In both metazoan development and metastatic cancer, migrating cells must carry out a detailed, complex program of sensing cues, binding substrates, and moving their cytoskeletons. The linker cell in Caenorhabditis elegans males undergoes a stereotyped migration that guides gonad organogenesis, occurs with precise timing, and requires the nuclear hormone receptor NHR-67. To better understand how this occurs, we performed RNA-seq of individually staged and dissected linker cells, comparing transcriptomes from linker cells of third-stage (L3) larvae, fourth-stage (L4) larvae, and nhr-67-RNAi–treated L4 larvae. We observed expression of 8,000–10,000 genes in the linker cell, 22–25% of which were up- or down-regulated 20-fold during development by NHR-67. Of genes that we tested by RNAi, 22% (45 of 204) were required for normal shape and migration, suggesting that many NHR-67–dependent, linker cell-enriched genes play roles in this migration. One unexpected class of genes up-regulated by NHR-67 was tandem pore potassium channels, which are required for normal linker-cell migration. We also found phenotypes for genes with human orthologs but no previously described migratory function. Our results provide an extensive catalog of genes that act in a migrating cell, identify unique molecular functions involved in nematode cell migration, and suggest similar functions in humans.

Functional transcriptomics of a migrating cell in Caenorhabditis elegans

Erich M. Schwarz1, a,b, Mihoko Kato1, a,b, 2 and Paul W. Sternberg3, a, b

1Division of Biology and 2Howard Hughes Medical Institute, California Institute of Technology, Pasadena, CA 91125

Edited by Ruth Lehmann, New York University Medical Center, New York, NY, and approved August 24, 2012 (received for review February 22, 2012)

Cell migration is pervasive in animal development and is also an unwanted part of human cancer. Several different varieties of cell migration, ranging from social protozoa to the human immune system, have been extensively studied (1), but there remains a need for detailed analysis of the different components of migration in a simple cell type that can be exhaustively probed through anatomy, genetics, and genomics. We have undertaken to develop the linker cell (LC) of Caenorhabditis elegans as such a model. The LC is a single cell, attached to the front of a proliferating male gonad, which carries out a stereotypic, long-range migration to generate the shape of the mature gonad (Fig. 1). During 24 h of the second to fourth larval stages (L2 through L4), the LC moves anteriorly along the ventral body wall, turns dorsally, proceeds posteriorly, and requires the nuclear hormone receptor NHR-67. To better understand how this occurs, we performed RNA-seq of individually staged and dissected linker cells, comparing transcriptomes from linker cells of third-stage (L3) larvae, fourth-stage (L4) larvae, and nhr-67-RNAi–treated L4 larvae. We observed expression of 8,000–10,000 genes in the linker cell, 22–25% of which were up- or down-regulated 20-fold during development by NHR-67. Of genes that we tested by RNAi, 22% (45 of 204) were required for normal shape and migration, suggesting that many NHR-67–dependent, linker cell-enriched genes play roles in this migration. One unexpected class of genes up-regulated by NHR-67 was tandem pore potassium channels, which are required for normal linker-cell migration. We also found phenotypes for genes with human orthologs but no previously described migratory function. Our results provide an extensive catalog of genes that act in a migrating cell, identify unique molecular functions involved in nematode cell migration, and suggest similar functions in humans.

Results

RNA-seq of Individual Migrating Linker Cells. We harvested individual yellow fluorescent protein (YFP)-labeled LCs for RNA-seq by cutting them free of the gonad with a laser microbeam, opening the midbody (14), and pipetting individual LCs (Fig. 1B). We ran single-cell RT-PCRs (13) on wild-type LCs from the mid-L3 and mid-L4 larval stages, as well as mid-L4 LCs with RNAi-inactivated nhr-67 (3). As a control for housekeeping genes (15) and genes expressed in other differentiated cell types (16), we also amplified RNA from whole mixed-stage hermaphrodites (primarily larvae). To minimize variable amplification of transcripts from five individual cells, we initially pooled aliquots of RT-PCRs from each cell type before sequencing. To determine cell-to-cell variability, we later sequenced RT-PCRs of each isolated cell individually and to greater depth. Gene expression was measured as reads per kilobase of exon model per million mapped reads (RPKM; ref. 17).

We detected 20-fold expression for 8,000–10,000 C. elegans genes in wild-type L3- and L4-stage LCs (Fig. 2, and SI Appendix, Fig. S1 and Tables S1–S5). In particular, 8,011 genes showed expression in RNA-seq of pooled L3- and L4-stage LCs (SI Appendix, Table S3). This dataset, which was acquired first, largely overlapped with our later data from individual LCs (with 7,915 genes redetected in the latter dataset) and was used in most analyses below. A total of 10,064 genes were detected in all wild-type L3- or L4-stage LCs, pooled and individual. In any individual L3- or L4-stage LC, we generally detected fewer genes on average (mean, 3,531 and 5,083) than in the L3- or L4-stage pools (5,740 and 6,603). Conversely, we detected more expressed genes in the aggregated, deeper RNA-seq data from 10 individual LCs (6,528 in five L3s; 9,153 in five L4s; 9,968 in 10 L3s and L4s) than in RNA-seq of their pools. RNA-seq of any individual cell was thus less sensitive than that of pools, which in turn was less sensitive than aggregating data from populations of border cells in Drosophila (7, 8) and heart cells in Ciona (9). More recently, RNA-seq (10) and functional genomics (11) have been used to identify mammalian genes involved in the transition to invasive mesenchymal cell types. In C. elegans, it should be possible to combine such RNA-seq with single-cell analysis to achieve a comprehensive inventory of genes involved in migration, with precise temporal resolution. Because the LC is a solitary and highly dynamic postembryonic cell, neither embryonic cell cultures nor cell-specific pulldown from mass worm cultures (12) were promising approaches to obtaining LCs in bulk. We thus sequenced single-cell RT-PCRs (13) from individually dissected wild-type LCs from the mid-L3 and mid-L4 larval stages, along with mid-L4-stage LCs undergoing RNAi of nhr-67 (Fig. 1B and C; ref. 3).

Author contributions: E.M.S. and M.K. designed research; E.M.S. and M.K. performed research; E.M.S., M.K., and P.W.S. analyzed data; and E.M.S., M.K., and P.W.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequence reported in this paper has been deposited in the GenBank SRA database (accession no. SRA058596).

1E.M.S. and M.K. contributed equally to this work.

2To whom correspondence should be addressed. E-mail: pws@caltech.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1203045109/-/DCSupplemental.
several individual cells. Gene-expression values for the mean individual L3, L4, and nhr-67(RNAi) LCs correlated well with expression values from pooled LCs of the same type ($r^2 = 0.56–0.93$; SI Appendix, Table S6). Correlation for mean individual LCs vs. pooled whole larvae was lower ($r^2 = 0.15–0.20$), as expected.

Although we detected expression levels from $1.2 \times 10^{-1}$ to $9.8 \times 10^5$ RPKM in pooled LCs, only 208 genes (2.6%) had $\geq 10^4$ RPKM, and almost all of these were ubiquitously expressed housekeeping genes (Fig. 2A). The LC position and gonad shape during LC migration. The LC (red), located at the tip of the gonad (blue), begins migrating toward the head, performs a U-turn while switching to the dorsal bodywall, migrates posteriorly and turns downward, and continues to migrate toward the tail until stopping at the cloaca (blue line). The bottom panel shows LC migration in an nhr-67(RNAi) animal at the same L4 stage timepoint as the wild-type animal. Stars indicate the types of LCs collected for profiling. (B) A YFP-labeled LC is dissected from an animal glued to an agar pad by extruding the gonad with a cutting needle and collecting the LC into a patch pipet. (Scale bar, 20 μm.) (C) Three classes of LC microdissected for RT-PCR and RNA-seq. Tailgonds are outlined in yellow. The nhr-67(RNAi) L4-stage LC is morphologically retarded, still showing the spheroid shape normally seen in L3-stage LCs. (Scale bar, 20 μm.) (D) The RNA-seq profile for nhr-67 showed tissue-specificity of transcripts and efficacy of RNAi. Our RT-PCR protocol (13) deliberately limited amplification to the 3' exons of genes.

Fig. 1. LC biology, dissection, and transcriptional profiling. (A) The LC position and gonad shape during LC migration. The LC (red), located at the tip of the gonad (blue), begins migrating toward the head, performs a U-turn while switching to the dorsal bodywall, migrates posteriorly and turns downward, and continues to migrate toward the tail until stopping at the cloaca (blue line). The bottom panel shows LC migration in an nhr-67(RNAi) animal at the same L4 stage timepoint as the wild-type animal. Stars indicate the types of LCs collected for profiling. (B) A YFP-labeled LC is dissected from an animal glued to an agar pad by extruding the gonad with a cutting needle and collecting the LC into a patch pipet. (Scale bar, 20 μm.) (C) Three classes of LC microdissected for RT-PCR and RNA-seq. Tailgonds are outlined in yellow. The nhr-67(RNAi) L4-stage LC is morphologically retarded, still showing the spheroid shape normally seen in L3-stage LCs. (Scale bar, 20 μm.) (D) The RNA-seq profile for nhr-67 showed tissue-specificity of transcripts and efficacy of RNAi. Our RT-PCR protocol (13) deliberately limited amplification to the 3’ exons of genes.

Functional Validation by RNAi. To validate the LC genes and find new genes required for LC migration, we inactivated 204 LC-enriched genes with postembryonic RNAi (Fig. 3 and SI Appendix, Table S8), and looked for defects in the L4 stage near the end of LC migration. We defined the overall LC-enrichment of a gene as being the ratio of its highest wild-type LC expression (at either L3 or L4 stages) to its expression in whole larvae; 1,097 (14%) of LC genes had a LC/larval ratio of $\geq 20$ (Fig. 2B and SI Appendix, Table S3). For 405 of these genes, preferential LC expression was highly significant (false-discovery rate = 0.01; Dataset S1; ref. 18).

For RNAi, we selected LC-enriched genes, the products of which might be biologically interesting: transcription factors; unambiguous orthologs of human genes, particularly disease-associated genes; highly conserved genes of unknown function; and otherwise nondescript genes whose LC-enriched expression was strikingly high. RNAi gave phenotypes for 45 genes (22% of genes assayed; Fig. 3 and SI Appendix, Table S9). The frequency of phenotypes was only moderately dependent on which genes were chosen as targets: it was 17% (19 of 111) when transcription factors and genes with strict human orthologs were removed.
from the list, and 17% (2 of 12) among genes lacking known homologies or protein domains. Phenotypes included complete failure of the LC to develop, abnormal LC shapes, abnormal migratory paths, delayed migration, LCs detaching from their gonad, and stretched-out gonads (Fig. 3 and SI Appendix, Table S9). Delayed migration and abnormal shapes were the most prevalent phenotypes, seen for 16 and 39 genes respectively; RNAi of single genes usually caused multiple phenotypes, with only three genes having a single phenotype. Because 16 genes with LC defects had no phenotypes in previous genome-wide RNAi screens (Dataset S1), their defects were likely to be LC-specific rather than pleiotropic.

Some genes with RNAi phenotypes had homologs required for cell migration in mammals or Drosophila. Among these were two transcription factors: hih-8, encoding a helix–loop–helix (HLH) protein orthologous to TWIST, which regulates cellular movements in Drosophila and mammals (19, 20); and crh-2, paralogous to SLBO, a C/EBP protein required for Drosophila border cell migration (7, 8). Others genes were regulators of the actin cytoskeleton, such as toca-1 and infr-1/inverted formin (21, 22). RNAi against infr-1 caused abnormal cell shapes, which paralleled the mutant phenotype of infr-1’s mammalian ortholog INF2 (23). Finally, spik-1/sphingosine kinase and lim-9 were required for normal LC migration. Sphingosine kinase participates in several instances of vertebrate cell migration (24), lim-9’s mammalian ortholog FHIL2 inhibits sphingosine kinase (25) and is involved in colon cancer invasion (26) and dendritic cell migration (27).

However, we also found migration phenotypes for highly conserved genes with no previously known role in cell migration, such as tep-1, a seven-transmembrane receptor whose mammalian ortholog TPRA40 inhibits embryonic cell division (28), and maea-1, whose mammalian ortholog MAEA/EMP is required for developing erythroblasts to bind macrophages and mature into erythrocytes (29). Other unexpected migration phenotypes arose...
from structural maintenance of chromosome (SMC) genes, the products of which mediate both chromosome segregation and long-distance gene regulation (30). RNAi against him-1/SMC1, smc-3/SMC3, and smc-4/SMC4 induced the LC to separate from the gonad and migrate without it. Because these three genes are expressed both in the LC and in somatic gonadal cells (SI Appendix, Table S9), their products are likely to be required in either the LC or the somatic gonad for them to bind one another.

**Cis-Regulatory Gene Assays.** We made YFP reporters for the proximal promoter regions of 21 genes, and observed LC expression for 11 (SI Appendix, Tables S9 and S10). Because the transgenic reporters were driven solely by 5'-ward flanking regulatory sequences and not by intronic or 3'-ward ones, the expression frequency of 11 of 21 represents a lower limit on LC expression of these genes in vivo. Four genes showed YFP expression in LCs but not RNAi phenotypes, and five showed the reverse; however, six genes showed both LC expression (via transgene or antibody staining) and RNAi phenotypes, suggesting that the RNA-seq data reflect in vivo function in LCs. These six genes included arx-7, him-1, msp-3, smc-4, sphk-1, and srsx-18.

One example of an NHR-67–regulated gene is srsx-18, which encodes a putative G protein-coupled receptor (Fig. 4). In our RNA-seq data, srsx-18 is transcriptionally active solely in wild-type L4-stage LCs, not in L3-stage LCs or in nhr-67(RNAi) L4-stage LCs. This L4-stage expression is regulated by nhr-67. (C and D) srsx-18::YFP is absent in the L3-stage but expressed in the L4-stage LC. (E) srsx-18::YFP expression in the L4-stage LC is abolished by nhr-67(RNAi). Other gene expression reporters are described in SI Appendix, Table S10.

**Biological Functions Encoded by the LC Transcriptome.** To find biological functions preferentially expressed by the LC transcriptome, we searched for Gene Ontology (GO) terms statistically over-represented among genes with high LC-enrichment or among genes strongly up-regulated in the L4 stage (31). Functions enriched in highly LC-enriched genes included transcriptional regulation, cytoskeletal protein binding, intracellular protein...
transport, cell adhesion, G protein-coupled receptor signaling, cholinergic synaptic transmission, and axon components (SI Appendix, Table S11 and Dataset S1). Transcription factors included the Eda/daughters ortholog hli-2, hli-8, and hli-19. HLH-8 and HLH-19 both bind HLH-2 in two-hybrid assays (32); because we observed expression of hli-19 and hli-8 at the L3 and L4 stages respectively, their protein products could be successive heterodimeric partners of HLH-2 in the developing LC. One LC gene associated with G protein-coupled receptor signaling and required for normal migration was srx-18 (Fig. 4 and SI Appendix, Table S9). LC genes associated with synaptic transmission included neurotransmitter receptors, which are generally considered neuronal; however, YFP reporters for the acetylcholine receptor genes acr-16 and gar-3 and the glutamate receptor gfr-2 were expressed in the LC (SI Appendix, Table S9). An approximate subset of axon components were protein kinases, such as cam-1, pak-1, sax-1, unc-51, and vab-1, which are required for normal migration or morphology of axons and neurons (33–37).

Genes up-regulated by NHR-67 from L3- to L4-stage LCs disproportionately encoded protein kinases and phosphatases, regulators of muscle contraction, and potassium channels (Fig. 2D, SI Appendix, Table S12, and Dataset S1). Genes associated with regulating muscle contraction included the myosin regulatory light-chain gene acr-16 and the tropomyosin msp-2 (SI Appendix, Dataset S1); their up-regulation by NHR-67 is consistent with the observed changes of LC shape and size in L4 larvae (Fig. 1A and C).

Among the genes preferentially up-regulated at the L4 stage were a set of 389 sperm-enriched genes, identified by Reinke et al. (38), which we also detected in wild-type LCs and found were disproportionately common among strongly up-regulated genes (SI Appendix, Fig. S2 and Table S13). A subset of these genes (Fig. 2C and SI Appendix, Fig. S2) encoded major sperm proteins (MSPs), motor proteins required for amoeboid movement of nematode sperm (39). LCs also expressed a broader class of MSP-domain (MSD) genes (Fig. 2C), the Ascaris suum homologs of which include regulators of MSP polymerization (39). Both sets of genes were much more strongly up-regulated in wild-type L4-stage LCs than in nhr-67(RNAi) ones (Fig. 2C and SI Appendix, Fig. S2 and Table S13). Although we cannot fully exclude the possibility that sperm-enriched genes detected in LCs arose from contamination during dissection of the male gonad, we can rule out sporadic contamination (because we detected them in five individual wild-type L4-stage LCs), and we reproducibly observed transcriptional up-regulation in L4-stage LCs (Dataset S1). We observed msp-3/YFP expression in the LC and distal tip cells, but not in other gonadal cells (SI Appendix, Fig. S3 and Table S9). RNAi against the most strongly expressed MSP gene, msp-3, produced abnormally round and delayed L4-stage LCs (SI Appendix, Fig. S3 and Table S9).

We tested five tandem pore potassium channel (twk) genes with RNAi, because 13 of 16 potassium channels expressed in the LC were from this class (40). Two twk genes yielded migration defects, in which the LC was abnormally slow to turn back from the dorsal to the ventral body wall (SI Appendix, Table S9). This finding was striking because C. elegans has 46 paralogous twk genes, which we had expected to be individually redundant. It is also notable that delayed ventral return is one of several defects seen for nhr-67 (RNAi) LCs (3). These RNA-seq and RNAi data suggest that twk channels are significant effectors of NHR-67 in the migrating LC.

Discussion

We have used single-cell transcriptional profiling to capture the gene-expression profile of a single migrating cell required for organogenesis in C. elegans, both at two time points during migration and at a single point defined by the transcription factor NHR-67. We detected the expression of 5,000 to 7,000 genes for each assay of pooled LCs, and a total of 8,000 to 10,000 wild-type LC genes. All of these disparate genes can be genetically perturbed, and their interactions studied, within the LC (Fig. 5). We found gene regulation to be highly dynamic during the course of cell migration, with half of the genes showing 20-fold changes of expression between two larval stages that are 10 h apart. Using both RNAi and GO-term enrichment analysis on the LCs assayed, we were able to identify several different classes of genes required for migration, including tandem pore potassium channels.

MSPs are best known for acting in nematode sperm cells, which use MSRs rather than actin for amoeboid motility (39). We observed expression of msp-3 in L4-stage LCs both with RNA-seq and with YFP transgenes, along with msp-3(RNAi) phenotypes in migrating LCs. A previous screen for genes required in axonal migration found one MSP domain-encoding gene (41). These data are consistent with the hypothesis that MSPs are not restricted to sperm cells, but also function in somatic cell and axonal migration. Such a function might explain why MSPs are conserved in parthenogenetic nematodes that lack sperm (42).

For many genes required in LC migration, their roles may be conserved in other organisms. In the case of SMCs, mutation of cohesion accessory factors can impair neuronal and axonal migration in C. elegans and mice (30). Non-twk potassium channels are required for mammalian cell migration (43), and mammalian TWK channels are activated by both chemical and mechanical stimuli (44); mammalian TWK channels, like their C. elegans orthologs, could also have a migratory function. Similarly to NHR-67 in the LC, regulators of the actin and microtubule cytoskeleton are transcriptionally up-regulated by SLBO in migrating Drosophila border cells (7, 8). Although mammals do not have an exact equivalent to the MSP family, they have two genes encoding MSP domains (45), MOSPD1 and MOSPD3, with one isoform of MOSPD1 predicted to be cytoplasmic. The LC allows migration genes and their interplay to be studied within a single, precisely timed migrating cell, and may reveal novel effectors of cell migration in metazoans.

Materials and Methods

The strains PS4730 [lys12; unc-119(+);100 µg/mL] and PS4864 [lys12; unc-119(+);100 µg/mL] are used for LC dissections and for RNAi assays of LC migration. Feeding RNAi was carried out as in ref. 3. Worms were scored for LC defects by Nomarski microscopy, allowing more detailed screening than dissection microscopy previously used (3). Dissected cells were subjected to 3′-tailed RT-PCR by the method of ref. 13. RNA-seq was performed as in ref. 17; reads were mapped to the W5190 sequence of the C. elegans genome with bowtie (46); and RPKM counts were computed with ERANGE 3.1 (17). GO terms associated with LC-enriched genes were identified with Func (31). Further details are given in the SI Appendix, Materials and Methods.

ACKNOWLEDGMENTS. We thank M. Goodman, M. Chaffie, and A. Mortazavi for early support in developing single-cell RT-PCR; B. Williams for providing control spike poly(A) RNA; L. Schaeffer, D. Trout, and F. Antoshechkin of the Jacobs Genome Center for Illumina sequencing; A. S. Nasrallah for help setting up the dissection rig; J. Downes for scoring linker cell migration phenotypes; J. Thomas for his list of transcription factors; WormBase for setting up the dissection rig; and the Howard Hughes Medical Institute, with which P.W.S. is an Investigator.