Comparative analysis of gene expression profiles for several migrating cell types identifies cell migration regulators

Young-Kyung Bae, Frank Macabenta, Heather Leigh Curtis, Angelike Stathopoulos

PII: S0925-4773(16)30126-5
DOI: doi: 10.1016/j.mod.2017.04.004
Reference: MOD 3454

To appear in: *Mechanisms of Development*

Received date: 7 January 2017
Revised date: 13 April 2017
Accepted date: 13 April 2017

Please cite this article as: Young-Kyung Bae, Frank Macabenta, Heather Leigh Curtis, Angelike Stathopoulos , Comparative analysis of gene expression profiles for several migrating cell types identifies cell migration regulators. The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. Mod(2017), doi: 10.1016/j.mod.2017.04.004

Comparative analysis of gene expression profiles for several migrating cell types identifies cell migration regulators

Young-Kyung Bae[1,2], Frank Macabenta[1,†], Heather Leigh Curtis[1,†], and Angelike Stathopoulos[1,*]

[1] Division of Biology & Biological Engineering, California Institute of Technology, 1200 East California Blvd., Pasadena, CA 91125

[2] Korea Research Institute of Standards and Science, Center for Bio-Analysis, Yuseung-gu, Gajung-ro 267, Daejeon, Republic of Korea

† These authors contributed equally to this work.

*Corresponding author: angelike@caltech.edu

Keywords: Cell migration; *Drosophila melanogaster*: caudal visceral mesoderm; hemocytes; fluorescence activated cell sorting (FACS); Zfh1; Neyo; Singed; Border cells; Neural crest

ABSTRACT

Cell migration is an instrumental process that ensures cells are properly positioned to support the specification of distinct tissue types during development. To provide insight, we used fluorescence activated cell sorting (FACS) to isolate two migrating cell types from the *Drosophila* embryo: caudal visceral mesoderm (CVM) cells, precursors of longitudinal muscles of the gut, and hemocytes (HCs), the *Drosophila* equivalent of blood cells. ~350 genes were identified from each of the sorted samples using RNA-seq, and *in situ* hybridization was used to confirm expression within each cell type or, alternatively, within other interacting, co-sorted cell types. To start, the two gene expression profiling datasets were compared to identify cell migration regulators that are potentially generally-acting. 73 genes were present in both CVM cell and HC gene expression profiles, including the transcription factor Zinc finger homeodomain-1 (Zfh1). Comparisons with gene expression profiles of *Drosophila* border cells that migrate during oogenesis had a more limited overlap, with only the genes *neyo (neo)* and *singed (sn)* found to be expressed in border cells as well as CVM cells and HCs, respectively. *neo* encodes a protein with Zona pellucida domain linked to cell polarity, while *sn* encodes an actin binding protein. Tissue specific RNAi expression coupled with live *in vivo* imaging was used to confirm cell-autonomous roles for Zfh1 and Neo in supporting CVM cell migration, whereas previous studies had demonstrated a role for Sn in supporting HC migration. In addition, comparisons were made to migrating cells from vertebrates. Seven genes were found expressed by chick neural crest cells, CVM cells, and HCs including extracellular matrix (ECM) proteins and proteases. In summary, we show that genes shared in common between CVM cells, HCs, and other migrating cell types can help identify regulators of cell migration. Our analyses show that Neo in addition to Zfh1 and Sn studied previously impact cell migration. This study also suggests that modification of the extracellular milieu may be a fundamental requirement for cells that undergo cell streaming migratory behaviors.

INTRODUCTION

Cell migration is an essential cellular behavior associated with various biological processes including morphogenesis during development, wound healing, and immune response. Although cell migration supports important in vivo functions under normal conditions, it can become the basis for various pathological situations when dysregulated. For example, malignant cancer cells gain invasiveness by reprogramming their gene expression programs in order to migrate (Nguyen and Massagué, 2007). Thus, mechanisms governing cell migration have been of general interest in biology and pathobiology. Studies of how cell migration acts to support development have highlighted the importance of dynamic

regulation of this process. Migrating cells within a developing animal encounter dynamic microenvironments, including changing interactions with different cell types as development proceeds. In the *Drosophila melanogaster* embryo, there are many migrating cell types, known to exhibit migration patterns that differ in timing, path, and cellular organization. Studying how cell types accomplish these migrations during normal development may provide insights into how to regulate aberrant migration associated with disease states.

The caudal visceral mesoderm (CVM) cells originate from the posterior end of the ventral mesoderm and their anterior movement persists approximately six hours, encompassing the migration of longest duration in *Drosophila* embryogenesis (Kusch and Reuter, 1999). CVM cells initiate anterior movement at the completion of germ band elongation, at late stage 10. They begin by separating into two bilateral groups, which maintain this symmetry and move synchronously throughout the course of their migration. As a result, CVM cells become aligned along the transverse visceral mesoderm (TVM) cells. The CVM and TVM cell types subsequently specify longitudinal and circular muscles, respectively, which ensheath the midgut. In addition, fibroblast growth factor (FGF) signaling through the Heartless (Htl) FGF receptor is required for both CVM cell survival and directional CVM cell movement (Kadam et al., 2012; Reim et al., 2012). However, CVM cell migration initiates even in the absence of FGF signaling suggesting that additional guidance mechanisms are acting. CVM cells migrate as a streaming collective, exhibiting analogous behavior to the vertebrate neural crest cells, which is distinct from other migration models studied in *Drosophila.* Therefore, novel insights into the regulation of cell migration collectives may be uncovered from understanding this particular migratory mode.

Hemocytes (HCs) are another embryonic cell type exhibiting directional migration in *Drosophila* (Evans and Wood, 2011). The HC cell population, which is composed of 95% plasmatocytes and 5% crystal cells, are blood cells that engage in phagocytosis and melanization in response to wounding (Honti et al., 2014). In the early embryo, HC precursors are present within two bilateral clusters that originate from the ventral head mesoderm. From this anterior position within the head, subsets of HCs initiate a posterior migration along the ventral midline at stage 11. This migration is dependent on the proper development of the ventral nerve cord, which provides survival and guidance cues such as PDGF/VEGF receptor signaling (PVR in *Drosophila*) to pave a physical pathway upon which the HCs crawl (Brückner et al., 2004; Evans et al., 2010). The remaining HCs initiate migration from the head mesoderm moving toward the ventral midline located at the posterior of the embryo. These HCs exhibit an invasive transepithelial migration, translocating themselves in a Rho-L-dependent manner to the posterior end of the embryo, which as a result of germ-band extension, is located at the top of the embryo (Siekhaus et al., 2010). HCs align along the ventral midline, spanning the length of the embryo, and then disperse laterally

to populate the entire embryo. A number of genes supporting HC migration have been identified including PVR, Rho GTPase, Rac GTPases, and LanB1 (laminin B1) (Evans and Wood, 2011). This developmental HC migration utilizes a mechanism distinct from the movement induced by immune response (Wood et al., 2006). However, their proper developmental dispersal is prerequisite for their later immune functions. HCs migrate throughout the embryo, and while their movement is directed the cells move as individuals, presumably, independent of their cohorts (Pocha and Montell, 2014).

To provide additional insight into the shared and distinct molecular mechanisms relating to cell migration of CVM cells and HCs, a comparative gene expression profiling approach was used. Both migrations involve loose streams of cells that appear to be controlled in their movement, and yet the mechanisms guiding these directional migrations remain unclear. Gene expression profiling has proven very useful for understanding other cell migration systems. We focus on three studies here: two of *Drosophila* border cells (BCs) (Borghese et al., 2006; Wang et al., 2006) and one of neural crest cells from chick (Simões-Costa et al., 2014). In the *Drosophila* ovary, the BCs exhibit a collective migratory behavior, which serves as a model for studies of invasive cell migration (Montell, 2001; Rørth, 2002). BCs delaminate from the follicular epithelium, migrate through the germline tissue, and then eventually move towards the oocyte while maintaining the tight cell collective (Montell, 2001; Rørth, 2002). Their migration is important for anterior-posterior patterning in the oocyte as well as morphogenesis of the micropyle, a structure through which sperm enters the egg (Savant-Bhonsale and Montell, 1993). In vertebrates, neural crest cells undergo a long-range migratory behavior to specify a variety of cell types including cartilage, enteric neurons, and melanocytes along the length of the embryo from cranium to trunk (Theveneau and Mayor, 2012). Neural crest cells emigrate from the tube after a regulated epithelial to mesenchymal transition (EMT), and many studies have focused on understanding how these cells interact to support their dispersal throughout the body (Theveneau and Mayor, 2012). This includes differences in gene expression programs associated with anterior-posterior position. Gene expression profiling experiments have successfully provided insights into the factors supporting migration in the specific cell types studied. However, due to the lengthy list of genes produced and because it is not practical to test all enriched genes for a role in supporting migratory behavior, we hypothesized that additional insight might be more easily achieved from comparative studies of gene expression profiles for diverse migratory cell types, within and across species, which could highlight conserved genes supporting the migratory behavior.

Here, we investigated two migrating cell populations in the *Drosophila* embryo, CVM cells and HCs, using gene expression profiling to provide insights into these cells' identity, function and behavior, especially with regard to mechanisms supporting cell migration. CVM cells move as a streaming

collective, similar to the vertebrate neural crest cells; in comparison, HCs appear to move more independently, as individuals, despite also being directed in their movement. We isolated cell populations via fluorescence activated cell sorting (FACS), and the RNA extracted was processed by RNA-Seq to reveal potential gene expression programs. Genes expressed at higher levels within sorted samples compared to the unlabeled cells were presumed to be expressed specifically in tissues of interest, and this was confirmed for a subset of genes using *in situ* hybridization. In addition, the role of several genes expressed within the CVM was assessed using tissue-specific RNAi. To identify genes that are broadly required for migration, we performed a comparative survey of gene expression profiles associated with migrating cell types from *Drosophila* or vertebrate neural crest. This approach led to the identification of putative general regulators of cell migration, genes playing a role in the regulation of multiple cell types.

RESULTS & DISCUSSION

Transcriptome analysis of CVM cells and hemocytes

In *Drosophila*, distinct cell populations actively migrate within the developing embryo. We focused on two migrating cell types: caudal visceral mesoderm (CVM) cells and hemocytes (HCs) as their migrations were found to overlap spatially and temporally, at least in part, and these cells both exhibit a cell streaming behavior (Figure 1A-A''). Originating from the caudal mesoderm located at the posterior of the embryo, CVM cells initiate movement at the end of germ band elongation as two bilaterally symmetric groups, with only one group apparent in lateral view. These cells move in a directed fashion toward the anterior of the embryo from stage 11 (RFP, Figure 1A), hugging the transverse visceral mesoderm (TVM) as they continue anteriorly through stage 12 (RFP, Figure 1A'). By stage 13, the CVM cells, which are muscle founder cells, are positioned along the entire length of the TVM allowing them to fuse with fusion competent muscle cells and, subsequently, to spread to ensheath the entire developing midgut (RFP, Figure 1A"). On the other hand, HCs are first specified in the head mesoderm and initiate their migration out of the head region at stage 11 (GFP, Figure 1A) (Evans and Wood, 2011). HCs exhibit either a posterior movement along the ventral midline or a dorsal transepithelial invasion into the extended germ band at stage 12 (GFP, Figure 1A'). Later, HCs migrate laterally from the midline at stage 13/14 as they disperse throughout the embryo (GFP, Figure 1A"). These observations demonstrate that the movements of CVM cells and HCs significantly overlap both temporally and spatially.

To gain insight into the molecular mechanisms directing active cell migration, we hypothesized that cells' gene expression profile may be reflective of their behaviors. In order to perform transcriptome

analyses, CVM cells and HCs were isolated from embryos during their migration (Figure 1B). Each cell type was detected using fluorescent reporters: HLH54F-Gap-Venus (GV2) (Stepanik et al., 2016) (Sup. Figure 1) for CVM cells and Srp-Gal4>UAS-Gap-Venus (Figure 6) for HCs. The Srp-Gal4 driver is specific to embryonic hemocytes (plasmatocytes and crystal cells)(Brückner et al., 2004). Cells were isolated from staged embryos encompassing each cell type's migratory process: stages 10 to stage 13 to detect CVM cells, and stages 12 to stage 16 to detect HCs. Though the stages of CVM cell and HC active migration partially overlap, cells were isolated from two different embryos collections to represent the full temporal window of active migration and sorting performed as two separate experiments. In addition, cells from control embryos (i.e. *yw*) were used to set the baseline to identify the fluorescent marker positive population within each experiment. The FACS scatter plots represent the total number of cells from embryos at different stages (heterogeneous pool; Figure 1C-D, 1F-G). However, the Venus-positive populations were distinctly present in each experiment (Figure 1D and 1G) albeit at different proportions (CVM: 1.59%, HC: 9.68%) representing the relative abundance of CVM cells and HCs within the embryo.

To identify differentially expressed genes in each migratory cell type, we first performed an RNA-seq analysis of samples isolated from Venus marker-positive and marker-negative populations for each experiment (Figure 1B). Sequenced reads were mapped to the *Drosophila melanogaster* reference genome using TopHat (Trapnell et al., 2009), and the list of differentially expressed genes was obtained using CuffDiff (Trapnell et al., 2013).

For the CVM transcriptome analysis, differentially expressed genes were identified by comparing RNA-seq reads from the HLH54F-Gap-Venus experiment, sorting Venus marker-positive and marker-negative populations. This transcriptome analysis revealed 825 enriched and 73 downregulated genes in the marker positive (i.e. CVM enriched) sample compared to the marker negative sample (i.e. other cell types) (Figure 1E, red and blue dots, respectively). Several genes known to be expressed in the CVM such as *HLH54F* (Ismat et al., 2010), *heartless* (*htl*) (Kadam et al., 2012; Mandal et al., 2004), and *beat IIa* (Ismat et al., 2010) were significantly enriched in the CVM sample (Figure 1E and 1E', green). The fold-change associated with the *htl* gene is lower than other CVM-specific genes and likely relates to its broad expression at the targeted stages (i.e. expression in tissues besides CVM; see *htl in situ* in Figure 3). These results demonstrate the validity of our analysis scheme and highlighted the potential for discovering novel CVM genes.

The HC transcriptome analysis identified 596 enriched and 689 downregulated genes from the Srp-Gal4>UAS-Gap-Venus experiment, sorting Venus marker-positive (i.e. HC-enriched) and marker-negative (i.e. other cell types) populations (Figure 1H, red and blue dots, respectively). Genes known to

be expressed in the embryonic hemocytes (such as *proPO-A1, ppn, Cg25C, Vkg, srp*) are ranked high in the HC list (Figure 1H and 1H', green), supporting the validity of the approach in identifying genes expressed in HCs as well as the ability to obtain distinct, relevant gene expression profiles for distinct cell types (i.e. CVM vs. HCs). Analogous to the case of the *htl* gene in CVM cells, *srp* gene is only moderately enriched within hemocytes as its expression is not hemocyte-exclusive but also found in the fat body during HC migration (Abel et al., 1993). Therefore, this list contains both genes expressed exclusively within HCs as well as more broadly expressed gene expression profiles that include expression in other tissues in addition to HCs.

Genes enriched within CVM cells

In order to focus on transcripts specifically expressed in the CVM cell type for further analysis, we narrowed down the initially defined CVM-enriched transcriptome from 825 genes to 324 genes using the following criteria: selection for RPKM (reads per kilobase per million reads) on the CVM marker positive sample of greater than 10 counts, and fold-change relative to wildtype of greater than 2.8 (see Sup. Table 1 for the complete list). Among these, select genes exhibit specific expression within the CVM, confirmed by conducting *in situ* hybridization using riboprobes (this study), through the BDGP (Berkeley Drosophila Genome Project) database (Tomancak et al., 2002; Tomancak et al., 2007), or by previous studies (Table 1). The top-listed, CVM-enriched gene, *HLH54F*, is a basic helix-loop-helix transcription factor that is required for specification and migration of CVM cells (Ismat et al., 2010). Kon (Kon-tiki or Perdido) and Grip (DGrip) cooperate to regulate muscle tube adhesion and fusion, and the expression of *kon* in CVM cells has been previously noted (Estrada et al., 2007; Schnorrer et al., 2007). While our analysis identified many genes that were previously reported, thus corroborating published research and lending credence to our methodology, we also identified a number of novel genes that present new avenues for further investigation.

To identify overrepresented molecular and biological processes in the CVM-enriched gene list, we performed a GO term analysis using DAVID (Database for Annotation, Visualization and Integrated Discovery v 6.8) (Dennis et al., 2003). With the short CVM list of 324 genes, three GO terms (Biological Process, Molecular Process, and Cellular Components) were independently searched for functional annotations associated with an over-representation of genes of this classification within the listed CVM-enriched gene lists. For each category, the ten most significantly enriched groups were selected (see Figure 2A with corresponding *p*-values). Among biological process terms assayed, 'muscle organ development' has the lowest *p*-value (highest score for enrichment) and is associated with 12 genes

(Figure 2C). Among these, the involvement of Mef2 (myocyte enhancer binding factor-2) and Htl in visceral muscle formation has been reported (Lilly et al., 1995), and a potential role for Kon at earlier stages in the development of these muscles, in the migration of CVM cells, has been proposed (Trisnadi and Stathopoulos, 2014). However, the roles of the other 9 'muscle organ development' genes in supporting formation/function of this particular muscle type have not been studied. Moreover, three other biological process terms also are relevant to muscle structure and function: 'myofibril assembly', 'skeletal muscle tissue development', and 'sarcomere organization'. Thus, muscle-associated activities comprise 4 out of the 10 most significant biological process categories. This trend is consistent even when all 3 GO terms assayed (i.e. Biological Process, Molecular Process, and Cellular Components) are considered, as about half of all the significant GO term categories identified relate to muscle (Figure 2A, *). This result illustrates that a significant number of genes that are enriched in CVM cells relate to CVM cells' ultimate function as longitudinal visceral muscle.

The GO term with the highest enrichment score (lowest *p*-value), out of the 30 identified, is 'integral component of plasma membrane' (Figure 2A, **). It represents 33 genes present in the CVM-enriched set (~10 % of the entire list). This set of genes was further categorized into several functional subclasses using information on molecular function obtained from Flybase (Figure 2B, Sup. Table 2 for list of genes). The largest subclass 'major facilitator superfamily domain protein' contains eight genes that encode members of an evolutionarily conserved family of proteins responsible for transport of solutes (sugars and ions) across the plasma membrane (Pao et al., 1998). Other subclasses of relevance to migrating cell types are 'cell adhesion' (4 genes, 12 %) and 'receptor' (2 genes, 6 %), as such genes could play a direct role in regulating CVM cell migration.

Transcription factors comprise another class of proteins of interest, which likely act to help define the CVM cell type and regulate expression of genes that support cell migration and differentiation. Transcripts for 17 transcription factors were enriched in the CVM RNA samples. Previous studies have shown that Doc2, HLH54F, and Bin function to support CVM specification (Ismat et al., 2010; Zaffran et al., 2001), whereas Mef2 is more generally required for muscle formation (Lilly et al., 1995). It is possible that the 13 other transcription factors identified cooperatively regulate expression of genes in the CVM together with previously characterized Doc2, HLH54F, Bin, and Mef2 transcription factors, which are required for CVM specification as well as migration.

CVM-specific gene expression patterns highlight genes involved in cell adhesion and proteolysis

To test if genes in the CVM list are specifically expressed in CVM cells during their migration, we first examined the expression patterns associated with 15 genes in control (i.e. *yw*) embryos by *in situ*

hybridization using riboprobes to respective genes (Figure 3). In addition to confirming seven known CVM genes (such as *Doc2, HLH54F, kon, htl*), we also identified nine novel CVM-enriched genes including *CG9416* (endoplasmic reticulum metallopeptidase 1), *CG5080* (plasma membrane protein with unknown function), *CG17124* (protein phosphatase 1 regulatory inhibitor subunit 14B), *beat IIb* (immunoglobulin domain protein, see below), *Grip* (*Drosophila* glutamate receptor binding protein), *Syn2* (dystrophin glycoprotein), *taspase 1* (endopeptidase), *tey* (E3 ubiquitin-protein ligase domain containing protein), and *tok* (protein with peptidase domain) (Figure 3). We confirmed these transcripts are specifically present in CVM cells throughout their migratory stages: stage 10 as two lateral groups, stage 11-12 as the two groups exhibit collective migration, and stage 13 as the cells prepare to fuse with the circular visceral muscle cells (Figure 3A). Their CVM-specific expression was further confirmed by loss of expression in the *HLH54F* mutant embryos, as CVM cells are absent in this genetic background (Figure 3B). Additionally, we investigated if any of these are FGF-dependent genes by examining their expression in the *htl* FGF receptor mutant. Expression of all the genes tested persists in the *htl* mutant background (Figure 3C, compare with 3A); albeit the CVM cells' arrangement is abnormal, exhibiting the *htl* mutant phenotype (Kadam et al., 2012). Therefore, it seems that none of the genes tested are FGF targets, suggesting that FGF signaling does not involve batch activation of target genes within the CVM but instead may direct CVM's cellular behavior in other ways beyond control of transcriptional programs.

Regulation of cell adhesion is an important aspect of migrating cells. *beat IIa* and another CVM enriched gene, *beat IIb*, are members of the *beaten path* gene family, encoding immunoglobulin superfamily domain proteins that regulate cell adhesion (Pipes et al., 2001). The related *beat Ia*, along with its binding partner *sidestep*, promote motor neuron fasciculation that is required to innervate the body wall muscles during *Drosophila* development (Siebert et al., 2009; Zinn, 2009). Other members of the *beaten path* family appear to be expressed in neurons as well (Pipes et al., 2001), and may also serve roles in neuronal guidance. We confirmed that both *beat IIa* and *beat IIb* are expressed in the non-neuronal, yet migrating, CVM cells (Figure 3). Future studies testing their roles in cell migration/adhesion and to identify their binding partners, likely present on interacting tissues or within the extracellular matrix (ECM), are warranted.

We also verified the expression of genes encoding several proteases in migrating CVM cells, including *taspase 1, tok,* and *CG9416.* Proteolysis is an essential process in biology with roles as diverse as protein turnover, signaling regulation, and ECM modification; moreover, it is also the basis of various pathologies including cancer. One of the highly conserved proteases Threonine Aspartase 1 (Taspase 1) is overexpressed in multiple cancer conditions (Stauber et al., 2016). As current knowledge regarding the substrate specificity and regulatory mechanism of Taspase 1 is limited, CVM cell migration can

potentially serve as a valuable model system to study this gene. *tok* and *CG9416* each encode proteins with peptidase domains, but their roles during *Drosophila* embryogenesis and cell migration remain undefined.

CVM gene list includes those highly expressed in neighboring tissues

Cells migrating through the developing embryo encounter a number of different tissues and these physical interactions influence cell behavior (Evans et al., 2010; Gline et al., 2015; Stepanik et al., 2016). Although we have identified novel CVM specific genes from our CVM transcriptome analysis, we also identified a significant number of genes that are expressed within other tissues in the vicinity of the CVM (Figure 4). Even in the shortened CVM list, we have identified genes that are predominantly expressed in the developing gut, TVM, yolk and germ cells (Figure 4). We cannot formally exclude the possibility that these genes are also expressed at low levels in CVM cells, especially for TVM-enriched genes as these two tissues are closely associated. Alternatively, we reason that there are at least two other possible sources from which these transcripts are isolated. The first possibility is that the CVM neighboring cells were co-sorted (i.e. piggy backed) with CVM cells expressing the reporter during cell sorting. The sorting protocol used could have permitted acquisition of cell-cell doublets (co-associated cells), which potentially exposed meaningful tissue interactions that are important for CVM migration. The other possible reason is that the CVM reporter GV2 used for CVM sorting is predominantly, but not exclusively, CVM-specific. However, we have not observed GV2 expression in the four tissues shown here nor any other non-CVM tissue (Figure 3; Sup. Figure 1). Therefore, the genes identified in the screen that are not expressed in the CVM are likely to be due to a failure to completely separate the CVM cells from other cells in the embryo, especially since the experiment was only done a single time. Nevertheless, isolation of genes in this concentration of tissues gives us potential insight into interacting cell types, which is useful information.

Genes enriched within hemocytes include regulators of the immune response

For consistency, the short list of highly enriched HC genes (386 genes) was generated using the same criteria used for the CVM analysis (i.e. HC marker positive RPKM greater than 10 and fold-change greater than 2.8) (Table 2; Sup. Table 3 for the complete list). We used this short list to produce the top listed genes with confirmed HC expression and GO term clustering (see below). When these data related to the HC marker positive sample are sorted by RPKM (Table 2), the top listed genes *NimC4* and *PPO2*

(*Prophenoloxidase*) are those known to be expressed in the embryonic HCs (BDGP, Figure 6) and crystal cells, a small subset of cells within the HC population (BDGP). Both *NimC4* and *PPO2* are involved in the immune response as they encode a putative phagocytosis receptor (Kurucz et al., 2007) and melanization enzyme (Binggeli et al., 2014), respectively, suggesting that our HC transcriptome is reflective of these cells' specific cellular function.

We also ran GO term enrichment analysis using DAVID on the HC gene expression results, as described above for the CVM list, to identify functional annotations that are overrepresented in the HC-enriched gene list. Again, the GO term enrichment for three categories (biological process, molecular process, and cellular component) was individually tested, this time using the short list of HC-enriched genes, and the resulting 10 most-significantly overrepresented terms for each category are shown (Figure 5A). Multiple GO terms reflecting HC function and specification appeared in this list such as 'phagocytosis', 'hemocyte development', 'wound healing', and 'apoptotic cell clearance' (Figure 5A, *). At minimum, *pvr* and *RhoL* (2 out of 4), genes present in the 'hemocyte development' subclass (Figure 5C), are known to directly influence HC migration (Brückner et al., 2004; Siekhaus et al., 2010). Additionally, *Drosophila* HCs are known to produce reactive oxidant intermediates (ROI) and mediators for melanogenic encapsulation in response to infection (Nappi and Vass, 1998). Genes relating to these specialized cellular functions are present also in the HC-enriched gene list. For example, genes involved in 'oxidation-reduction process', 'hydrogen peroxide catabolic process', 'glutathione peroxidase activity', and 'peroxiredoxin activity' are most prevalent (Figure 5A, *). In addition, functional terms relating, possibly, to the migratory behavior of HCs include genes present in 'extracellular matrix organization' and 'metallopeptidase activity' categories (for gene lists, see Figure 5C).

Similar to the CVM list, the 'plasma membrane' cellular component GO term category is associated with the score of highest significance (lowest *p*-value) in reference to the HC gene list. The 51 genes within this 'plasma membrane' category designation were further categorized into several functional subclasses (Figure 5B, Sup. Table 4 for the gene list). The two largest subclasses are relevant to 'transmembrane transport' (6 genes, 13%) and 'ABC transporters' (5 genes, 10%). The subclass of next highest ranking is related to adhesion (5 genes, 10%). These results suggest that regulation of cell transport and adhesion may play a role in HC function and behavior.

To confirm the specific expression of HC genes identified, we performed *in situ* hybridization on control (i.e. *yw*) embryos (Figure 6). Many of these genes have been shown to be present in HCs with varying specificity. Expression is found within the head mesoderm at stage 10 and then observed broadly within the embryo as HCs disperse. For example, *ilp4* (*insulin-like peptide 4*) is expressed in HCs (most evident at stage 13) but its earlier expression is broader, expanded beyond HCs and including head

mesoderm (hemocyte precursors) as well as the trunk mesoderm. On the other hand, we confirmed that expression of *NimC4*, *su(r)*, *CG6310*, and *CG8501* is clearly present in HCs (Figure 6). *NimC4*, which encodes a transmembrane phagocytic receptor, is a member of the highly conserved Nimrod gene cluster (Kurucz et al.). Its expression in embryonic HCs suggests that phagocytosis in these cells is mediated by Nimrods at the embryonic stages. *su(r)* (*suppressor of rudimentary*) encodes a protein involved in the oxidation-reduction process, which is relevant to HCs' function as immune cells. *CG6310* and *CG8501* encode proteins of unknown function, however their specific expression within HCs suggests that they may act to support HC differentiation and/or migration.

Comparisons between CVM and HC enriched genes

We hypothesized that distinct migratory cell populations may express similar genes to support their migratory behavior. Genes shared in common are more likely to support shared behavioral properties rather than cell type specific functions. When the short lists for CVM- (324 genes) and HC- (386 genes) enriched gene sets were compared, a total of 73 genes were found in common (Sup. Table 5 for the gene list). To identify overrepresented functional annotations, we performed a clustering analysis using DAVID program's default IDs that include GO-term biological process, molecular process, cellular components, and protein domain databases. This 'clustering' maximizes the pool of information from our query of relatively small number (n=73), while minimizing the intrinsic redundancy of individual GO terms.

The most significantly enriched functional cluster is 'transmembrane transporter' comprised of eight genes (Sup. Figure. 2, Sup. Table 5), which play an important role in cellular metabolism and cell-environment communication by transporting essential nutrients, ions, and other metabolic products across the plasma membrane. This cluster includes specific GO terms such as 'monocarboxylic acid transporter (MCT)' and 'major facilitator superfamily'. MCT acts on the plasma membrane to transport monocarboxylates such as L-lactate, pyruvate, and the ketone bodies (Halestrap and Wilson, 2012). Members of MCT family have been implicated in epithelial cell migration via its interaction with beta1-integrin (Gallagher et al., 2009) and lung cancer cell invasion (Izumi et al., 2011). The major facilitator superfamily is an evolutionarily conserved family responsible for inward and outward transport of solutes (Pao et al., 1998). This result suggests that genes involved in active cell metabolism and both cell-autonomous and non cell-autonomous interactions may contribute to the control of migratory behavior for CVM cells and HCs.

Of the genes common to the CVM and HC datasets, Zfh1, a member of the zinc-finger E-box-binding family transcription factor, has been linked to the regulation of cell migration in several contexts. Its mammalian homolog, Zeb2/Sip1, has been shown to regulate EMT and cell motility (Vandewalle et al., 2009). Interestingly, both *Drosophila zfh1* and its mammalian homolog *SIP1* are required for proper development of the cardiogenic lineage; so highly conserved is this function that expression of mouse *SIP1* in *Drosophila zfh1* null mutants that it completely rescues heart development phenotypes (Liu et al., 2006). In *Drosophila*, *zfh1* is additionally required for specification of lateral mesodermal lineages, including somatic gonadal mesoderm to which germ cells are attracted (Broihier et al., 1998) and also for short distance migration of neuronal cells (Bhat, 2007). Furthermore, it also has been reported that CVM migration in the *zfh1* mutant embryos is disrupted (Broihier et al., 1998). Although *zfh1*'s expression in embryonic hemocytes has been noted (Broihier et al., 1998), its role in HC migration remains to be investigated.

Comparisons with border cell gene expression programs

We next searched for gene expression profiles from other migrating cell populations in *Drosophila* other than CVM cells and HC. In the ovary, a group of ~10 cells, the border cells (BCs), migrates as a tight collective within the developing oocyte. We reasoned that comparing genes enriched in these three cell groups (i.e. CVM cell, HCs, and BCs) which engage in different, developmental migratory behaviors in *Drosophila* may provide insights to common mechanisms, if any, governing the cell migration process. We utilized already existing BC gene expression profiles from two independent sources [Lists BCborghese (Borghese et al., 2006) and BCwang (Wang et al., 2006)] and compared these with our CVM cell and HC datasets based on their Flybase IDs (Figure 7A). The BCwang list included 413 genes that are enriched in the migratory BCs (fold change: 1.02 to 16.29); whereas the BCborghese list contained 392 genes that are significant enriched in the wildtype BCs compared to follicle cells (fold change: 1.3 to 266). The overlap between the two BC lists was limited encompassing only ~10% (i.e. 37 genes) (Figure 7A). We searched the Dresden Ovary Table (DOT) database for gene expression patterns of the shared genes within *Drosophila* ovaries (Jambor et al., 2015) and found *in situ* hybridization images for 20 out of 37 genes, with only six that had been annotated with BC expression (i.e. *tsl, cln3, Argk, B4, neo, sn*). The limited overlap of these two BC datasets likely reflects the different experimental approaches, and also suggests that genes found in common in comparisons of multiple, large datasets may help define key players.

When we compared the two BC lists with our CVM and HC short lists, no gene was common to all four lists. Nevertheless, two genes were found for each intersection of the following comparisons: *neyo* (*neo*) was common to CVM + BCborghese + BCwang datasets, and *singed* (*sn*, also known as *fascin*) was common to HC + BCborghese + BCwang datasets (Figure 7B).

neo encodes a protein with Zona pellucida domain (ZPD) that is involved in the regulation of cell polarity (Fernandes et al., 2010). In *Drosophila* embryos, *neo* is weakly expressed in CVM cells during their migration (Figure 8F, arrowheads) in addition to more prominent expression in a posteriorly-localized tissue, possibly within the hindgut proper and anal pad precursor cells (Figure 8F; BDGP). We also examined *neo*'s expression in *HLH54F* mutants, which confirmed the weak expression was indeed CVM-specific as this signal was absent in mutant embryos (Sup. Figure 3). Physically-associated with the plasma membrane, Neo and other ZPD proteins cooperatively influence epidermal cell size and shape as they remodel cellular interaction with local ECM environment on the apical compartment (Fernandes et al., 2010). In *Drosophila*, distinct ZPD proteins support wing and tracheal development (Jaźwińska et al., 2003; Roch et al., 2003). However, a role for Neo in supporting cell movement has not been reported previously (see below).

On the other hand, Sn is known to be required for hemocyte migration via its function in actin filament organization and cell polarization (Zanet et al., 2009). However, the role for Sn in border cell migration remains rather tenuous. In a previous study, *sn* mutant border cell clones did not exhibit migration defects (Borghese et al., 2006). However, as the *sn* mutant used is a hypomorph, this leaves open the possibility that a null mutant could affect BC migration.

zfh1 and neo act cell-autonomously to support CVM migration

Next we asked whether the identified genes shared between migration systems regulate CVM cell migration. We first examined the CVM migration phenotype in fixed embryos expressing 5053-Gal4>UAS-RNAi for *zfh1* and *neo* (Figure 8B and E). In control embryos, the CVM membrane marker (GV2) labels two separate, bilateral CVM cell groups when visualized from the dorsal side at stage 11-12 (Figure 8A and D). In contrast, embryos expressing *zfh1* RNAi in the CVM exhibit CVM migration defects such as asynchronous migration (Figure 8B, arrowheads) and meandering of aggregated cells towards the midline (Figure 8B). Similar migration defects were observed when *neo* RNAi is expressed in the CVM, resulting in abnormal cell migration (Figure 8E, arrowhead).

To further elucidate the individual requirements of Zfh1 and Neo in mediating CVM cell migration, we performed live *in vivo* imaging of embryos expressing a CVM nuclear reporter (HC2; Kadam et al., 2012) as well as the aforementioned RNAi constructs specifically in the CVM via the 5053-GAL4 driver. In a control embryo expressing HC2, two lateral groups of CVM cells migrate forward synchronously from a dorsal view (Sup Movie 1). Consistent with the findings from fixed embryos (Figure 8B and E), the CVM-specific expression of either RNAi lines induced severe migratory defects, such as temporal delay and loss of synchrony between the two lateral groups. Specifically, in the *neo* RNAi expressing embryo, CVM cells exhibit seemingly dysregulated cell division as they stall and divide at an earlier time point compared to control (Sup Movie 2). In contrast, we did not observe a similar stalling and cell division defect in embryos expressing *zfh1* RNAi, and instead observed a general loss of directionality, as some cells within each migrating cohort wandered laterally and towards the midline (Sup Movie 3). We conclude that *neo* and *zfh1* are required for proper CVM migration and act in cell-autonomous manners, possibly by modulating cell-environment interactions and expression of genes related to EMT, respectively. For example, in other contexts, the vertebrate homolog of *zfh1* has been shown to negatively regulate cadE expression to influence EMT (Rogers et al., 2013).

The identification of genes expressed in two or more tissues brings about the intriguing possibility that we have identified a few genes that function as general regulators of cell migration. For instance, *neo* was found to be expressed in both CVM and border cells (BCs) (Sup. Figure 3), a group of 6-8 migratory cells found in individual egg chambers within a mature ovary. BCs are a well-studied model for collective cell migration. It has been found that their distinctive migratory behavior requires both epidermal growth factor receptor (EGFR) and JAK/STAT signaling (Beccari et al., 2002; Duchek and Rørth, 2001; Silver and Montell, 2001), while polarized leader-follower interactions within a migrating cluster are mediated by differential activity of endocytic Rab proteins (Assaker et al., 2010). In order to determine whether *neo* also plays a role in supporting BC migration, we used a BC-specific Gal4 (i.e.*slbo*-GAL4) to drive expression of either *neo* RNAi or controls within BCs (Figure 8G; Sup. Figure 4). In the negative and unexpressed controls (lacZ and *zfh1* RNAi) the BCs successfully migrate through the nurse cells to the periphery of the developing oocyte in stage 10 egg chambers. In contrast, egg chambers expressing *neo* RNAi in the BCs often failed to complete migration while others had serious to mild delays. This phenotype is statistically significant, yet less severe than other RNAi phenotypes for other genes [e.g. *brk*; (Luo et al., 2015)]. These results suggest that *neo* plays a general, previously uncharacterized role in supporting migration in two distinct cell types; further characterization of the similarities and differences between the mechanisms involving *neo* in CVM and BCs would be an interesting avenue for future study.

Comparisons with the vertebrate neural crest cells highlight ECM-related genes

We also extended our comparative analysis to include a vertebrate migratory cell population, the chick neural crest cells. First, the list of genes enriched (fold change > 2) in migrating neural crest cells was produced by a previous study (Simões-Costa et al., 2014). We converted the chick ensembl gene IDs associated with each gene in this list to the respective Flybase gene IDs by searching for closest *Drosophila melanogaster* orthologs using Ensembl Biomart. In total 1,031 chick genes were queried and yielded 1,019 genes with Flybase IDs. This converted list comprehensively covers corresponding *Drosophila* orthologs as one chick gene occasionally yielded multiple, related fly genes, maximizing the capacity of our comparative analysis. This resulting list of *Drosophila* orthologs for chick neural crest genes was used to query overlap with our CVM cell and HC gene expression profiling datasets (Figure 7C). Seven genes were found to comprise the overlap between these three gene expression profiles (i.e. neural crest + CVM + HC lists) (Figure 7C and 7D). These seven genes related to three functional classes either ECM-related (n=4), oxidation-reduction process (n=2), or amino acid metabolism (n=1) (Figure 7D).

The ECM-related genes are most common as they include four out of the seven (Mmp2, Crag, Vkg, Spn27A). These genes encode either modulators or components of the extracellular environment (or basement membrane). Mmp 2 is one of two secreted matrix metalloproteases (MMPs) in the *Drosophila* genome and has been shown to play an essential role in motor axon fasciculation in fly embryos (Miller et al., 2008). Another *Drosophila* MMP, Mmp1, promotes wound healing by permitting cell migration via ECM repair (Stevens and Page-McCaw, 2012), suggesting that Mmp2 similarly may act as a positive regulator of cell migration. In vertebrate embryos, MMPs (MMP-2 and Mmp17b) are important for neural crest cell migration (Duong and Erickson, 2004; Leigh et al., 2013). *vkg* (collagen IV) is known to be secreted from HCs which in turn promotes renal tubule morphogenesis in *Drosophila* embryos (Bunt et al., 2010). Vkg is also likely to be pro-migratory factor of other cell types in the vicinity of renal tubules, as well, as it is a major constituent of the ECM. Crag (calmodulin-binding protein related to a Rab3 GDP/GTP exchange protein) regulates epithelial cells' polarized secretion of basement membrane molecules including Collagen IV, Laminin and Perlecan (Denef et al., 2008). Spn27A (serine-type endopeptidase inhibitor) encodes a highly conserved secreted factor that is known to function in the dorsoventral patterning of the *Drosophila* embryo by antagonizing Toll signaling (Hashimoto et al., 2003). In sum, four out of the seven genes, shared in expression by neural crest + CVM + HCs, function in the extracellular space, and this result highlights the importance of cell-matrix interactions and regulation during cell migration.

Two other genes uncovered, *CG8080* and *CG9629*, relate to the oxidation-reduction process, but their link to migratory cell types is unclear (Figure 7C). However, CG8080 shows homology to NAD (Nicotinamide Adenine Dinucleotide) kinase, which produces NADP+ that can be reduced to NADPH (Tedeschi et al., 2016), and NADPH mediates various metabolic and biosynthetic processes including neutralizing ROS species. Whereas Oat (Ornithine aminotransferase precursor) is involved in ornithine metabolic process, which has been implicated in cell migration. Ornithine, an amino acid, is found only rarely in proteins but is important in living organisms as an intermediate in the urea cycle and arginine biosynthesis. L-Arginine and its metabolism by OAT activates intestinal cell migration and epithelial cell wound repair, respectively (Rhoads et al., 2004). Whether the ornithine metabolic process via Oat influences migration of CVM cells and hemocytes is unknown.

Ten additional genes were absent from the HC list but were common to the neural crest and CVM lists: *AdamTS-A* (ADAM metallopeptidase with thrombospondin type 1 motif A), *csw* (tyrosine phosphatase), *Ahcy89E* (Adenosylhomocysteinase 89E), *Gdh*, *CG7255*, *CG32732*, *Nmdmc*, *Prat2* (Phosphoribosylamidotransferase 2), *DIP-epsilon* (Dpr-Interacting Protein epsilon), and *CG4733*. In *Xenopus* embryos, ADAM13 is required for cranial neural crest cell migration (Alfandari et al., 2001). Furthermore, AdamTS-A is reported to be expressed by hemocytes and CVM cells, in addition to other migratory cells (i.e. salivary gland, germ cells, and tracheal visceral branch) in the *Drosophila* embryo where it is promotes their migration (Ismat et al., 2013). The AdamTS-A transcript is clearly present in the head mesoderm at stage 10 but not in dispersed HCs at later stages (Figure 3), and this pattern (i.e. on only in HCs at early stages) may have contributed to the fact that this gene is not significantly enriched in our HC analysis. Nevertheless, identification of ADAMTS-A in these comparative analysis of neural crest and CVM cells highlights, again, the importance of ECM regulation during cell migration. ADAMTS-A, as proposed before (Ismat et al., 2013), may be a general regulator of diverse cell migratory mode, ranging from cells moving in tight collectives to those moving individually.

CONCLUSIONS

In this study, we first investigated two migrating cell types (CVM and hemocytes) within *Drosophila* embryos and identified enriched genes in each by employing RNA-Seq of RNA samples isolated from FACS-sorted cells. Validating our approach, each gene list includes a number of previously known genes: *HLH54f, htl, beat IIa, kon,* and *mew* for CVM; *proPO-A1, ppn, Cg25C, Vkg, srp* for HCs. In addition, we confirmed expression of several new genes within each cell type: *CG9416*, *CG5080*, *CG7124*, *beat IIb, grip, Syn2, taspase 1*, and *tey* for CVM cells; and *nimC4, CG6310, CG8501* for HCs.

We also noticed that the CVM list contains some genes that are predominantly expressed within neighboring tissues such as the TVM, gut primordium, yolk, and germ cells, which highlighted many potential cell-cell interactions impacting migration of CVM cells. Furthermore, when the CVM and HC gene lists were individually analyzed for functional term enrichment, we found that terms representative of subsequent cell-type specific functions were enriched, rather than functional terms that might otherwise relate to their shared behavioral, migratory property. The CVM gene list was associated with multiple terms related to muscle formation, clearly representing CVM cells' subsequent specification to longitudinal visceral muscle. Alternatively, the HC gene list was associated with terms reflecting HC function as immune cells. However, one gene in common to both CVM and HC lists is *zfh1*, which here we show using tissue specific RNAi and live imaging analysis acts cell-autonomously in CVM cells to support their migration.

In order to gain insight into shared mechanisms governing cell migration, we took a comparative approach to include enriched gene lists from two other migrating cell types: border cells (BCs) in the *Drosophila* ovary and chick neural crest cells. Neo, the only common gene to the two BC and CVM datasets, is required for proper migration in both cell types. Sn, present in the two BC and HC datasets, is known to regulate HC migration but its role in BC migration remains unclear. Whereas, when the vertebrate NC gene expression profile is compared to our CVM cell and HC list, 8 genes are common (including AdamTS-A if included) and 5 of these are ECM-related. This particular result illustrates the importance of ECM regulation and cell-environment interactions, which is evidently reflected in gene expression profiles of migratory cells. In sum, our comparative approach of gene expression profiles for multiple migratory cell populations, using our data and previously published data (Wang et al., 2006), successfully identified key players that are broadly required to support cell migration.

Additionally, we propose that the presence of *zfh1*, *neo*, and *sn* genes in some, but not all, migratory cell types examined suggests that these genes may support different migratory mechanisms. It was proposed that a spectrum exists in cell migratory modes ranging from tight collective to individual cell movement, relating to regulation of EMT (Campbell and Casanova, 2016). Within the spectrum, the cell streaming in a loose collective (exhibited by CVM cells and HC) may be placed at a moderately individual migratory mode. We propose that *zfh1* (and its vertebrate homologs) may be a key regulator that determines the mode of migration as it seems to be specifically required for streaming behavior. We show that *zfh1* is required for CVM migration, but expressed both by CVM cells and HCs, leaving open the possibility that it also has a role in HC migration. Zeb2/Sip1, the vertebrate ortholog of *zfh1*, positively regulates EMT, and thereby is required for neural crest cell delamination supporting migration (Rogers et al., 2013). In contrast, *zfh1* is not expressed within cells migrating as tight collectives such as

BCs (determined by *in situ* hybridization and antibody staining, data not shown), salivary gland, and tracheal cells (Lai et al., 1991). Neo, on the other hand, is expressed within CVM cells as well as BCs, and we observed migratory defects in both tissues when we ablated *neo* function via tissue-specific expression of RNAi. Similarly, Sn is required for HC migration (Zanet et al., 2009) but no phenotypes were detected when hypomorph mutants were assayed for border cell migration (Borghese et al., 2006). In sum, Zfh1 seems to be a general regulator of cell migration associated with those cell types that exhibit cell streaming behavior; whereas, whether Neo and Sn are general regulators is outstanding. At minimum, they are important for migration of CVM cells/BCs and HCs, respectively.

In conclusion, our comparative study encompassing four migratory cell populations across species identified multiple genes that potentially general and specific regulators for cell migration. Overlaps between gene expression profiling experiments can provide a quick mechanism to uncover important developmental regulators, in general.

EXPERIMENTAL PROCEDURES

Drosophila strains and genetic crosses. *Drosophila melanogaster* flies were maintained in standard condition at 25°C. Fly stocks used are following: HLH54F-Gap-Venus (GV2) (Stepanik et al., 2016), HLH54F-H2A-mCherry (Kadam et al., 2012), *HLH54F$^{\Delta 598}$* (Ismat et al., 2010), Srp-Gal4 (Brückner et al., 2004), UAS-Gap-Venus (Mavrakis et al., 2009), UAS-lacZ (Bloomington stock #8410), 5053-Gal4 (Bloomington stock #2702) and slbo-Gal4 (Bloomington stock #6458). The UAS-RNAi stocks are as follows: *zfh1* RNAi (TRiP, Bloomington #38929), *neo* RNAi (VDRC 32499) and *brk* RNAi (Kyoto 229-01).

Whole mount in situ hybridization and antibodies. For *in situ* hybridization and antibody staining, embryos were collected and processed by standard methods. Antisense RNA probes were labeled with digoxigenin (DIG) (Roche Diagnostics, 11277073910, 11464221), biotin (Roche Diagnostics, 11685597910, 11840821), or fluorescein (FITC) (Roche Diagnostics, 11685619910, 11852120). Antibodies used in this study are sheep anti-DIG Alkaline Phosphatase (AP) (Roche Diagnostics, 11093274910), goat anti-biotin AP (Sigma, 1002231758, SLBN8917V), rabbit anti-Green Fluorescent Protein (GFP) (1:1000, Life Technologies, A11122, 1753594), goat anti-GFP (1:5000, Rockland,

600101215, 25297), rabbit anti-Red Fluorescent Protein (RFP) (1:1000, MBL, PM005, 044), mouse anti-Fas3 (1:200, DSHB, 7G10), rat anti-Vasa (1:200, DSHB), rabbit anti-Vasa (1:1000, Santa Cruz), mouse anti-Lamin (1:100, DSHB, LC28.26), rabbit anti-FITC (Life Technologies, A889, 1661268), sheep anti-DIG (Roche Diagnostics, 11333089001), donkey anti-sheep 555nm (Life Technologies, A21436, 1719641), donkey anti-rabbit 647nm (Life Technologies, A31573, 1693297). DAB staining was performed using standard procedures and the VECTASTAIN Elite ABC kit (Vector Laboratories) followed by detection with DAB Peroxidase Substrate Kit (Vector Laboratories). Primer sequences for making riboprobes are listed in Sup. Table 6. For the cross-section of Figure 4, stained embryos were arranged and embedded in acetone-araldite (Electron Microscopy Science) and blocks were allowed to harden at 65° overnight. Sections at 8 µm were obtained by using a microtome (LKB Bromna 2218 Historange) and mounted in 1:1 acetone:araldite solution. Fluorescent in situ hybridization embryos were mounted in VECTASHIELD mounting medium (Vector Laboratories).

Ovary dissection, fixation, and immunostaining. Ovaries were collected, fixed, and immunostained as previously described (Zimmerman et al., 2013). *slbo*-GAL4, UAS-GFP females were crossed to either *yw* males for controls or the appropriate RNAi line. The resulting F1 progeny were allowed to develop and eclose, and the adult flies were transferred to 29°C, where they were incubated and allowed to feed on yeast paste for at least two days. Between 10-20 adult female flies were dissected according to standard protocols in EBR solution, and the collected ovaries were subsequently fixed in 4% paraformaldehyde in 1X PBT with 1% DMSO for 20 min with rocking at RT. The fixed ovaries were then rinsed with 1X PBT three times of 5 minutes each, and were permeabilized in 1% Triton X-100 in PBT for one hour with gentle separation of ovaries by pipetting up and down. The ovaries were then washed three times in 1X PBT for 5 minutes each, and then incubated for one hour in blocking solution (20% WBR in PBT). After the blocking step, the ovaries were incubated in diluted primary antibody overnight (~18 hours) at 4°C with rocking. The ovaries were subsequently washed in blocking solution four times for 10 minutes each at RT, and were then incubated in diluted secondary antibodies for 2-3 hours in the dark (as with all subsequent steps). After antibody incubation, the ovaries were washed in blocking solution an additional four times for 10 minutes each, and were subsequently incubated in a diluted solution of ActinRed™ ReadyProbes™ reagent (Molecular Probes, Lot#1771007) for 30 minutes. The ovaries were then washed in blocking solution four times for 10 minutes each and then mounted in VECTASHIELD mounting medium (Vector Laboratories). Imaging was performed using a Pascal confocal microscope (Zeiss). P-values were calculated to test the null hypothesis that neo RNAi and the three controls are identically distributed. Three test statistics were used: the difference in means, the symmetrized Kullback-Leibler divergence, and the Komolgorov-Smirnov statistic.

Live imaging and image processing. Live imaging was conducted as described previously (Kadam et al., 2012) with the following modifications. In short, embryos were collected for 1 hour and aged for ~4.0-4.5 hours in a humidified 24°C incubator to reach stage 10. These staged embryos were hand dechorionated and placed on a coverslip coated with dried heptane-glue mixture. ddH_2O was added onto the embryos, which were subsequently imaged using a 40x water lens objective on a Zeiss LSM5 Pascal confocal microscope. The time-course of images collected by LSM software was compiled using ImageJ, where linear adjustments such as brightness/contrast and levels were made and temporal color-coded projections were generated from compiled hyperstacks. Movie files were generated using ImageJ after adding the appropriate time stamps.

Embryo collection for FACS. Healthy young adult flies were kept in 4-5 plastic cages where 100 mm plastic petri dishes are fitted at the bottom. Petri dishes contained apple juice agar with brewer's yeast paste. For embryo collection, agar plates from first cycle were discarded to maximize the number of synchronized embryos. For CVM marker expressing embryo collection, HLH54F-Gap-Venus (GV2; Stepanik et al., 2016) expressing embryos were collected for 5 hours at 24°C and aged until the population reaches embryonic stage 10 through 13 at either 24 or 18 degrees. For hemocyte marker-expressing embryo collection, Srp-Gal4; UAS-Gap-Venus embryos were collected for 5 hours at 24°C and aged until the population reaches stage 12-16 at either 24 or 18 degrees. When needed, properly aged embryos were kept at 4°C for no more than a maximum of 4 hours until processing.

Cell isolation. Staged embryos were dechorionated with fresh 50% bleach solution for 3 min and thoroughly washed with running water for > 1 min on a strainer. For final wash step, Millipore filtered DI water followed by ice cold Schneider's medium (Gibco, cat. #21720) was used. From this step, embryos and cell suspension were kept on ice until sorting. The following cell isolation method is slightly modified from (Estrada and Michelson, 2008). 20~30 embryos were transferred with a fine paintbrush into a dounce homogenizer (Wheaton, cat. #357542, loose fitting pestle, 7 mL) filled with ice cold Schneider's medium. The embryos were homogenized with 7 gentle strokes and the resulting cell suspension was transferred to chilled 50mL conical tube. When filled, tubes were centrifuged at 40 g for 5 min to pellet the tissue debris or clumps. The single cell suspension in the supernatant was transferred to a new 50 mL tube and centrifuged at 380 g for 10 min to pellet cells. After discarding the supernatant, cell pellets were resuspended and pooled with 8% FBS in Schneider's Medium. To remove cell clumps or cell debris for sorting, the cell suspension was gravity filtered through 40 micron nylon mesh prior to sorting.

Sorting migrating cells by FACS. The cell suspension was sorted using iCyt Mission Technology Reflection Cell Sorter. Machine setting were optimized to the *Drosophila* cell sorting mode. The sheath solution (Seecof saline) was freshly made the day before sorting and filtered 0.22 micron filter. The

composition of Seecof saline is 6 mM Na_2HPO_4, 3.67 mM KH_2PO_4, 106 mM NaCl, 26.8 mM KCl, 6.4 mM $MgCl_2$, 2.25 mM $CaCl_2$ and adjusted to be pH 6.8. A small fraction of cell suspension was used to confirm that dead cell population is negligible (<0.1%) by 10 μg/ml Propidium Iodide (PI) staining. Cell suspension from *yw* flies served as negative control to set the baseline to select the Venus positive cell population. Positive and negative cells were sorted directly into cell lysis solution to maximize total RNA extraction efficiency. The total RNA from each population was extracted using Qiagen RNeasy Mini Kit (Cat. #74104) with on column DNase treatment as suggested in the protocol. The integrity of RNA samples were verified using Bioanalyzer (Agilent Cat. #5067) and the RNA concentration was measured using Qubit fluorometer (Thermo Fisher, Cat. #32866) for each sample.

Bioinformatics and GO term analysis. The RNA-Seq using lllumina HiSeq2500 sequencer using the isolated RNA is performed by standard manufacturer's protocol at the Millard and Muriel Jacobs Genetics and Genomics Laboratory. The raw .fastq files were transferred and mapped to the *Drosophila melanogaster* reference genome using TopHat via Bowtie. The resulting .bam files from marker positive and negative samples were compared using CuffDiff program to obtain the list of differentially expressed genes. The resulting data were further processed in Microsoft Excel worksheets. For GO term enrichment of gene lists, DAVID web server (Dennis et al., 2003) for functional annotation search tools was used with flybase IDs of a given gene set. Terms enriched at the gene number count greater than 2 and *p*-value higher than 0.1 were considered significant. For individual 'Biological Process', 'Molecular Process' and 'Cellular Component' GO term searches, GOTERM_BP_DIRECT, GOTERM_MP_DIRECT, and GOTERM_CC_DIRECT category was selected, respectively. For clustering enrichment analysis, default DAVID categories, including UP_keywordsm SEQ_features, Interpro protein domains, and GO terms (Molecular Process, Biological Process, Cellular Component), were searched.

AUTHOR CONTRIBUTIONS

Y.-K.B. and A.S. devised the overall question and experimental approach; Y.-K.B conducted the FACS experiments and data analysis; F.M., H.C., and Y.-K.B. performed in situ hybridization and RNAi experiments; Y.-K.B., F.M, H.C., and A.S. analyzed the data; and Y.-K.B., F.M, H.C, and A.S. wrote the paper.

ACKNOWLEDGMENTS

We thank Katja Brückner, Manfred Frasch, and Jennifer Lippincott-Schwartz for sharing *Drosophila* stocks and probes. We are grateful to Igor Antoshekien in the Millard and Muriel Jacobs Genetics and Genomics Laboratory at California Institute of Technology for sequencing support and answering questions on data analysis. We also thank Rochelle Diamond at the Caltech Flow Cytometry Facility for excellent assistance in cell sorting. We additionally appreciate Justin Bois for his data analysis advice. This work was supported by National Institute of Health grants R01GM1043838 to A.S., 1F32GM119395-01 to F.M., and 5T32GM007616-38 to H.C., and by American Heart Association postdoctoral fellowship grant 11POST7600181 to Y.-K.B.

REFERENCES

Abel, T., Michelson, A.M., Maniatis, T., 1993. A Drosophila GATA family member that binds to Adh regulatory sequences is expressed in the developing fat body. Development (Cambridge, England) 119, 623-633.

Alfandari, D., Cousin, H., Gaultier, A., Smith, K., White, J.M., Darribère, T., DeSimone, D.W., 2001. Xenopus ADAM 13 is a metalloprotease required for cranial neural crest-cell migration. Current biology : CB 11, 918-930.

Assaker, G., Ramel, D., Wculek, S.K., González-Gaitán, M., Emery, G., 2010. Spatial restriction of receptor tyrosine kinase activity through a polarized endocytic cycle controls border cell migration. Proceedings of the National Academy of Sciences of the United States of America 107, 22558-22563.

Beccari, S., Teixeira, L.s., Rørth, P., 2002. The JAK/STAT pathway is required for border cell migration during Drosophila oogenesis. Mechanisms of Development 111, 115-123.

Bhat, K.M., 2007. Wingless activity in the precursor cells specifies neuronal migratory behavior in the Drosophila nerve cord. Developmental biology 311, 613-622.

Binggeli, O., Neyen, C., Poidevin, M., Lemaitre, B., 2014. Prophenoloxidase activation is required for survival to microbial infections in Drosophila. PLoS pathogens 10, e1004067-e1004067.

Borghese, L., Fletcher, G., Mathieu, J., Atzberger, A., Eades, W.C., Cagan, R.L., Rørth, P., 2006. Systematic analysis of the transcriptional switch inducing migration of border cells. Developmental cell 10, 497-508.

Brückner, K., Kockel, L., Duchek, P., Luque, C.M., Rørth, P., Perrimon, N., 2004. The PDGF/VEGF receptor controls blood cell survival in Drosophila. Developmental cell 7, 73-84.

Broihier, H.T., Moore, L.A., Van Doren, M., Newman, S., Lehmann, R., 1998. zfh-1 is required for germ cell migration and gonadal mesoderm development in Drosophila. Development (Cambridge, England) 125, 655-666.

Bunt, S., Hooley, C., Hu, N., Scahill, C., Weavers, H., Skaer, H., 2010. Hemocyte-secreted type IV collagen enhances BMP signaling to guide renal tubule morphogenesis in Drosophila. Developmental cell 19, 296-306.

Campbell, K., Casanova, J., 2016. A common framework for EMT and collective cell migration. Development (Cambridge, England) 143, 4291-4300.

Cripps, R.M., Zhao, B., Olson, E.N., 1999. Transcription of the myogenic regulatory gene Mef2 in cardiac, somatic, and visceral muscle cell lineages is regulated by a Tinman-dependent core enhancer. Dev Biol 215, 420-430.

Denef, N., Chen, Y., Weeks, S.D., Barcelo, G., Schüpbach, T., 2008. Crag regulates epithelial architecture and polarized deposition of basement membrane proteins in Drosophila. Developmental cell 14, 354-364.

Dennis, G., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., Lempicki, R.A., 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome biology 4, P3-P3.

Duchek, P., Rørth, P., 2001. Guidance of cell migration by EGF receptor signaling during Drosophila oogenesis. Science (New York, N.Y.) 291, 131-133.

Duong, T.D., Erickson, C.A., 2004. MMP-2 plays an essential role in producing epithelial-mesenchymal transformations in the avian embryo. Developmental dynamics : an official publication of the American Association of Anatomists 229, 42-53.

Estrada, B., Gisselbrecht, S.S., Michelson, A.M., 2007. The transmembrane protein Perdido interacts with Grip and integrins to mediate myotube projection and attachment in the Drosophila embryo. Development (Cambridge, England) 134, 4469-4478.

Estrada, B., Michelson, A.M., 2008. A genomic approach to myoblast fusion in Drosophila. Methods in molecular biology (Clifton, N.J.) 475, 299-314.

Evans, I.R., Hu, N., Skaer, H., Wood, W., 2010. Interdependence of macrophage migration and ventral nerve cord development in Drosophila embryos. Development (Cambridge, England) 137, 1625-1633.

Evans, I.R., Wood, W., 2011. Drosophila embryonic hemocytes. Current biology : CB 21, R173-174.

Ferjoux, G., Auge, B., Boyer, K., Haenlin, M., Waltzer, L., 2007. A GATA/RUNX cis-regulatory module couples Drosophila blood cell commitment and differentiation into crystal cells. Dev Biol 305, 726-734.

Fernandes, I., Chanut-Delalande, H., Ferrer, P., Latapie, Y., Waltzer, L., Affolter, M., Payre, F., Plaza, S., 2010. Zona pellucida domain proteins remodel the apical compartment for localized cell shape changes. Developmental cell 18, 64-76.

Gallagher, S.M., Castorino, J.J., Philp, N.J., 2009. Interaction of monocarboxylate transporter 4 with beta1-integrin and its role in cell migration. American journal of physiology. Cell physiology 296, C414-421.

Georgias, C., Wasser, M., Hinz, U., 1997. A basic-helix-loop-helix protein expressed in precursors of Drosophila longitudinal visceral muscles. Mech Dev 69, 115-124.

Gline, S., Kaplan, N., Bernadskaya, Y., Abdu, Y., Christiaen, L., 2015. Surrounding tissues canalize motile cardiopharyngeal progenitors towards collective polarity and directed migration. Development 142.

Halestrap, A.P., Wilson, M.C., 2012. The monocarboxylate transporter family--role and regulation. IUBMB life 64, 109-119.

Hashimoto, C., Kim, D.R., Weiss, L.A., Miller, J.W., Morisato, D., 2003. Spatial regulation of developmental signaling by a serpin. Developmental cell 5, 945-950.

Honti, V., Csordás, G., Kurucz, É., Márkus, R., Andó, I., 2014. The cell-mediated immunity of Drosophila melanogaster: hemocyte lineages, immune compartments, microanatomy and regulation. Developmental and comparative immunology 42, 47-56.

Ismat, A., Cheshire, A.M., Andrew, D.J., 2013. The secreted AdamTS-A metalloprotease is required for collective cell migration. Development (Cambridge, England) 140, 1981-1993.

Ismat, A., Schaub, C., Reim, I., Kirchner, K., Schultheis, D., Frasch, M., 2010. HLH54F is required for the specification and migration of longitudinal gut muscle founders from the caudal mesoderm of Drosophila. Development (Cambridge, England) 137, 3107-3117.

Izumi, H., Takahashi, M., Uramoto, H., Nakayama, Y., Oyama, T., Wang, K.-Y., Sasaguri, Y., Nishizawa, S., Kohno, K., 2011. Monocarboxylate transporters 1 and 4 are involved in the invasion activity of human lung cancer cells. Cancer science 102, 1007-1013.

Jambor, H., Surendranath, V., Kalinka, A.T., Mejstrik, P., Saalfeld, S., Tomancak, P., 2015. Systematic imaging reveals features and changing localization of mRNAs in Drosophila development. eLife 4.

Jaźwińska, A., Ribeiro, C., Affolter, M., 2003. Epithelial tube morphogenesis during Drosophila tracheal development requires Piopio, a luminal ZP protein. Nature cell biology 5, 895-901.

Kadam, S., Ghosh, S., Stathopoulos, A., 2012. Synchronous and symmetric migration of Drosophila caudal visceral mesoderm cells requires dual input by two FGF ligands. Development (Cambridge, England) 139, 699-708.

Kobayashi, M., Aita, N., Hayashi, S., Okada, K., Ohta, T., Hirose, S., 1998. DNA supercoiling factor localizes to puffs on polytene chromosomes in Drosophila melanogaster. Mol Cell Biol 18, 6737-6744.

Kramerova, I.A., Kramerov, A.A., Fessler, J.H., 2003. Alternative splicing of papilin and the diversity of Drosophila extracellular matrix during embryonic morphogenesis. Dev Dyn 226, 634-642.

Kurucz, E., Márkus, R., Zsámboki, J., Folkl-Medzihradszky, K., Darula, Z., Vilmos, P., Udvardy, A., Krausz, I., Lukacsovich, T., Gateff, E., Zettervall, C.-J., Hultmark, D., Andó, I., 2007. Nimrod, a putative phagocytosis receptor with EGF repeats in Drosophila plasmatocytes. Current biology : CB 17, 649-654.

Kusch, T., Reuter, R., 1999. Functions for Drosophila brachyenteron and forkhead in mesoderm specification and cell signalling. Development (Cambridge, England) 126, 3991-4003.

Lai, Z.C., Fortini, M.E., Rubin, G.M., 1991. The embryonic expression patterns of zfh-1 and zfh-2, two Drosophila genes encoding novel zinc-finger homeodomain proteins. Mechanisms of development 34, 123-134.

Leigh, N.R., Schupp, M.-O., Li, K., Padmanabhan, V., Gastonguay, A., Wang, L., Chun, C.Z., Wilkinson, G.A., Ramchandran, R., 2013. Mmp17b is essential for proper neural crest cell migration in vivo. PloS one 8, e76484-e76484.

Lilly, B., Zhao, B., Ranganayakulu, G., Paterson, B.M., Schulz, R.A., Olson, E.N., 1995. Requirement of MADS domain transcription factor D-MEF2 for muscle formation in Drosophila. Science (New York, N.Y.) 267, 688-693.

Liu, M., Su, M., Lyons, G.E., Bodmer, R., 2006. Functional conservation of zinc-finger homeodomain gene zfh1/SIP1 in Drosophila heart development. Development genes and evolution 216, 683-693.

Luo, J., Zuo, J., Wu, J., Wan, P., Kang, D., Xiang, C., Zhu, H., Chen, J., 2015. In vivo RNAi screen identifies candidate signaling genes required for collective cell migration in Drosophila ovary. Sci China Life Sci 58, 379-389.

Mandal, L., Dumstrei, K., Hartenstein, V., 2004. Role of FGFR signaling in the morphogenesis of the Drosophila visceral musculature. Developmental dynamics : an official publication of the American Association of Anatomists 231, 342-348.

Mavrakis, M., Rikhy, R., Lippincott-Schwartz, J., 2009. Plasma membrane polarity and compartmentalization are established before cellularization in the fly embryo. Dev Cell 16, 93-104.

Miller, C.M., Page-McCaw, A., Broihier, H.T., 2008. Matrix metalloproteinases promote motor axon fasciculation in the Drosophila embryo. Development (Cambridge, England) 135, 95-109.

Montell, D.J., 2001. Command and control: regulatory pathways controlling invasive behavior of the border cells. Mechanisms of development 105, 19-25.

Nappi, A.J., Vass, E., 1998. Hydrogen peroxide production in immune-reactive Drosophila melanogaster. The Journal of parasitology 84, 1150-1157.

Nguyen, D.X., Massagué, J., 2007. Genetic determinants of cancer metastasis. Nature reviews. Genetics 8, 341-352.

Pao, S.S., Paulsen, I.T., Saier, M.H., 1998. Major facilitator superfamily. Microbiology and molecular biology reviews : MMBR 62, 1-34.

Pipes, G.C., Lin, Q., Riley, S.E., Goodman, C.S., 2001. The Beat generation: a multigene family encoding IgSF proteins related to the Beat axon guidance molecule in Drosophila. Development (Cambridge, England) 128, 4545-4552.

Pocha, S.M., Montell, D.J., 2014. Cellular and molecular mechanisms of single and collective cell migrations in Drosophila: themes and variations. Annual review of genetics 48, 295-318.

Rørth, P., 2002. Initiating and guiding migration: lessons from border cells. Trends in cell biology 12, 325-331.

Reim, I., Hollfelder, D., Ismat, A., Frasch, M., 2012. The FGF8-related signals Pyramus and Thisbe promote pathfinding, substrate adhesion, and survival of migrating longitudinal gut muscle founder cells. Dev Biol 368, 28-43.

Rhoads, J.M., Chen, W., Gookin, J., Wu, G.Y., Fu, Q., Blikslager, A.T., Rippe, R.A., Argenzio, R.A., Cance, W.G., Weaver, E.M., Romer, L.H., 2004. Arginine stimulates intestinal cell migration through a focal adhesion kinase dependent mechanism. Gut 53, 514-522.

Roch, F., Alonso, C.R., Akam, M., 2003. Drosophila miniature and dusky encode ZP proteins required for cytoskeletal reorganisation during wing morphogenesis. Journal of cell science 116, 1199-1207.

Rogers, C.D., Saxena, A., Bronner, M.E., 2013. Sip1 mediates an E-cadherin-to-N-cadherin switch during cranial neural crest EMT. The Journal of cell biology 203, 835-847.

Savant-Bhonsale, S., Montell, D.J., 1993. torso-like encodes the localized determinant of Drosophila terminal pattern formation. Genes & development 7, 2548-2555.

Schnorrer, F., Kalchhauser, I., Dickson, B.J., 2007. The transmembrane protein Kon-tiki couples to Dgrip to mediate myotube targeting in Drosophila. Developmental cell 12, 751-766.

Siebert, M., Banovic, D., Goellner, B., Aberle, H., 2009. Drosophila motor axons recognize and follow a Sidestep-labeled substrate pathway to reach their target fields. Genes & development 23, 1052-1062.

Siekhaus, D., Haesemeyer, M., Moffitt, O., Lehmann, R., 2010. RhoL controls invasion and Rap1 localization during immune cell transmigration in Drosophila. Nature cell biology 12, 605-610.

Silver, D.L., Montell, D.J., 2001. Paracrine Signaling through the JAK/STAT Pathway Activates Invasive Behavior of Ovarian Epithelial Cells in Drosophila. Cell 107, 831-841.

Simões-Costa, M., Tan-Cabugao, J., Antoshechkin, I., Sauka-Spengler, T., Bronner, M.E., 2014. Transcriptome analysis reveals novel players in the cranial neural crest gene regulatory network. Genome research 24, 281-290.

Stauber, R.H., Hahlbrock, A., Knauer, S.K., Wünsch, D., 2016. Cleaving for growth: threonine aspartase 1--a protease relevant for development and disease. FASEB journal : official publication of the Federation of American Societies for Experimental Biology 30, 1012-1022.

Stepanik, V., Dunipace, L., Bae, Y.-K., Macabenta, F., Sun, J., Trisnadi, N., Stathopoulos, A., 2016. The migrations of Drosophila muscle founders and primordial germ cells are interdependent. Development (Cambridge, England) 143, 3206-3215.

Stevens, L.J., Page-McCaw, A., 2012. A secreted MMP is required for reepithelialization during wound healing. Molecular biology of the cell 23, 1068-1079.

Tedeschi, P.M., Bansal, N., Kerrigan, J.E., Abali, E.E., Scotto, K.W., Bertino, J.R., 2016. NAD+ Kinase

as a Therapeutic Target in Cancer. Clinical cancer research : an official journal of the American Association for Cancer Research 22, 5189-5195.

Theveneau, E., Mayor, R., 2012. Neural crest delamination and migration: from epithelium-to-mesenchyme transition to collective cell migration. Developmental biology 366, 34-54.

Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S., Lewis, S.E., Richards, S., Ashburner, M., Hartenstein, V., Celniker, S.E., Rubin, G.M., 2002. Systematic determination of patterns of gene expression during Drosophila embryogenesis. Genome biology 3, RESEARCH0088-RESEARCH0088.

Tomancak, P., Berman, B.P., Beaton, A., Weiszmann, R., Kwan, E., Hartenstein, V., Celniker, S.E., Rubin, G.M., 2007. Global analysis of patterns of gene expression during Drosophila embryogenesis. Genome biology 8, R145-R145.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., Pachter, L., 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nature biotechnology 31, 46-53.

Trapnell, C., Pachter, L., Salzberg, S.L., 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics (Oxford, England) 25, 1105-1111.

Trisnadi, N., Stathopoulos, A., 2014. Ectopic expression screen identifies genes affecting Drosophila mesoderm development including the HSPG Trol. G3 (Bethesda, Md.) 5, 301-313.

Urbano, J.M., Dominguez-Gimenez, P., Estrada, B., Martin-Bermudo, M.D., 2011. PS integrins and laminins: key regulators of cell migration during Drosophila embryogenesis. PLoS One 6, e23893.

Vandewalle, C., Van Roy, F., Berx, G., 2009. The role of the ZEB family of transcription factors in development and disease. Cellular and molecular life sciences : CMLS 66, 773-787.

Wang, X., Bo, J., Bridges, T., Dugan, K.D., Pan, T.-c., Chodosh, L.A., Montell, D.J., 2006. Analysis of cell migration using whole-genome expression profiling of migratory cells in the Drosophila ovary. Developmental cell 10, 483-495.

Wood, W., Faria, C., Jacinto, A., 2006. Distinct mechanisms regulate hemocyte chemotaxis during development and wound healing in Drosophila melanogaster. J Cell Biol 173, 405-416.

Zaffran, S., Küchler, A., Lee, H.H., Frasch, M., 2001. biniou (FoxF), a central component in a regulatory network controlling visceral mesoderm development and midgut morphogenesis in Drosophila. Genes & development 15, 2900-2915.

Zanet, J., Stramer, B., Millard, T., Martin, P., Payre, F., Plaza, S., 2009. Fascin is required for blood cell migration during Drosophila embryogenesis. Development (Cambridge, England) 136, 2557-2565.

Zimmerman, S.G., Peters, N.C., Altaras, A.E., Berg, C.A., 2013. Optimized RNA ISH, RNA FISH and protein-RNA double labeling (IF/FISH) in Drosophila ovaries. Nature protocols 8, 2158-2179.

Zinn, K., 2009. Choosing the road less traveled by: a ligand-receptor system that controls target recognition by Drosophila motor axons. Genes & development 23, 1042-1045.

FIGURES WITH LEGENDS

Figure 1. Transcriptome analysis of the CVM cells and hemocytes (HC) from *Drosophila* embryos. (A-A'') Confocal images of CVM cells (red, HLH54F-H2A-mCherry) and HC (green, Srp-Gal4>UAS-Gap-Venus) within a live embryo. (B) Experimental scheme. (C) In order to set the threshold for Venus signal (CVM marker HLH54F-Gap-Venus), the cells from control (*yw*) embryos of matching stages was analyzed. (D) Flow cytometric analysis of cells from the CVM marker (HLH54F-Gap-Venus) expressing embryos shows a small but distinct population with Venus signal. The upper (R2) and lower (R3) populations are sorted as CVM marker (+) and (-) cells. (E) The CVM transcriptome analysis by comparing RNA-Seq reads from CVM marker positive (Y-axis) and negative samples (X-axis). Each dot represents a gene and plotted with the RPKMs (reads per kilobases per million reads) from CVM marker positive and negative populations. Red dots are enriched genes in CVM, blue dots are downregulated genes in CVM, and grey dots represent genes that did not show significant difference between these two populations. (E') Differentially expressed genes between CVM and non-CVM samples are plotted with their fold change (CVM / non CVM) and enrichment significance ($-\log_{10}$(*p*-value)). To plot *p*-values of 0 (highest significance) in log scale, the maximum was set for enrichment significance as 16. (F) The threshold for Venus signal (HC marker Srp-Gal4>UAS-Gap-Venus) was set using the cells from control (*yw*) embryos of matching stages. (G) Flow cytometric analysis of cells from the HC marker expressing embryos shows a distinct population with Venus signal. The upper (R2) and lower (R3) populations are sorted as HC marker (+) and (-) cells. (H) The HC transcriptome analysis by comparing RNA-Seq reads from HC marker positive (Y-axis) and negative samples (X-axis). Each dot represents a gene and plotted with the RPKMs (reads per kilobase million reads) from HC marker positive and negative populations. Red dots are enriched genes in HC; blue dots are downregulated genes in HC; and grey dots represent genes that did not show significant difference between these two populations. (H') Differentially expressed genes between HC and non-HC samples are plotted with their fold change (HC/ non HC) and enrichment significance ($-\log_{10}$(*p*-value)). To plot *p*-values of 0 (highest significance) in a log scale, the maximum was set for enrichment significance as 16.

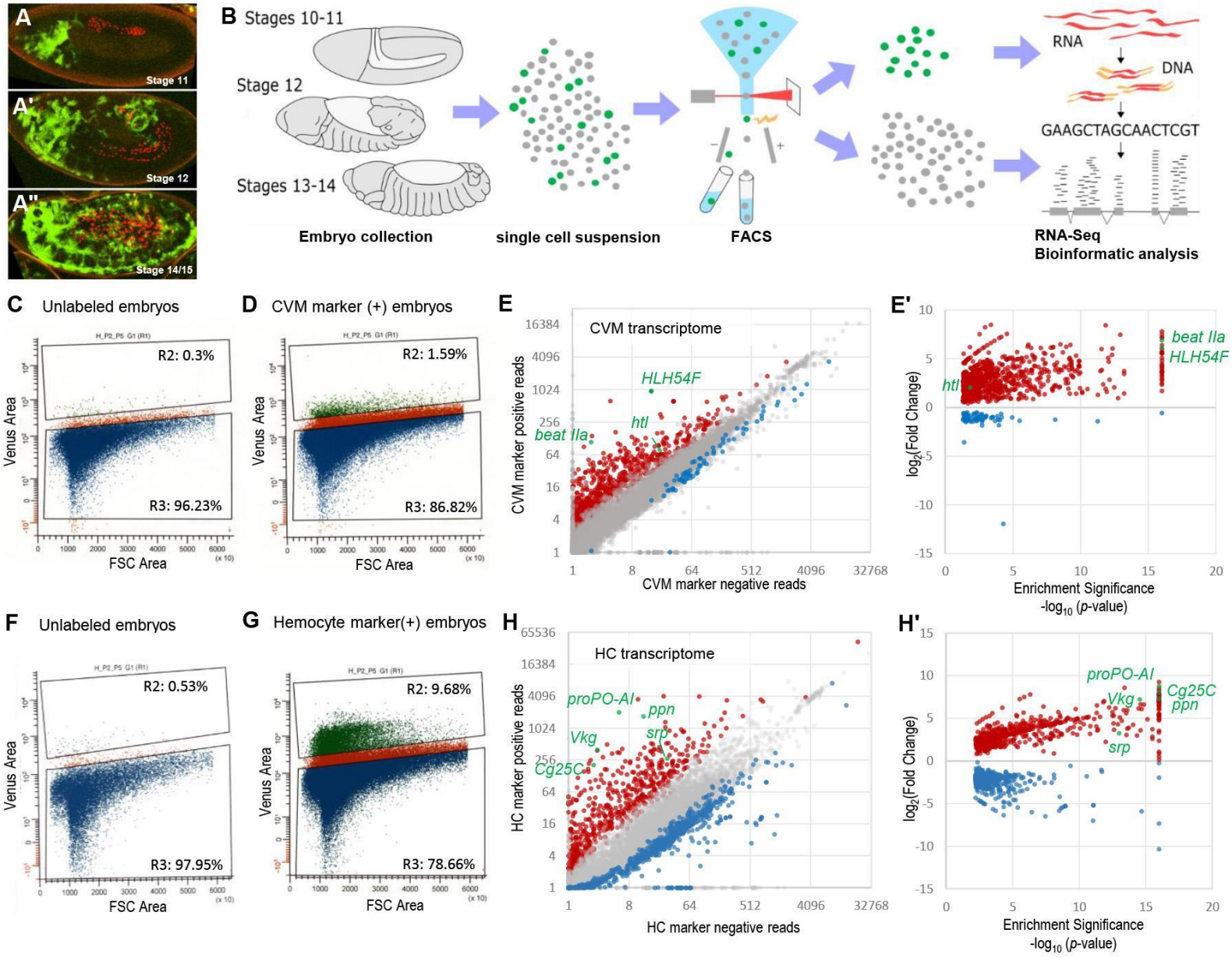
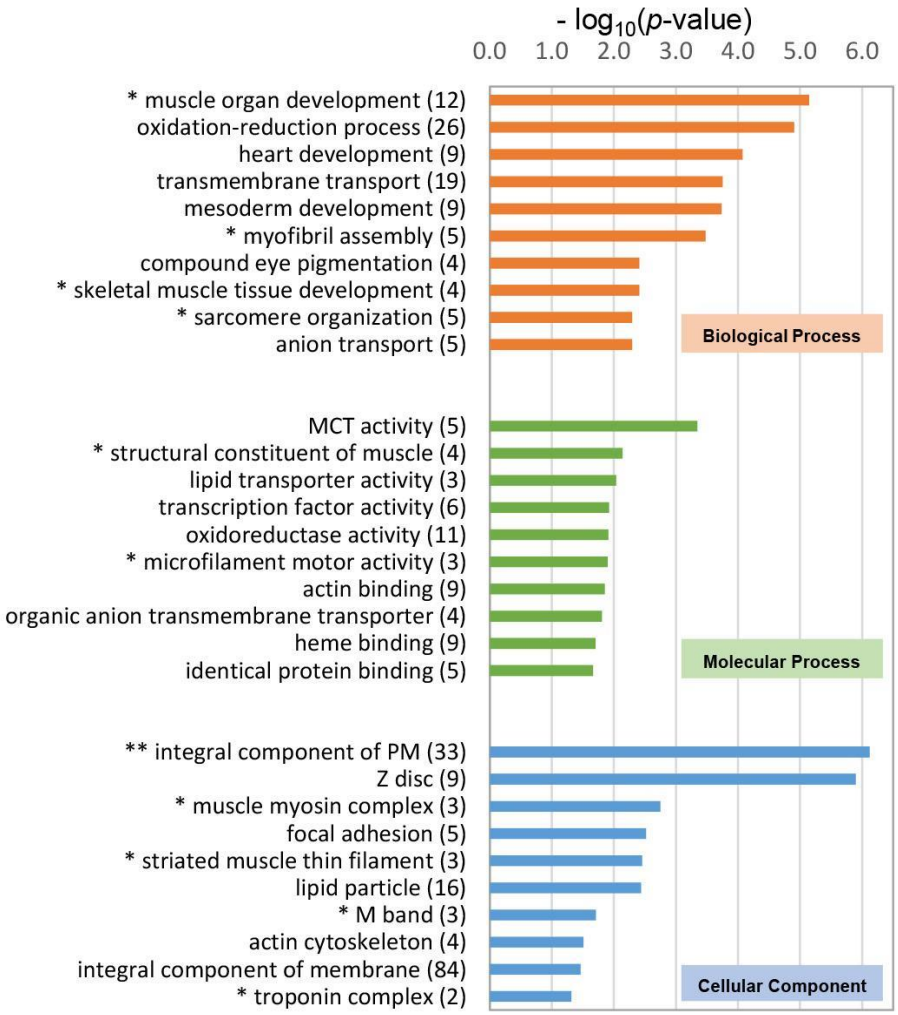
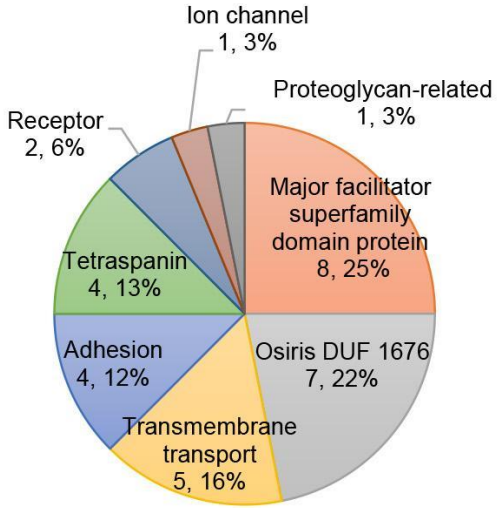


Figure 2. GO term analysis of CVM enriched genes. (A) The top 10 list of each GO term category (biological process, molecular process, cellular component) in CVM short list (n=324) using DAVID database. Each term is sorted with its enrichment significance (- $\log_{10}$(*p*-value)). Asterisk (*) denotes terms relevant to CVM function as muscle founder cells and ** denotes the most enriched term (see next). (B) A pie chart showing subclasses within the "integral component of plasma membrane**" (n=33). The subclass name, number of genes, and percentage within this class is labeled. (C) List of genes from enriched GO terms of interest. The term "Transcription factor activity, sequence specific DNA binding" is not shown in (A), but contains the most number of transcription factors. PM: plasma membrane, MCT: monocarboxylate transporter.

A GO term analysis on CVM upregulated genes



B Subclasses of "Integral component of plasma membrane **"



C **Members of other classes of interest**

Muscle organ development	*Mef2, wupA, Dys, zfh1, sls, Mhc, htl, eve, kon, Mlp60A, mbl, nvy*
Transcription factor activity, sequence specific DNA binding	*Mef2, Sox14, eve, croc, Hr4, Doc2, Kr-h1, H2.0, Fer1, so, Doc1, HLH54F, luna, bin, nvy, cad, sisA*

Figure 3. *In situ* hybridizations confirm gene expression in the CVM. The top panel is a representative antibody staining specifically marking CVM cells. All embryos are oriented with the anterior to the left; those marked with a 'D' are shown from the dorsal view, while all others are seen from the lateral view. Arrowheads refer to CVM-specific staining, in those cases where this staining could be deemed ambiguous due to additional staining in other tissues, levels of expression, or viewpoint. Gene expression patterns are shown in (A) wild-type embryos, stages 10-13 (B) *HLH54F*$^{\Delta 598}$ mutant embryos lacking CVM, stages 11, 12, or 13 and (C) *htl*^{AB42} mutant embryos with CVM displaying various migratory defects, stages 11 or 12.

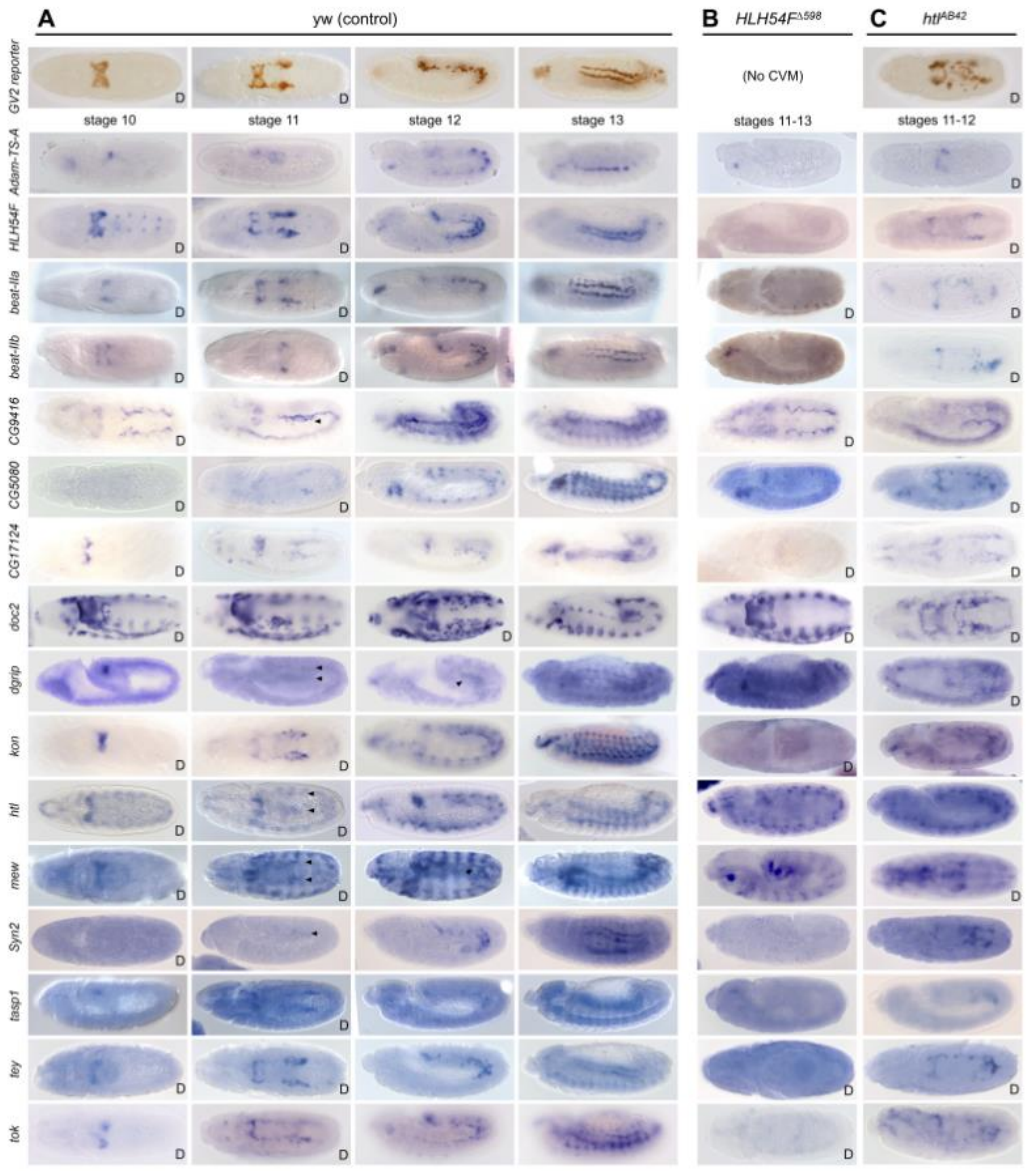


Figure 4. Unexpected trends in our transcriptional profiling data highlight potentially meaningful tissue interactions. (A) Diagram and cross-section of a stage 12 *Drosophila* embryo that illustrate the main tissues that potentially interact with the CVM. mRNA expression patterns in the TVM (B), gut primordium (C), yolk (D), germ cells (E), and the less common anal pad precursors (F), which reside above the CVM before they begin their migration in stage 10.

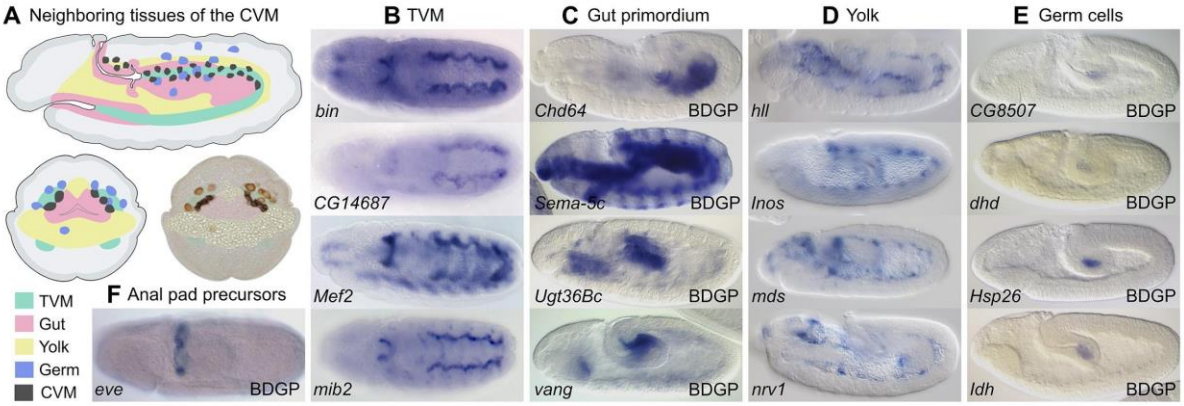


Figure 5. GO term analysis of hemocyte enriched genes. (A) The top 10 list of each GO term category (biological process, molecular process, cellular component) in hemocyte short list (n=386) using DAVID database. Each term is sorted with its enrichment significance (- $\log_{10}$(*p*-value)). * denotes terms relevant to HC function as immune cells and processing reactive oxygen species. ** denotes the most enriched term "plasma membrane". (B) A pie chart showing subclasses within the "plasma membrane" (n=51). The subclass name, number of genes, and percentage for each subclass is shown. (C) List of genes from enriched GO terms of interest. MCT: monocarboxylate transporter, ECM: extracellular matrix.

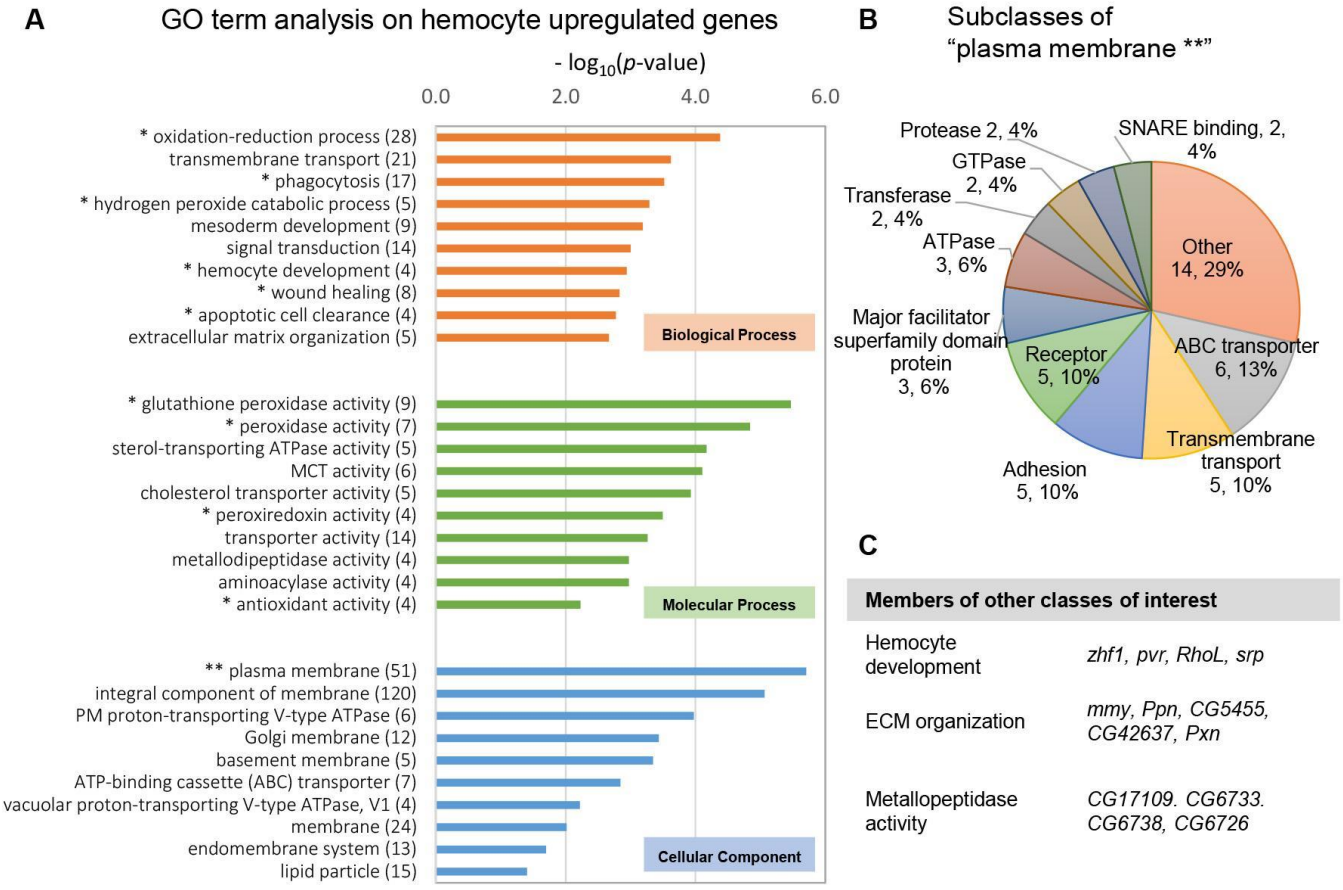


Figure 6. In situ hybridizations confirm gene expression in the hemocytes. Embryos shown from dorsal view. Top panels depict immunostained embryos expressing the Venus reporter under the control of the HC-specific *srp*-GAL4 driver. *In situ* hybridization using indicated riboprobes to detect expression in embryos spanning stages 9-13.

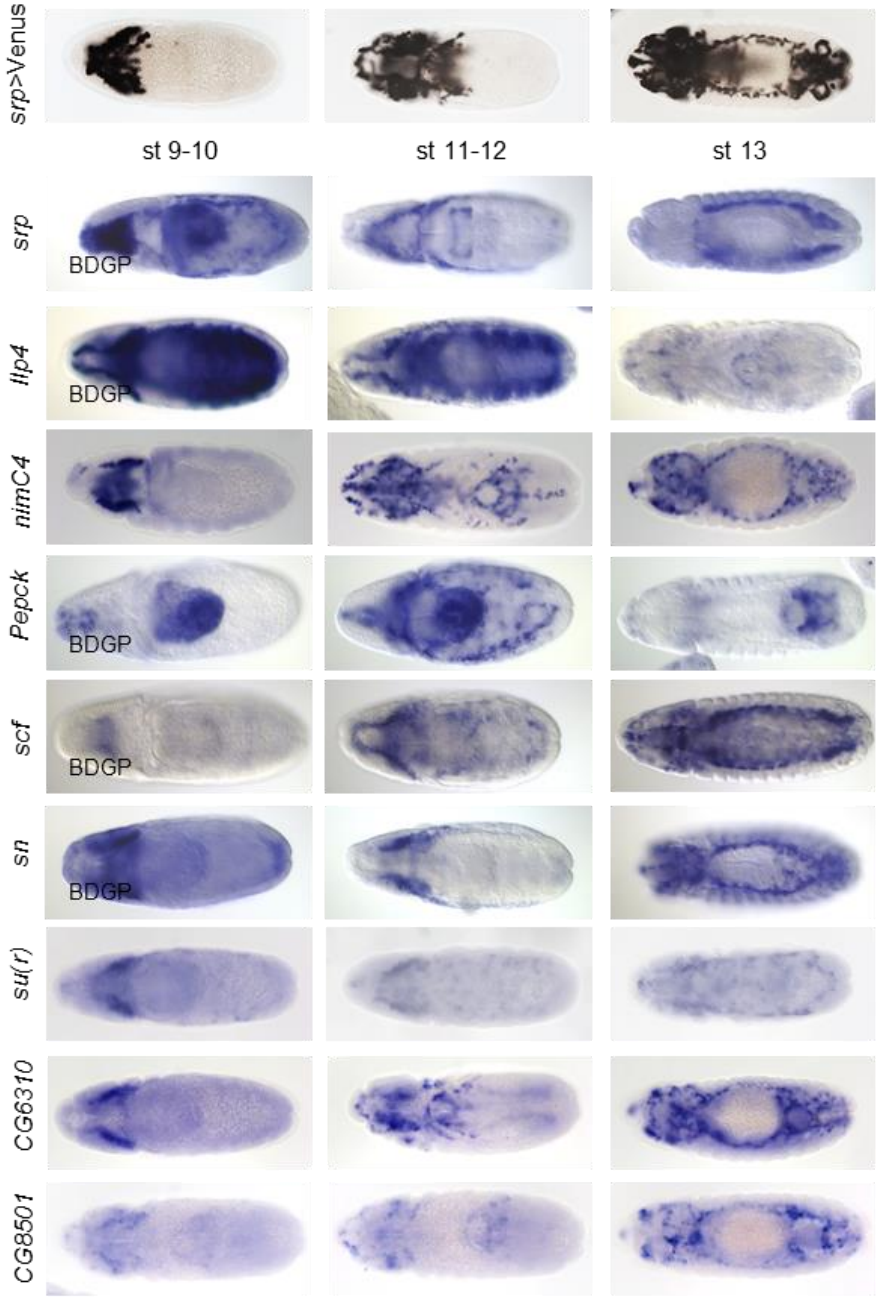
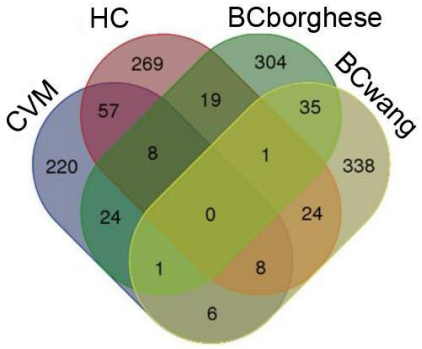


Figure 7. Comparative analysis of gene expression profiles of migratory cell populations within and across species. (A) A Venn diagram illustrating overlaps between four *Drosophila* datasets: CVM, hemocyte, and two border cell (BCwang, BCborghese) gene expression profiles. For CVM and hemocyte profiles, the short lists described previously were used. For BCwang list, gene names were taken from Table S1 (Wang et al., 2006) "Genes enriched in the migratory border cells". The BCborghese list was extracted from Table S1 (Borghese et al., 2006) "Genes significantly up-regulated in the WT border cells compared to follicle cells". (B) The list genes in overlaps. (C) A Venn diagram showing overlaps in gene expression profiles across species: CVM, hemocyte, and chick neural crest cells. (D) The list of genes that are common in all three dataset (CVM, hemocyte, and chick neural crest cells), which are categorized into three functional groups.

A

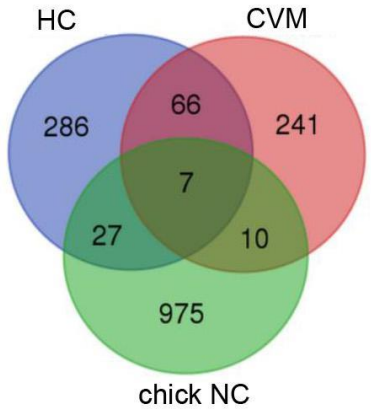


B

Genes in overlap

CVM + HC + BCwang	CVM + HC + BCborghese	CVM + BCwang + BCborghese	HC + BCwang + BCborghese
8	8	1	1
w, anxB9, nemy, nrv1, CG5002, CG5973, Mnd, MESR3	*CG9629, Tig, CG8468, Hsp23, Hsp26, CG3246, GILT2, CG5958*	***neo***	***Sn*** *(=fascin)*

C



D

Overlap between CVM + HC + chick NC (n=7)

	Gene name	Function	Closest human orthologs
ECM-related	*spn27A*	Serine-type endopeptidase inhibitor activity	SERPINA11
	Mmp2	Matrix metalloproteinase	MMP17
	Vkg	Extracellular matrix (collagen)	COL4A5
	Crag	Calmodulin-binding protein related to a Rab3 GDP/GTP exchange protein	DENND4A
Amino acid metabolism	*Oat*	Ornithine aminotransferase precursor	OAT
oxidation-reduction	*CG9629*	Oxidation-reduction process	ALDH7A1
	CG8080	NAD metabolic process	NADK2

Figure 8. Tissue-specific RNAi against *zfh1* and *neo* results in CVM and BC migration defects. (A-E) Immunostained embryos expressing the GV2 reporter. In WT embryos, the CVM cells undergo synchronous migration as two closely associated yet dynamic clusters (A,D). RNAi-mediated ablation of *zfh1* specifically in the CVM results in a loss of directionality and synchrony between the two migrating groups of cells (B). Tissue-specific knockdown of *neo* via RNAi results in a similar asynchronous migration phenotype (E). *In situ* hybridization confirms expression of *zfh1* (C) and *neo* (F) in the CVM. (G,H) Immunostained egg chambers stained against GFP (green), Lamin (blue), and Phalloidin (red) to mark the BCs, nuclear membrane, and cell membranes, respectively. In stage 10 control egg chambers, BCs migrate through the nurse cells to the periphery of the developing oocyte (G). In contrast, tissue-specific expression of a *neo* RNAi construct in the BCs via the *slbo*-GAL4 driver often results in a failure to migrate to the oocyte periphery (H; Sup. Figure 4). (I-K) Temporal color-coded projections of CVM migration in control (A), *5053*-GAL4>*zfh1* RNAi (B), and *5053*-GAL4>*neo* RNAi (C) embryos expressing the HC2 reporter. Each projection is compiled from 80 movie stills of time points taken every 3 minutes over a 4-hour span (A',B'C'), with each still assigned a unique color code corresponding to a specific time point. (I,I') In WT embryos, CVM cells migrate in a closely associated yet dynamic fashion. In contrast, tissue-specific knockdown of *zfh1* in the CVM via expression of a hairpin construct using the *5053*-GAL4 driver results in reduced cohesion within each migrating cohort, such that individual cells wander and approach each other more closely at the midline (J,J'), while RNAi-mediated knockdown of *neo* in the CVM results in stalling, dysregulated cell division, and concomitant asynchronous migration of the two groups of cells (K,K').

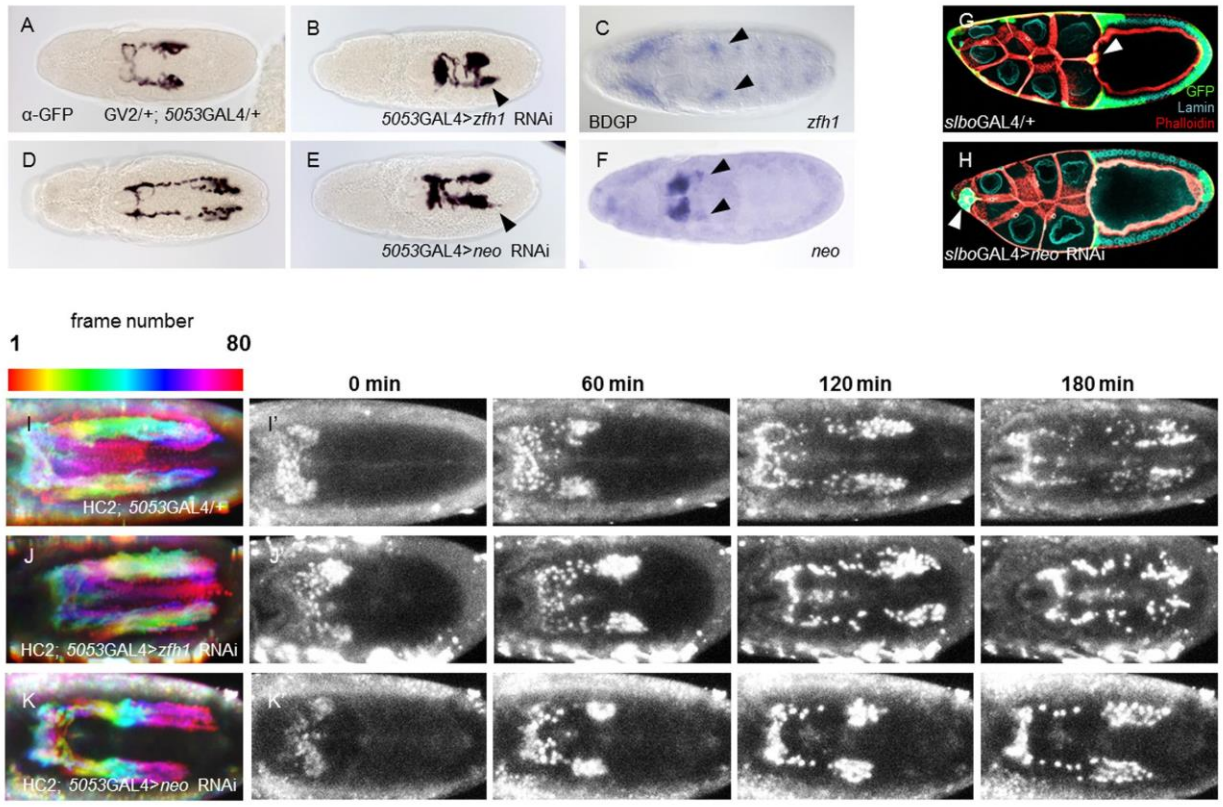


Table 1. Genes enriched in CVM cells with confirmed expression

Flybase gene_id	Gene name	Biological & Molecular Process	RPKM in CVM	RPKM in Non-CVM	Fold change	Expression pattern, confirmed
FBgn0022740	*HLH54F*	visceral muscle development, transcription factor	960.7	14.6	65.8	BDGP, Figure 2 (Georgias et al., 1997)
FBgn0036899	*tey*	negative regulation of transcription, DNA-templated	631.9	2.8	228.9	BDGP, Figure 2
FBgn0032683	*kon*	positive regulation of filopodium assembly; muscle organ development	359.9	20.4	17.6	BDGP, Figure 2 (Schnorrer et al., 2007)
FBgn0011656	*Mef2*	muscle organ development; transcriptional activator activity	295.5	89.1	3.3	BDGP, (Cripps et al., 1999)
FBgn0004456	*mew*	cell adhesion; cell migration	166.9	43.3	3.9	BDGP, Figure 2 (Urbano et al., 2011)
FBgn0034275	*CG5002*	sulfate transport	160.1	18.1	8.8	BDGP
FBgn0038494	*beat-IIb*	heterophilic cell-cell adhesion via PM cell adhesion molecule	141.5	1	145.1	BDGP, Figure 2
FBgn0036564	*Taspase1*	endopeptidase activity	133.2	24.8	5.4	BDGP, Figure 2
FBgn0004606	*zfh1*	transcription factor activity	128.5	25.9	5	BDGP (Lai et al.), Figure 2
FBgn0037835	*CG14687*	myosin light chain binding	109.2	37	3	BDGP (early)
FBgn0038498	*beat-IIa*	heterophilic cell-cell adhesion via PM cell adhesion molecule	108.5	0.9	119.5	BDGP, Figure 2 (Ismat et al., 2010)
FBgn0031313	*CG5080*	unknown	103.3	7.7	13.3	BDGP, Figure 2
FBgn0035956	*Doc2*	transcription factor activity	98.6	23.2	4.2	BDGP, Figure 2, (Ismat et al., 2010)
FBgn0010389	*htl*	fibroblast growth factor-activated receptor activity	76.4	20	3.8	BDGP, (Mandal et al., 2004), Figure 2
FBgn0004885	*tok*	metalloendopeptidase activity	74.7	5.5	13.6	Figure 2
FBgn0029830	*Grip*	muscle attachment	63.3	5.7	11.2	BDGP, Figure 2
FBgn0034135	*Syn2*	structural constituent of muscle	11.7	0.18	65.4	BDGP, Figure 2

Selected sixteen CVM enriched genes from CVM short list to show only those with confirmed positive expression in CVM. The genes were sorted by the RPKM (Reads Per Kilobase of transcript per Million mapped reads) of CVM marker positive population. BDGP refers to a database supported by Berkeley Drosophila Genome Project (Tomancak et al., 2002; Tomancak et al., 2007), PM : plasma membrane

Table 2. Genes enriched in hemocytes with confirmed expression

Flybase gene_id	Gene name	Biological & Molecular Process	RPKM HC	RPKM NON-HC	Fold change	Expression Pattern
FBgn0260011	*nimC4*	phagocytosis, engulfment	4099.8	87.9	46.64	Figure 6
FBgn0033367	*PPO2*	defense response to fungus and bacteria	3607.9	9.9	364.51	BDGP (cc),(Ferjoux et al., 2007)
FBgn0025682	*scf*	chromatin organization; positive regulation of transcription, DNA-templated	3435.5	228.1	15.06	BDGP (pc, cc), (Kobayashi et al., 1998),Figure 6
FBgn0261362	*PPO1*	defense response to fungus and bacteria	2027.3	4.7	434.12	BDGP (cc)
FBgn0010470	*Fkbp13*	imaginal disc development; regulation of Notch signaling pathway	2022.1	212.2	9.53	BDGP (pc)
FBgn0015221	*Fer2LCH*	cellular iron ion homeostasis; response to fungus	1742.4	397.2	4.39	BDGP (cc, pc)
FBgn0003137	*Ppn*	extracellular matrix organization	1702.0	12.2	139.08	cc, pc (Kramerova et al., 2003)
FBgn0026084	*cib*	brain development; actin filament organization	1516.1	189.2	8.01	BDGP (pc)
FBgn0036121	*CG6310*	NA	1339.7	26.6	50.33	Figure 6
FBgn0030245	*CG1637*	acid phosphatase activity	1269.1	106.1	11.96	BDGP (pc)
FBgn0030955	*CG6891*	actin binding	1133.1	147.6	7.68	BDGP (head meso)
FBgn0030796	*CG4829*	neuron projection morphogenesis	1065.8	26.6	40.05	BDGP (pc)
FBgn0015222	*Fer1HCH*	cellular iron ion homeostasis; response to fungus	1026.9	224.2	4.58	BDGP (cc)
FBgn0003447	*sn/fascin*	hemocyte migration; actin filament organization	966.7	47.0	20.59	BPGP (pc), Figure 6 (Zanet et al., 2009)
FBgn0003067	*Pepck*	gluconeogenesis	886.7	51.5	17.21	BDGP (pc), Figure 6
FBgn0040398	*CG14629*	NA	877.7	49.6	17.69	BDGP (cc, pc)
FBgn0044049	*Ilp4*	insulin receptor signaling pathway; larval feeding behavior	844.5	120.7	7.00	BDGP (cc, pc), Figure 6

Selected HC enriched genes from HC short list to show only those with confirmed positive expression in HC. The genes were sorted by the RPKM (Reads Per Kilobase of transcript per Million mapped reads) of HC marker positive population. BDGP refers to a database supported by Berkeley Drosophila Genome Project (Tomancak et al., 2002; Tomancak et al., 2007), pc: plasmatocytes; cc: crystal cells; NA: not available

Highlights

• RNA profiling was performed for two migrating cell types from the *Drosophila* embryo

• Genes specific to caudal visceral mesoderm (CVM) cells or hemocytes (HCs) were found

• Comparison of CVM and HC datasets uncovered limited overlap: 73 genes including zfh1

• Genes also shared by *Drosophila* border cells or chick neural crest were identified

• Neyo regulates migration of CVM and *Drosophila* border cells