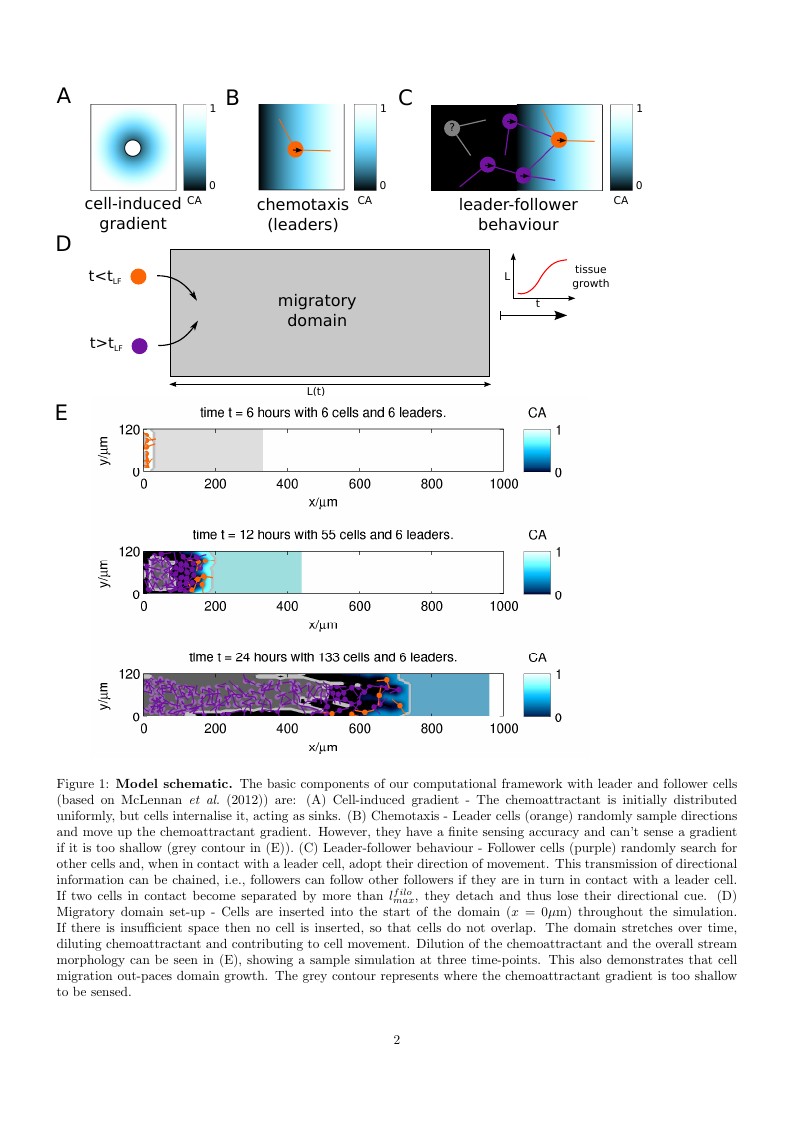


**Figure 1 – Distinctive regional expression profiles exist within the cranial NC migratory stream.** (A)Cranial region of an avian embryo at HH 15. (B) NC form stereotypical migratory streams within the developing vertebrate embryo. (C) HNK1 IHC labels NC within a transverse cryosection of the rhombomere 4 migratory stream at HH 15. (D) The rhombomere 4 NC stream at HH St 15 divided into 8 subregions by laser capture microdissection. (Subregion 1 corresponds to the distal-most invasive front, and subregion 8 refers to the proximal-most, newly emigrated NC cells.) (E) Subregions of the HH15 cranial NC stream were analyzed for the expression of 77 genes. (F) Hierarchical clustering of the 8 subregions from the HH St 15 cranial NC stream by Pearson dissimilarity based upon the 77-gene profile. All samples shown relative to subregion 8 with genes ordered by their level of linear expression in subregion 1. (G) Hierarchical clustering of 3 subregions at the invasive front of the HH St 15 cranial NC stream by Pearson dissimilarity based upon the 77-gene profile. (H) Heat map of 34 significant (p<0.1) linear gene expression differences between HH15 leading subregion 1-3 relative to trailing subregion 6-8. (I) Heat map of 14 significant (p<0.1) linear gene expression differences between HH15 leading subregion 1 relative to subregion 2-3. (J) Heat map of 19 significant (p<0.1) linear gene expression differences between HH15 leading subregion 1 relative to the rest of the stream (subregion 2-8).



**Figure 2 – Individual genes have regional expression differences within the cranial NC migratory stream.** (A) HAND2 and GPC3 expression is increased at the invasive front of the cranial NC migratory stream. All graphs are linear expression with error bars depict the standard error of the mean. Each subregion is represented by 3-6 biological replicates. (B) PDGFRL and CXCR1 expression is increased in the middle of the cranial NC migratory stream. (C) EPHA4 and FOXD3 expression is increased at the trailing portion of the cranial NC migratory stream. (D) RHOA expression is consistent throughout the cranial NC migratory stream. (E) The rhombomere 4 NC stream at HH St 15 divided into 8 subregions by laser capture microdissection.

**Figure 3 – HCR confirms of regional expression differences**



**Figure 4 – Model schematic.** The basic components of our computational framework with leader and follower cells (based on McLennan et al., 2012) are: (A) Cell-induced gradient - The chemoattractant is initially distributed uniformly, but cells internalize it, acting as sinks. (B) Chemotaxis - Leader cells (orange) randomly sample directions and move up the chemoattractant gradient. However, they have a finite sensing accuracy and cannot sense a gradient if it is too shallow (grey contour in (E)). (C) Leader-follower behavior - Follower cells (purple) randomly search for other cells and, when in contact with a leader cell, adopt their direction of movement. This transmission of directional information can be chained, i.e., followers can follow other followers if they are in turn in contact with a leader cell. If two cells in contact become separated by more than *a distance lfilomax*, they detach and thus lose their directional cue. (D) Migratory domain set-up - Cells are inserted into the start of the domain (x = 0µm) throughout the simulation. If there is insufficient space then no cell is inserted, so that cells do not overlap. The domain grows over time, diluting chemoattractant and contributing to cell movement. Dilution of the chemoattractant and the overall stream morphology can be seen in (E), showing a sample simulation at three time-points. This also demonstrates that cell migration out-paces domain growth. The grey contour represents where the chemoattractant gradient is too shallow to be sensed.

![A description...](data:None;base64,)

**Figure 5 - Multicellular stream migration can be more successful with smaller leader fractions.** To determine the effect of different (fixed) leader fractions on migration, we varied the time *tLF* after which newly inserted cells are followers (purple) instead of leaders (orange). Each stochastic simulation was repeated n=20 times (A). The migratory stream density (B) was then constructed from summing cell positions at the end of the n individual simulations. Leader fractions *fL* are the average for a given *tLF*. The greyscale bars are boxplots summarizing the distribution of distances migrated (in the x direction), with circles marking the median. Each boxplot is colored according to the relative stream density, calculated by dividing the total number of cells in n repeats by the maximum over all parameter combinations. A histogram of cell counts vs. distance migrated for the different leader fractions is shown in (C).

![A description...](data:None;base64,)

**Figure 6 – Trailblazer NC cells at the invasive front of the stream have a partially conserved molecular profile across distinct phases of migration.** (A) Isolation of single trailblazer NC cells, at distinct phases of long distance migration, by manual dissection and FACS. (B) HH13 & 15 trailblazer NC cells were analyzed for the expression of 96 genes. (C) Heat map of single trailblazer NC cells at HH 13 (n=72) & 15 (n=76) with genes ordered by mean log2 expression in all cells based upon the 96-gene profile. (D) Principal Component Analysis (PCA) weights of 96 genes analyzed in single HH 13 & 15 trailblazer NC cell profiling. Blue, shaded squares within each plot represent an arbitrary PC weight (PC1-2) of less than 0.15. PC weights (PC1-2) ranged from 0 to 0.4573. (E) Violin plots of selected genes representing profiling of single HH 13 & 15 trailblazer NC cells.

![A description...](data:None;base64,)

**Figure 7 – Single trailblazer NC cells have a molecular profile that is distinct from the rest of the cranial NC stream.** (A) Isolation of single NC cells from each quartile of the cranial NC stream by manual dissection and FACS. (B) Intensity plot displaying the similarity of each cell’s 96-gene profile to all other single cell profiles as measured by Pearson’s correlation coefficient. (n=318 cells total; n=72 HH13 trailblazers, n=76 HH15 trailblazers, n=43 HH15 1st quartile, n=41 2nd quartile, n=44 3rd quartile & n=42 4th quartile) (C) Hierarchical clustering of single trailblazer NC cells at HH 13 & 15 as well as single NC cells from each quartile of the HH 15 cranial NC stream by Euclidean distance based upon averages of the 96-gene profiles. Genes are ordered by mean log2 expression in HH 15 trailblazer NC cells. (D) Hierarchical clustering of single trailblazer NC cells at HH 13 & 15 as well as single NC cells from each quartile of the HH 15 cranial NC stream by Pearson dissimilarity based upon averages of the 96-gene profiles. Genes are ordered by mean log2 expression in HH 15 trailblazer NC cells. (E) PCA of single trailblazer NC cells at HH 13 & 15 as well as single NC cells from each quartile of the cranial NC stream based upon a 96-gene profile. (F) Violin plots of selected genes representing profiling of single trailblazer NC cells at HH 13 & 15 as well as single NC cells from each quartile of the cranial NC stream.

![A description...](data:None;base64,)

**Figure 8 – Knockdown of key transcription factors perturbs cranial NC cell migration.** (A) Cranial NC stream electroporated with fluorescein-labeled control morpholino (green) and nuclear H2B mCherry (red) at HH St 15, n=8 embryos. (B) Cranial NC stream electroporated with fluorescein-labeled HAND2 morpholino (green) and nuclear H2B mCherry (red) at HH St 15, n=13 embryos. (C) Cranial NC stream electroporated with fluorescein-labeled TFAP2A morpholino (green) and nuclear H2B mCherry (red) at HH St 15, n=11 embryos. (D) Distribution of the percentage of morpholino transfected NC cells along the migratory route at HH St 13, n=8 embryos for control morpholino, n=6 embryos for HAND2 morpholino, n=8 embryos for TFAP2A morpholino. (E) Distribution of the percentage of morpholino transfected NC cells along the migration route at HH St 15. (F) Box plot of the distance migrated at HH St 13. (G) Box plot of the distance migrated at HH St 15. (H) Box plot of the area covered at HH St 13. (I) Box plot of the area covered at HH St 15. For box plots (F-I), the central mark is the median, the edges of the box represent the 25th and 75th percentiles, the whiskers reach the furthest data points not considered outliers, and outliers are plotted as crosses.

![A description...](data:None;base64,)

**Figure 9 – Upregulating leader behavior in trailing cell population disrupts model migration.** To represent an upregulation of a leader-like profile in the trailing portion of the stream, after time *tLF* every other cell inserted is a leader cell (orange) rather than a follower cell (purple) (A, mimicking 50% electroporation efficiency). Each stochastic simulation was repeated n=20 times (A), and the final cell positions summed up to construct the migratory stream densities (B) and cell count histograms (C).

![A description...](data:None;base64,)

**Figure 10 – Overexpression of HAND2 in trailing NC cells perturbs stream migration.** (A) Trailing NC cells transfected with control EGFP construct (green) and DiI (red) at HH St 15, n=20 embryos. (B) Trailing NC cells transfected with HAND2 FL construct (green) and DiI (red) at HH St 15, n=13 embryos. (C) Trailing NC cells transfected with TFAP2A FL construct (green) and DiI (red) at HH St 15, n=14 embryos. (D) Distribution of the percentage of transfected NC cells along the migratory route. (E) Distribution of the average number of transfected NC cells along the migratory route. (F) Raw data showing the positions of transfected NC cells oriented to neural tube as point of origin, n= 629 cells for control EGFP, n=226 cells for HAND2 FL, n=576 cells for TFAP2A.

![A description...](data:None;base64,)

**Figure 11 – Profile of trailblazer cranial NC cells**

![A description...](data:None;base64,)

**Table 1- Model parameters**

**Table 2 – Highest expressed genes in single HH 13 & 15 trailblazer NC cell profiling.** The average Ct value was calculated among HH13 & 15 single trailblazer NC cells that expressed each gene. The percentage of HH13 & 15 single trailblazer NC cells in which each gene was detected is also listed. Genes are ranked by level of expression (lowest to highest average Ct), with the maximal Ct value cut off set at 22. Bolded genes are those which are 1) detected in more than half of the single cells analyzed and 2) have an average Ct of less than 22 at both developmental time points.![A description...](data:None;base64,)

![A description...](data:None;base64,)

**Table 3 – Differential expression between trailblazers and quartiles at HH 15.** Averages from single trailblazer NC cells and single NC cells from the migratory stream were compared. The bolded genes are expressed at high levels by the trailblazers in more than half of the cells at both developmental time points (HH13 & 15). Thus, the trailblazer molecular signature of 16 genes was established.

![A description...](data:None;base64,)

**Table 4- Molecular profile after HAND2 or TFAP2A knockdown.** Highlighted genes are present in the trailblazer NC cell molecular signature, or are HAND2 or TFAP2A themselves.