Genome-wide view of cell fate specification: *ladybird* acts at multiple levels during diversification of muscle and heart precursors

Guillaume Junion,1 Laetitia Bataillé, Teresa Jagla, Jean Philippe Da Ponte, Romain Tapin, and Krzysztof Jagla2

Institut National de la Santé et de la Recherche Médicale (INSERM) U384, 63000 Clermont-Ferrand, France

Correct diversification of cell types during development ensures the formation of functional organs. The evolutionarily conserved homeobox genes from *ladybird/Lbx* family were found to act as cell identity genes in a number of embryonic tissues. A prior genetic analysis showed that during *Drosophila* muscle and heart development *ladybird* is required for the specification of a subset of muscular and cardiac precursors. To learn how *ladybird* genes exert their cell identity functions we performed muscle and heart-targeted genome-wide transcriptional profiling and a chromatin immunoprecipitation (ChIP)-on-chip search for direct Ladybird targets. Our data reveal that *ladybird* not only contributes to the combinatorial code of transcription factors specifying the identity of muscle and cardiac precursors, but also regulates a large number of genes involved in setting cell shape, adhesion, and motility. Among direct *ladybird* targets, we identified *bric-a-brac 2* gene as a new component of identity code and *inflated* encoding αPS2-integrin playing a pivotal role in cell–cell interactions. Unexpectedly, *ladybird* also contributes to the regulation of terminal differentiation genes encoding structural muscle proteins or contributing to muscle contractility. Thus, the identity gene-governed diversification of cell types is a multistep process involving the transcriptional control of genes determining both morphological and functional properties of cells.

[Keywords: Cell fate; microarray; ChEST; *Drosophila; ladybird; muscle; heart*]

Supplemental material is available at http://www.genesdev.org.

Received April 17, 2007; revised version accepted October 4, 2007.

Diversification of cell types is crucial for correct progression through development. Genetic studies highlight the essential role of the regulatory inputs, which set up the cell fate-specifying combinatorial code of transcription factor activities, the identity gene code (Carmena et al. 1998; Halfon et al. 2000; Jagla et al. 2002; Zaffran and Frasch 2002, Liu et al. 2006; Estrada et al. 2006). The identity gene code is commonly used to ensure diversification of cells in a broad range of developing tissues and metazoan organisms (Briscoe et al. 2000; Jagla et al. 2002; Buckingham 2003; De Graeve et al. 2004; Certel and Thor 2004). However, our understanding of the global gene expression program that operates downstream from the identity gene code and leads to the acquisition of a given cell fate remains very limited. A large number of key transcriptional regulators acting as cell fate-specifying genes have been identified in *Drosophila* [for review, see Skeath 1999; Baylies and Michelson 2001; Riechmann and Ephrussi 2001; Fichelson et al. 2005; Olson 2006]. Particularly well suited to studying cell fate acquisition are *Drosophila* somatic muscles and cardiac cells for which the majority of cell types can be identified by specific molecular markers. Prior studies (Xu et al. 1998; Halfon et al. 2000; Lee and Frasch 2000) revealed that an interplay of Wg, Dpp, and RTK signals was required to switch on expression of the selector genes *twist* (twi) and *tinman* (tin), whose activity is essential for the initiation of somatic muscle and cardiac development programs. It has also been demonstrated that the same extrinsic regulatory inputs supplemented by the intrinsic Twi and Tin action activate the cell fate-specifying factors within the cardiac and somatic mesoderm (for review, see Frasch 1999a,b). For example, the regulatory requirements for the specification of a dorsally located DA1 muscle founder and a pericardial cell founder expressing *even-skipped* (*eve*) have been described in detail (Halfon et al. 2000). This led to the identification of *eve* enhancer carrying binding sites for the
effectors of Wg, Dpp, and RTK signaling pathways and for tissue-specific factors Twi and Tin. All these factors are required for the initiation of eve expression in dorsal mesodermal cells, whereas repression by identity genes ladybird (lb) and msh ensures restriction of eve expression to the DA1 muscle founder and two pericardial cells [Jagla et al. 2002]. The finding that a tissue domain is determined by the coordinated action of exogenous signaling cues and endogenous tissue-specific selector genes inspired a CodeFinder approach [Philippakakis et al. 2006]. This approach was recently used to uncover a large number of regulatory modules regulated by the interplay of the signals mentioned above and muscle- or heart-specific factors. The CodeFinder and a few other approaches [Rebeiz et al. 2002; Zhou and Wong 2004] provided the first genome-wide data sets describing the activity of genes during the determination of muscular and cardiac tissue progenitors. Likewise, transcriptional profiling and chromatin immunoprecipitation (ChIP)-on-chip assays have identified a broad set of genes and their regulatory motifs regulated by the key muscle and heart differentiation factor Mef2 [Junion et al. 2005; Sandmann et al. 2006].

Altogether, the recently reported global approaches dedicated to the muscle developmental program have identified clusters of genes expressed specifically in founder cells, fusion-competent myoblasts, and in differentiating muscle fibers. However, no systematic genome-wide analysis has been performed to identify downstream targets of identity genes required for the diversification of muscle and heart cells. To address this issue, we set out to identify genes whose expression was regulated by Lb homeodomain factors known to be required for the specification of segmental border muscle and a small subset of cardiac precursors [Jagla et al. 1997, 1998].

It had been shown previously that ladybird early [lbe] and ladybird late [lbL] genes act as repressors contributing to the cell identity code during Drosophila founder cell diversification [Jagla et al. 1999, 2002]. Importantly, previous genetic analyses revealed that muscle founder cells carry all the information required for correct shape, innervation, and attachment of different muscle fibers strongly suggesting an instructive role of identity genes in these processes. The phenotypes observed in lb mutant embryos [Jagla et al. 1998] suggested a role of lb genes in axial positioning of muscle precursors and in the acquisition of fiber-specific shape and attachment to tendon cells. Interestingly, the vertebrate lb orthologs Lbx1 genes play similar roles during the specification of neural fates [Cheng et al. 2005] and during the setting of properties in mouse, chick, and zebrafish appendicular muscle precursors [Brohmann et al. 2000; Gross et al. 2000; Neyt et al. 2000; Mootoo samy and Dietrich 2002].

Thus, lb/Lbx1 genes emerge as evolutionarily conserved cell identity factors whose ability to specify cell fates is employed in different tissue contexts and in a broad range of animal species. Hence, the unveiling of the global gene expression program operating downstream from lb should provide a general view of how the acquisition of cell identity is executed at the genome level. To approach this issue we designed a targeted transcriptional profiling strategy and used it in combination with the recently described [Junion et al. 2005] ChIP-enriched in silico target [ChEST] technique. Our data show that lb acts at multiple levels and over a long time period. During myogenic determination steps it represses a subset of identity genes and contributes to the cell fate-specifying code of transcription factors. Later, it contributes to the acquisition of individual properties of muscle and cardiac cells by acting as a regulator of genes involved in setting cell shape, cell motility, and function.

Results

Uncovering transcriptional Lb targets using targeted expression profiling

The gene lb, like many other cell identity genes, is expressed in a very restricted population of embryonic cells, making the identification of its transcriptional targets in the whole embryo particularly challenging. In each hemisegment within the somatic and cardiac mesoderm, lb is expressed in one out of 30 precursors of somatic muscles and in four out of 16 cardiac precursors [see Fig. 1A; Jagla et al. 1997, 1998]. Unlike highly restricted spatial expression, timely, lb is expressed during a large time window [see Fig. 1A]. Its expression is initiated at 5 h AEL [after egg laying] when the diversification of mesodermal cell types starts, and is maintained until the late steps of muscle and heart morphogenesis indicating potential role of lb in progressive acquisition of specific muscle and cardiac cell properties. To optimize the identification of mesodermal lb targets we opted to use muscle and heart-targeted overexpression or RNA interference [RNAi]-based attenuation of lb. Thus, to design a targeted expression profiling adapted to muscular and cardiac tissues and to the timing of lb activity we opted for two different mesodermal drivers: (1) the pan-mesodermal 24B-GAL4 driver (H + M-targeting) and the heart-specific tin-GAL4 driver (H-targeting) [see Fig. 1A]. The 24B-GAL4 driver’s activity overlaps with all the steps of heart- and muscle-specific lb expression, whereas the cardiac tin-GAL4 driver is expressed from 9 h AEL onward and targets the later aspects of cardiogenic lb functions. Within the targeted tissues, and especially in the heart, the proportion of lb-positive cells is relatively high (~18% in the cardiac mesoderm), thus increasing the probability of targets identification.

To reduce the number of false positives, we performed three independent RNA isolations from staged 5- to 16-h-AEL wild-type [white] embryos or staged embryos collected from crosses of UAS-lb or UAS-lbRNAi [Maqbool et al. 2006] flies with one of the targeting drivers [Supplementary Fig. S1A]. Thus, for each experimental condition three independent cDNA probes were used and three global GeneChip data sets were generated [Supplementary Fig. S1A]. The data sets [ArrayExpress
accession no. E-TABM-297) were normalized, and classified using SAM analysis [Supplementary Tables S1–S4; see Materials and Methods for details]. Based on this selection, we found 2663 genes with significantly altered transcription levels in H-targeting conditions and 1931 in an H + M-targeting context. Comparison of H and H + M targets revealed that 1118 genes were identified in both targeting contexts (Supplementary Table S5). Within the groups, the candidate genes could be subdivided into those regulated positively and those whose expression was down-regulated by \( lb \). Comparison of candidate genes identified in \( lb \) gain-of-function (GOF) conditions with those from the \( lb \) RNAi conditions revealed that several genes display opposite changes in their transcription levels (see Supplementary Table S6). Moreover, as shown by Venn diagrams [Fig. 1B], an important number of candidate genes appear common when comparing pool of GOF versus wild-type candidates with LOF versus wild-type and with GOF versus LOF candidates. Overlapping populations of candidate genes were normalized, and classified using SAM analysis [Supplementary Tables S1–S4; see Materials and Methods for details]. Based on this selection, we found 2663 genes with significantly altered transcription levels in H-targeting conditions and 1931 in an H + M-targeting context. Comparison of H and H + M targets revealed that 1118 genes were identified in both targeting contexts (Supplementary Table S5). Within the groups, the candidate genes could be subdivided into those regulated positively and those whose expression was down-regulated by \( lb \). Comparison of candidate genes identified in \( lb \) gain-of-function (GOF) conditions with those from the \( lb \) RNAi conditions revealed that several genes display opposite changes in their transcription levels (see Supplementary Table S6). Moreover, as shown by Venn diagrams [Fig. 1B], an important number of candidate genes appear common when comparing pool of GOF versus wild-type candidates with LOF versus wild-type and with GOF versus LOF candidates. Overlapping populations of candidate genes were identified by expression profiling and ChEST strategies. [A] Muscle- and heart-specific expression of \( lb \) and the two drivers (24B > GFP and tin > GFP) used for targeted transcriptional profiling. Three different embryonic stages are presented to illustrate \( lb \) and 24B-GAL4 expression in biological samples used for microarray experiments. In the first column (5–9 h AEL) are shown embryos at stage 11 or early stage 12 [lateral views]. In the middle column (9–12 h AEL) are embryos at stage 13, and in the third column (12–16 h AEL) are shown embryos at stage 15–16. In the middle and right column, muscle (M) expression is illustrated by lateral views and heart (H) expression by dorsal views of embryos. Note that tin-GAL4 driver expression starts later, from stage 13, and persists until the late embryogenesis. To obtain wild-type biological samples for microarray and ChEST experiments, three collections of embryos aged from 5 to 16 h AEL have been used. In a similar manner, F1 embryos from the cross of 24B-GAL4 with UAS-lbe and with UAS-lbRNAi lines were collected and used as H + M GOF and H + M LOF samples, respectively. Embryos from the cross of tin-GAL4 with UAS-lbe or UAS-lbRNAi lines were aged according to tin-GAL4 expression and later labeled H GOF and H LOF, respectively. [B] Venn diagrams showing populations of candidate genes identified by comparison of GOF versus wild-type, LOF versus wild-type, and GOF versus LOF context. Note that in H conditions more candidate genes are common for these three candidate gene populations than in H + M context. [C] Main categories of \( lb \) targets identified by targeted expression profiling. GO-based statistical distribution of genes identified in H + M targeting conditions and H targeting conditions. In both contexts we find an important enrichment for ECM and cytoskeleton components, signaling molecules, proteins involved in proteolysis, ATP binding, and factors carrying a DNA- or RNA-binding domain. Note that >50% of identified genes have no GO annotations and ~17% have a GO, which does not fit into any of the five main categories. [D] Genome scanning strategy for \( lb \)-binding CRMs. Different heart (H) targeting and muscle (M) targeting genome scanning conditions were used [asterisk; see Supplemental Material]. A pool of predicted cardiac and muscular \( lb \)-dependent CRMs was filtered with respect of distance to adjacent genes and their annotations [asterisk; see Supplemental Material]. Selected CRMs were spotted to produce a computed \( lb \)-CRM array. In parallel, a pool of DNA fragments bound in vivo by \( lb \) was isolated by ChIP and used to probe the computed array. Among 74 CRMs spotted, 16 were found enriched in ChIP material. [E] Identification of the in vitro consensus binding site for Ladybird. Random 10-nt-long oligos flanked on 5' and 3' by 20-nt primer-compatible sequences were radiolabeled and used for the SELEX approach. Four cycles of incubation with 6xHis-\( lb \) homeodomain fusion protein followed by the amplification of a selected subset of sequences were applied to select motifs bound preferentially by \( lb \). As shown by shift assays, \( lb \) homeodomain [top panels] and both Lbe and Lbl proteins [bottom panel] recognize sequences containing TAAT and TAAC core motifs, and the deduced consensus sequence is RVYTAAYHAG.
### Table 1. Selected GO-enriched microarray candidate genes and the ChEST-identified direct Lb targets

| Affy id | Gene name | Function | Expression | Fold change | q value (%) | Validation |
|---------|-----------|----------|------------|-------------|-------------|------------|
| 1632744_a_at | if | Integrin receptor, muscle attachment | SM | +1.54 | 5.72 | RT/QRT/Ab |
| 1633428_at | prc | ECM component, cell migration | DV | +2.04 | 2.2 | Ab (n.s.) |
| 1639229_at | vkg | ECM component | DV HEM | +2.12 | 2.61 | QRT |
| 1633286_at | CG3272 | Cell adhesion | meso | +2.04 | 3.85 | ND |
| 1635370_at | Tsp42El | Integral to membrane | DV | +2.32 | 1.82 | ND |
| 1632968_at | CG6053 | Microtubule-based movement | ND | +8.35 | 3.71 | ND |
| 1637232_at | CG17150 | Microtubule-based movement | ND | +6.5 | 3.71 | ND |
| 1627859_at | CG31646 | Cell–cell adhesion signal transduction | ND | +4.17 | 2.61 | ND |
| 1633167_s_at | CG6788 | Cell adhesion molecule | ND | +1.98 | 5.72 | ND |
| 1628671_a_at | sls | Myosin light chain kinase activity, muscle contraction | SM | +2.17 | 4.52 | Ab |
| 1632945_at | Msp-300 | Actin binding, dsRNA binding | DV and SM | +1.93 | 2.18 | RT/Ab |
| 1630261_a_at | up | Troponin binding, troponin complex | SM | +1.72 | 5.72 | RT |
| 1630509_at | Mp20 | Actin binding, muscle contraction | SM | +1.78 | 2.93 | QRT |
| 1624746_at | βTub85D | Microtubule intracellular transport | SM | +4.76 | 1.4 | ND |
| 1626397_at | bt | Muscle contraction kinase activity, constituent of cytoskeleton | SM | +2.32 | 2.11 | ND |
| 1625825_at | SPARC | Calcium ion binding, structural molecule activity | Meso | +1.96 | 1.4 | ND |
| 1629235_s_at | wupA | Actin binding, troponin complex | SM | +1.88 | 2.47 | ND |
| 1629164_at | CG9825 | High-affinity inorganic phosphate:sodium symporter activity | ND | +2.44 | 1.42 | ND |
| 1627067_at | CanB2 | Calcium-binding neurotransmitter | SM | +2 | 1.18 | ND |
| 1630778_at | CG10116 | Lipoprotein lipase | ND | +1.88 | 2.61 | ND |
| 1636378_a_at | Syn | Neurotransmitter secretion | ND | +6.67 | 2.5 | ND |
| 1637674_at | Ace | Neuromuscular synaptic transmission | ND | +5.14 | 2.23 | ND |
| 1625065_s_at | SK | Small conductance calcium-activated potassium channel | ND | +3.22 | 2.11 | ND |
| 1627882_at | dlg1 | Plasma membrane, cell cortex EGFR binding | ND | +2.5 | 1.82 | RT |
| 1634580_at | bsk | MAPK dorsal closure, frizzled, and Jun pathway | ND | +1.92 | 5.72 | RT |
| 1630560_s_at | Gyc 76C | Intracellular signaling cascade | SM | +2.42 | 7.19 | ND |
| 1624389_at | CG7526 | Cell communication transmembrane receptor protein tyrosine kinase activity | SM | +2.5 | 1.82 | ND |
| 1622957_s_at | Rep3 | Apoptosis | ND | +16.67 | 5.09 | ND |
| 1640325_at | CG8641 | RAS small monomeric GTPase | ND | +14.28 | 1.42 | ND |
| 1624636_at | CG5712 | Adenylate cyclase | ND | +10 | 1.18 | ND |
| 1634436_at | Ast-C | Regulation of heart | ND | +3.7 | 1.06 | ND |

*continued on next page*
| Affy id     | Gene name       | Function                                      | Expression | Fold change | q value (%) | Validation |
|------------|-----------------|-----------------------------------------------|------------|-------------|-------------|------------|
| 1630285_at | RhoGAP100F      | GTPase activator intracellular signaling       | ND         | +2.27       | 3.85        | ND         |
| 1635009_at | CG8500          | RAS small monomeric GTPase                    | ND         | +2.27       | 3.22        | ND         |
| 1639502_at | CG8942          | Wnt receptor signaling pathway                | ND         | +2.17       | 1.82        | ND         |
| 1625478_at | Mtn1            | JNK cascade MAPKKK cascade                    | ND         | +1.54       | 2.5         | ND         |
| 1637488_at | Wnt2            | frizzled2 signaling                           | ND         | −1.6        | 5.72        | RT         |
| 1629448_at | rap             | Cyclin catalbolism                            | CB         | −1.55       | 3.22        | ND         |
| 164994_a_at | RacGAP84C       | RAS GTPase activator                          | ND         | −2.94       | 3.49        | ND         |

**DNA and RNA binding**

| Affy id     | Gene name       | Function                                      | Expression | Fold change | q value (%) | Validation |
|------------|-----------------|-----------------------------------------------|------------|-------------|-------------|------------|
| 1637943_at | how             | mRNA binding, mesoderm dev                    | SM DV      | +1.81       | 2.47        | ND         |
| 1627436_s_at | pdp1            | Par-domain protein leucine zipper factor mesoderm dev | SM         | +1.69       | 3.71        | QRT        |
| 1633789_at | Hand            | Regulation of transcription from RNA polymerase II promoter | meso DV Hem | +1.54       | 3.71        | ND         |
| 1624520_a_at | fitz-f1         | Regulation of transcription                    | CB         | +1.67       | 3.22        | ND         |
| 163351_at  | CG10384         | Apoptosis/cell proliferation                   | ND         | +5.88       | 2.23        | ND         |
| 1625880_at | CG7274          | Cell cycle                                    | ND         | +3.33       | 2.61        | ND         |
| 1636710_at | Neu2            | Transcription regulator activity                | ND         | +2.56       | 2.61        | ND         |
| 1627734_at | nau             | bHLH factor                                   | SM         | −2.67       | 2.61        | Ab         |
| 1629368_at | slou            | Homeobox gene cell fate                       | SM         | −1.62       | 1.82        | Ab         |

**Proteolysis**

| Affy id     | Gene name       | Function                                      | Expression | Fold change | q value (%) | Validation |
|------------|-----------------|-----------------------------------------------|------------|-------------|-------------|------------|
| 1635183_at | Spa43Ab         | Serine protease inhibitor                      | ND         | +1.72       | 2.47        | RT         |
| 1626394_at | fan99Fi         | Chymotrypsin                                  | ND         | +21.48      | 10.01       | ND         |
| 1630547_at | CG4723          | Endothelin-converting enzyme                   | ND         | +10         | 2.23        | ND         |
| 1629618_at | CG12256         | Serine-type endopeptidase                     | ND         | +6.26       | 1.32        | ND         |
| 1635966_at | aTry            | Proteolysis and peptidolysis                  | ND         | +2.94       | 1.62        | ND         |
| 1637966_at | CG15253         | Metalloendopeptidase activity                  | ND         | +2.78       | 3.85        | ND         |
| 1629588_at | CG2528          | Prolyl aminopeptidase                         | ND         | −3.36       | 5.72        | ND         |

**ATP binding/mitochondria**

| Affy id     | Gene name       | Function                                      | Expression | Fold change | q value (%) | Validation |
|------------|-----------------|-----------------------------------------------|------------|-------------|-------------|------------|
| 1623770_at | CG1673          | Transaminase activity                         | ND         | +1.59       | 2.47        | RT         |
| 1639825_at | comt            | Golgi organization and biogenesis              | SM DV      | +1.56       | 4.52        | ND         |
| 1635632_at | CG18363         | Mitochondrial membrane                        | ND         | +8.17       | 8.34        | ND         |
| 1629643_at | CG32186         | ATP-binding cassette (ABC) transporter         | ND         | +6.69       | 4.33        | ND         |
| 1640231_a_at | CG8908          | ATP-binding cassette (ABC) transporter         | ND         | +2.53       | 2.5         | ND         |
| 1631731_at | CG1494          | ATP-binding cassette (ABC) transporter         | ND         | −4.82       | 1.36        | ND         |

**(B) Microarray candidate genes from M targeting conditions**

**ECM and cytoskeleton**

| Affy id     | Gene name       | Function                                      | Expression | Fold change | q value (%) | Validation |
|------------|-----------------|-----------------------------------------------|------------|-------------|-------------|------------|
| 1631756_at | CG10859         | Microtubule-based movement                    | ND         | +6.25       | 2.37        | ND         |
| 1623412_a_at | CG17150         | Dynemin ATPase                               | ND         | +4.77       | 7.48        | ND         |
| 1637488_at | Cad88C          | Calcium-dependent cell adhesion               | ND         | +2.37       | 0           | ND         |
| 1629012_at | capt            | Actin polymerization and/or depolymerization  | ND         | +1.69       | 5.46        | ND         |
| 1630304_at | Dhc93AB         | Dynemin complex motor activity                | ND         | −3.33       | 6.29        | RT         |
| 1631881_a_at | sli             | ECM component, receptor binding               | DV and SM  | −3.27       | 5.46        | QRT        |
| 1641442_a_at | trol            | ECM cell-cell adhesion                        | DV         | −2.38       | 3.93        | RT         |
| 1633428_at | prec            | ECM component, cell migration                 | DV         | −2.44       | 0           | Ab (n.s.)  |
| Affy id      | Gene name | Function                                                                 | Expression | Fold change | q value (%) | Validation |
|--------------|-----------|---------------------------------------------------------------------------|------------|-------------|-------------|------------|
| 1639613_at   | shg       | Calcium-dependent cell adhesion, adherens junction                        | DV         | -1.86       | 3.75        | RT         |
| 1634290_a_at | mew       | Integrin complex receptor activity                                         | SM CB      | -1.53       | 5.46        | ND         |
| 1639229_at   | vkg       | ECM component                                                             | DV Hem     | -2.13       | 3.93        | ND         |
| 1637137_a_at | Dhc64C    | Microtubule-based motility                                                | ND         | -5.33       | 3.82        | ND         |
| 1637848_at   | CG6873    | Actin binding                                                             | ND         | -4.38       | 6.35        | ND         |
| 1632164_at   | scrib     | Establishment and/or maintenance of epithelial cell polarity              | ND         | -2.17       | 7.48        | ND         |

**Muscle structure and function**

| Affy id      | Gene name | Function                                                                 | Expression | Fold change | q value (%) | Validation |
|--------------|-----------|---------------------------------------------------------------------------|------------|-------------|-------------|------------|
| 1626452_at   | CG8698    | Muscle contraction                                                        | SM         | -1.8        | 4.6         | RT         |
| 1628671_a_at | sls       | Myosin-light-chain kinase activity, muscle contraction                     | SM         | -1.88       | 3.93        | Ab         |
| 1632945_at   | Msp300    | Actin binding, dsRNA binding                                              | DV and SM  | -2.28       | 4.07        | Ab         |
| 1630261_a_at | up        | Troponin binding, troponin complex                                        | SM         | -1.88       | 0           | RT         |
| 1638818_at   | Mlp60A    | Lim muscle protein, protein binding                                       | SM         | -2.08       | 0           | RT         |
| 1631217_a_at | TpnC73F   | Calmodulin-binding muscle contraction                                     | SM         | -3.45       | 3.93        | ND         |
| 1634436_at   | Ast2      | Regulation of heart contraction rate                                       | ND         | -3.12       | 3.61        | ND         |
| 1630509_at   | Mp20      | Actin binding, muscle contraction                                         | SM         | -2.38       | 6.54        | ND         |

**DNA and RNA binding**

**Cell fate specification**

| Affy id      | Gene name | Function                                                                 | Expression | Fold change | q value (%) | Validation |
|--------------|-----------|---------------------------------------------------------------------------|------------|-------------|-------------|------------|
| 1637278_at   | tin       | Homeodomain transcription factor                                          | DV and SM  | +1.74       | 3.7         | Ab (n.s.)  |
| 1630010_a_at | pnt       | ETS domain transcription factor                                           | DV         | +2.22       | 6.98        | RT         |
| 1629459_at   | Doc3      | Transcription factor                                                      | CB         | +1.63       | 7.48        | ND         |
| 1624060_at   | bab2      | BTB POZ factor                                                            | DV and SM  | -1.92       | 3.75        | QRT/Ab     |
| 1636558_a_at | Abd-A     | Heart proper cell fate commitment                                         | CB         | -1.98       | 4.6         | ND         |

**Others**

| Affy id      | Gene name | Function                                                                 | Expression | Fold change | q value (%) | Validation |
|--------------|-----------|---------------------------------------------------------------------------|------------|-------------|-------------|------------|
| 1636780_at   | meso18E   | Mesoderm development                                                      | SM         | -2.35       | 4.92        | RT         |
| 1632327_at   | Rx        | Paired-like homeobox                                                     | DV         | -7.14       | 3.93        | ND         |
| 1636679_at   | Mrtf      | Mesodermal cell migration, GO:0030036 actin cytoskeleton organization, and biogenesis | DV and SM  | -2.02       | 5.46        | ND         |
| 1631948_s_at | Clk       | bHLH factor                                                              | ND         | -9.46       | 2.49        | ND         |

**Cell communication**

| Affy id      | Gene name | Function                                                                 | Expression | Fold change | q value (%) | Validation |
|--------------|-----------|---------------------------------------------------------------------------|------------|-------------|-------------|------------|
| 1641460_at   | CanB      | Calcium-binding neurotransmitter                                          | ND         | +2.04       | 5.75        | QRT        |
| 1634498_at   | CG17152   | Glutamate-gated ion channel                                               | ND         | +4.65       | 4.65        | ND         |
| 1631846_at   | CG14857   | Organic cation porter                                                    | ND         | -8.01       | 2.18        | ND         |
| 1641495_at   | pHCL-A    | Glycine-gated chloride channel complex                                    | ND         | -6.52       | 4.38        | ND         |

**Rho/PI3K/CDK**

| Affy id      | Gene name | Function                                                                 | Expression | Fold change | q value (%) | Validation |
|--------------|-----------|---------------------------------------------------------------------------|------------|-------------|-------------|------------|
| 1635592_at   | vav       | Rho protein signal transduction, regulation of cell shape                 | ND         | +2.08       | 2.89        | RT         |
| 1641006_s_at | PI3K68D   | Intracellular signaling                                                   | ND         | +2.04       | 3.82        | ND         |
| 1635062_at   | Smg1      | Phosphatidylinositol 3-kinase                                             | ND         | +1.53       | 3.7         | ND         |
| 1625199_s_at | dap       | Cyclin-dependent protein kinase inhibitor                                 | meso       | -2.05       | 4.6         | RT         |
Table 1. [continued]

| Affy id   | Gene name | Function                        | Expression | Fold change | q value (%) | Validation |
|-----------|-----------|---------------------------------|------------|-------------|-------------|------------|
| 1641322_at| Pde9      | eGMP-specific phosphodiesterase | ND         | −5.66       | 4.09        | ND         |
| 1636914_at| Cdk5a     | Cyclin-dependent protein kinase  | ND         | −1.67       | 3.93        | ND         |
| 1640657_at| RhoGAP102A| Rho GTPase activator            | ND         | −1.56       | 7.48        | ND         |

**MAPK signaling**

| Affy id   | Gene name | Function                        | Expression | Fold change | q value (%) | Validation |
|-----------|-----------|---------------------------------|------------|-------------|-------------|------------|
| 1640325_at| CG8641    | RAS small monomeric GTPase      | ND         | +12.5       | 4.09        | ND         |
| 1629678_a_at| Aplip1 | Regulation of JNK cascade          | ND         | +1.58       | 4.86        | ND         |
| 1641253_at| Src64B    | Tyrosine kinase Jun pathway      | SM         | −1.64       | 3.48        | ND         |

**Other signalings**

| Affy id   | Gene name | Function                        | Expression | Fold change | q value (%) | Validation |
|-----------|-----------|---------------------------------|------------|-------------|-------------|------------|
| 1623324_at| lin       | Torso signaling pathway          | meso       | +1.89       | 2.37        | ND         |
| 1627328_at| scw       | BMP receptor signaling pathway   | ND         | +11.11      | 4.92        | ND         |
| 1626723_at| Takr86C   | G-protein-coupled receptor protein signaling pathway | ND         | +10         | 2.37        | ND         |
| 1637339_at| Ccap      | Positive regulation of heart contraction | ND         | +2.43       | 6.29        | ND         |
| 1625636_s_at| pan | Wnt receptor signaling pathway   | DV and SM  | −1.58       | 0           | ND         |
| 1632457_s_at| mam | Mesoderm cell fate determination Notch signaling pathway | meso       | −1.82       | 4.86        | ND         |
| 1628166_a_at| DmsR-2 | G-protein-coupled receptor protein signaling pathway | ND         | −5.82       | 3.39        | ND         |
| 1633461_at| slik      | Receptor signaling protein serine–threonine kinase | ND         | −5.55       | 5.46        | ND         |

**Apoptosis**

| Affy id   | Gene name | Function                        | Expression | Fold change | q value (%) | Validation |
|-----------|-----------|---------------------------------|------------|-------------|-------------|------------|
| 1633498_at| CG7188    | Apoptosis inhibitor             | ND         | +2.08       | 4.07        | ND         |
| 1639071_a_at| Dredd | Apoptosis caspase               | ND         | −1.63       | 5.75        | RT         |
| 1624441_at| CG18389   | Induction of apoptosis by hormones | ND         | −9.1        | 6.29        | ND         |

**Proteolysis**

| Affy id   | Gene name | Function                        | Expression | Fold change | q value (%) | Validation |
|-----------|-----------|---------------------------------|------------|-------------|-------------|------------|
| 1632355_at| lack      | TGFβ pathway ubiquitin ligase   | ND         | +1.85       | 3.75        | RT         |
| 1632993_at| CG6372    | Leucyl aminopeptidase           | ND         | +3.3        | 3.39        | ND         |
| 1631986_at| CG16998   | Serine-type endopeptidase       | ND         | +4.35       | 4.38        | ND         |
| 1624240_at| Uch       | Ubiquitin C-terminal hydrolase  | ND         | +3.7        | 5.46        | ND         |
| 1628288_s_at| Nep4 | Metalloendopeptidase            | DV         | −2.78       | 6.54        | ND         |
| 1626394_at| Iop99Fi   | Chymotrypsin                    | ND         | −10.84      | 5.75        | ND         |

**ATP binding/mitochondria**

| Affy id   | Gene name | Function                        | Expression | Fold change | q value (%) | Validation |
|-----------|-----------|---------------------------------|------------|-------------|-------------|------------|
| 1628659_at| Mdr49     | ATPase activity, coupled to transmembrane movement of substances | meso       | +3.12       | 6.98        | ND         |
| 1629643_at| CG32186   | ATPase activity, coupled to transmembrane movement of substances | ND         | +5.21       | 4.09        | ND         |
| 1634045_at| CG10748   | Tricarboxylic acid cycle, mitochondrial matrix | ND         | −4.54       | 4.38        | ND         |

(C) Candidate genes and CRMs identified by ChEST with H-targeting conditions

| Gene CRM position | Biological function |
|-------------------|---------------------|
| sli I: 11384323–11384549 | Receptor binding/mesoderm migration |
| bab2 I: 1146918–1147918 | Regulation of transcription |
| Dhc93AB I: 16851421–16852003 | Cell motility/microtubule-based movement |
| Dab D: 16689242–16689667 | Receptor binding/sevenless signaling |
| lbe D: 17266296–17266690 | Transcription factor/cell fate specification |
| krt1 Up: 20561047–20561799 | Transcription factor |
| NetA Up: 14488661–14489438 | Cell adhesion/synaptic target attraction |
| if I: 16607638–16607929 | Muscle attachment/myofibril assembly |

continued on next page
genes are comparatively smaller in H + M targeting context than in H targeting conditions, mainly due to the relatively restricted number of candidate genes whose expression was significantly altered in 24B > RNAIlib embryos [LOF vs. wild-type pool].

Main categories of Lb targets identified by expression profiling

We used gene ontology (GO) annotations to classify identified targets and an GO-based analysis (GObar) to identify the significantly enriched GO annotations. In both H + M and H-targeting conditions about one-half of the candidate genes were found to be GO-annotated. This enabled us to subdivide the candidates into six main categories [Fig. 1C] and to identify the overrepresented GO annotations in the gene list (Table 1A,B; Supplementary Fig. S2). The GO-annotated candidate genes with a mesodermal function or mesodermal expression pattern and with a high fold expression change value are listed in Table 1A,B for the entire list of GO-annotated candidates see Supplementary Tables S1–S3). As expected, we identified a group of already known muscle and heart identity genes [e.g., nau, slou, tin, Doc3] as well as potential new components of the cell identity code [e.g., bab2, Pdp1, CG32611]. We observed that Lb regulates the transcription of this category of genes either positively or negatively. The Lb-regulated genes identified by expression profiling also code for the extracellular matrix [ECM]/cell adhesion components [e.g., shg, if, vkg, trol, pcr, sli] and for proteins involved in the remodeling of actin filaments and in cell motility [e.g., Dhc93A, Mp20, Msp300, Pax, Vin]. A subset of candidate genes encoding cytoskeletal components is repressed by Lb in the H context and positively regulated in H + M conditions (Table 1A,B; Supplementary Table S3). To this category belong constituents of the sarcomeric cytoskeleton [e.g., up, Mlp60A, Ptm, sfs] suggesting that their activity contributes to specific properties of Lb-expressing muscles. A large number of GO-enriched identified targets fit into Wnt [e.g., Wnt2, pan, Cby], Rho/Rac [e.g., RhoGAP100F, vav], and RTK [e.g., Jon99Fj, CG15253] signaling pathways or are encoding proteins involved in energetic metabolism [e.g., cont, CG32186] and in protein degradation [e.g., Snp43Ab, gol].

Interestingly, a similar GO-based distribution of Lb targets was found in H and in H + M targeting conditions [Fig. 1C], indicating that the acquisition of heart and muscle cell identities involves the same categories of genes.

Direct Lb targets revealed by ChEST

To identify Lb-dependent cis-regulatory modules (CRMs) and genes directly regulated by Lb in muscle and heart cells we used a ChIP-enriched in silico target approach [ChEST] (Fig. 1D; Supplementary Fig. S1B; Junion et al. 2005). The ChEST assay is based on a combination of computational prediction of TF-binding regulatory modules with ChIP and enabled us to select the CRMs bound in vivo by the factor of interest, Lb genes code for homeodomain transcription factors thought to bind DNA. However, the DNA sequences to which Lb binds and genes directly regulated by Lb have not yet been described. As the Lb-binding site has not yet been determined we first used the in vitro SELEX approach [Roulet et al. 2002] to identify DNA motifs recognized by Lb protein [Fig. 1E]. We found that 6His-LbHomeodomain recombinant protein bound to 10-nucleotide [nt] motifs containing TAAT or TAAC core sequences [Fig. 1E, top panels]. These motifs were also specifically recognized by the full-length Lbe and Lbl proteins [Fig. 1E, bottom panel]. The alignment of all variants of Lb-bound motifs gave the consensus Lb-binding site RVYTAAYHAG [Fig. 1E], which was used for in silico genome scanning as previously described [Junion et al. 2005]. The ChEST strategy adapted for seeking Lb-dependent muscle and heart enhancers [Fig. 1D] consisted of a computer-as-
sisted genome-wide screen for CRMs containing Lb-binding sites clustered with motifs recognized by heart- and muscle-specific factors [see Supplemental Material]. The selected CRMs were amplified and spotted to produce a computed Lb-CRM chip. In parallel, a pool of genomic DNA fragments bound by Lb in vivo was isolated from whole 5- to 16-h-AEL embryos by ChiP with specific anti-Lb antibody [Fig. 1D]. A nonimmune serum was used in a control ChiP assay [see Supplementary Fig. S1B]. The ChiP-isolated DNA fragments from three independent experiments were labeled and used as a probe for the Lb-CRM chip to identify ChiP-enriched CRMs. Using heart [H] targeting and muscle [M] targeting conditions [see Materials and Methods] we predicted 336 potential H + M Lb-dependent CRMs, from which after stringent gene annotation-based filtering [see Supplemental Material for details] 74 CRMs were selected and spotted to produce an Lb-CRM array. From this pool, 16 CRMs were found to be significantly enriched after hybridization with the Lb-ChiP probe [Table 1C; Supplementary Table S7]. The restricted number of ChEST-revealed Lb targets results most probably from the applied stringent in silico filtering procedure and the small proportion of mesodermal cells expressing Lb proteins in ChiP samples. Nevertheless, the ChEST-identified target genes fit into the same categories as the microarray-revealed candidates, thus confirming the global view of Lb transcriptional targets. Importantly, eight out of 16 ChEST-CRMs lie in the vicinity of microarray-identified candidate genes [bold candidates in Table 1C]. Among the common ChEST/microarray targets are genes encoding transcription factors [bab2], setting cell shape, motility, and adhesion [sli, Dhc93AB, shg, CG10116], and also involved in later aspects of muscle differentiation and function [if, how, CG8698].

Lb regulates expression of microarray and ChEST-identified candidate genes

To validate the microarray data we tested the expression of candidates from each target’s category by RT–PCR, real-time RT–PCR [Fig. 2A, A’, B, Supplementary Table S8], or if available, by antibody staining [Fig. 2C–J]. In all the cases tested, changes in the expression of candidate genes agreed with microarray data. For example, shg expression decreased by about half in heart-specific Lb-GOF embryos in both microarrays and in RT–PCR [see Table 1B; Fig. 2A, Supplementary Table S8], whereas the level of slit transcript [detected by microarrays and by real-time RT–PCR] was consistently higher in the H + M Lb-LOF context [Table 1A, Supplementary Table S8]. The changes in candidate gene expression induced by Lb LOF or GOF are also seen at protein levels [Fig. 2C–J]. We found that the reduced expression of slouch [slo] and nautilus [nau] in Lb H + M GOF embryos detected in microarrays correlated with the loss of muscle-specific Slou and Nau stainings in embryos overexpressing Lbe under the control of a 24B-GAL4 driver [Fig. 2C–F]. The capacity of Lb to act as an activator and regulate expression of genes encoding muscle structure components is illustrated by accumulation of Sallimus [Sl] and Msp300 proteins in muscles ectopically expressing Lbe [Fig. 2G–J].

To confirm that the ChEST-identified CRMs act as bona fide enhancers in vivo we generated transgenic reporter lines carrying Lb-dependent modules close to bab2 and if genes [Fig. 2K, P], which are heart and muscle Lb targets common for ChEST and microarray assays [Table 1C]. The role of Lb in regulation of bab2 and if expression was validated by real-time RT–PCR [Fig. 2A’]. In the case of bab2 CRM [Fig. 2K–O] we found that GFP expression in the transgenic reporter line recapitated cardiac bab2 expression starting from stage 13. In microarray experiments [Table 1B] and in real-time RT–PCR validation tests [Fig. 2A’; Supplementary Table S8] bab2 was found to be negatively regulated by Lb, so we decided to test whether the bab2 CRM-driven expression of GFP [bab2-GFP] was influenced by pan-mesodermal overexpression of Lbe. Interestingly, we observe two distinct effects. In embryos at early stage 13, bab2-GFP expression was down-regulated in cardiac primordia [Fig. 2M] indicating that at the beginning of cardiogenesis Lb acts as direct repressor of bab2 excluding its activity from Lb-positive cells. In embryos at stage 14 and older Lb no longer acted as a bab2 repressor. We observed instead that GFP expression was enhanced in the Lb GOF context [Fig. 2O] indicating a switch from a repressive to an activating action of Lb. Overall, these data reveal that Lb directly represses bab2 expression during early steps of cardiac cell fate specification, and later on acts as its activator most probably by interacting with different cofactors.

The ChEST-identified if CRM [Fig. 2P–R] also acts as an enhancer in vivo. It drives lacZ expression in a restricted subset of lateral muscles with the most prominent expression in the Lb-positive SBM [segment border muscle] [Fig. 2Q]. To test whether Lb is able to regulate positively if via the identified CRM we analyzed if-lacZ expression in embryos expressing Lbe ubiquitously in all muscle cells. As shown in Figure 2R, lacZ expression is ectopically expanded indicating that Lb is able to act as an if activator. if codes for α-PS2 integrin and is known to play important roles in muscle attachment and sarcomeric assembly [Bloor and Brown 1998; Brown et al. 2000]. In embryos, if accumulates as a heterodimer with β-PS-integrin, Myospheroid [Mys] at extremities of growing myotubes where it is required for interaction with ECM ligands such as Tiggerin to facilitate muscle–tendon and muscle–muscle contacts [Brown et al. 2000]. It has been proposed that by interactions with the ECM-forming proteins [Collagens, Laminins] outside the cell and with actin-binding proteins such as Talin, Vinculin, and α-actinin within the cell if ensures a link between adhesion and cell shape. Hence, identifying if as a direct Lb target provided a strong line of evidence that Lb regulates morphological and dynamic properties of cells and is able to influence late aspects of muscle differentiation. Thus, Lb appears to directly regulate genes whose activity ranges from the cell fate determination to the late aspects of cell-type-specific differentiation.
Based on presented here data we propose a model [see Fig. 2S] for direct [e.g., if] and indirect [e.g., sls] Lb-dependent and fiber-type-specific adjustment of pan-muscular gene expression. In the direct mode Lb binds to the identified if CRM [Fig. 2S, red box] and together with previously described [Junion et al. 2005] generic dMef2-dependent CRMs [Fig. 2S, blue boxes] regulates fiber-type specific level of if expression. In an indirect mode Lb can
influence expression of another pan-muscular gene, sls previously found to be direct target of dMef2 (Junion et al. 2005). In this particular case Lb is expected to act indirectly via a so far unknown factor, which binds to the CRM1 (Fig. 2S, blue and pink boxes) and together with dMef2 regulates sls expression in lateral muscles including SBM (Fig. 2S, see Discussion for further comments; see Junion et al. 2005).

bab2 is directly regulated by Lb and contributes to diversification of cardiac fates

bab2 is one of the common ChEST/microarrays candidates. It has been previously reported that bab2 is involved in leg development (Couderc et al. 2002) and in sex-specific pigmentation of the adult fly abdomen (Gompel and Carroll 2003). Thus, it was surprising to find bab2 as an Lb target in embryos, and we decided to analyze the embryonic expression pattern of the Bab2 protein (Fig. 3A–D). Our data clearly show that Bab2 is dynamically expressed in mesodermal cells including cardiac precursors. Triple staining with Lbe and Eve (Fig. 3A–C) show that during the early steps of cardiac development, at early stage 12, Bab2 is present in Eve-positive pericardial and dorsal muscle precursors but is excluded from Lbe-positive cardiac cells. This specific expression underlies cardiac diversification events, which take place in stage 11 and early stage 12. Starting from the late stage 12 (Fig. 3B) weak Bab2 levels can be detected in one of the two Lb-positive cardioblasts, and slightly later (Fig. 3C) in all cardioblasts and pericardial cells. Interestingly, after cardiac tube formation Bab2 expression is maintained in all cardioblasts, and at a lower level in pericardial cells, indicating that it can also play a role in heart morphogenesis. The identification of Bab2 protein in a subset of cardiac precursors at early stage 12 (Fig. 3A) prompted us to determine whether its function was required for diversification of cardiac cells.

To address this issue we took advantage of previously described loss-of-function (LOF) and GOF alleles for bab2 (Couderc et al. 2002). As we observed that Bab2 was initially coexpressed with Eve but excluded from Lb-positive cardiac precursors, we wondered whether Bab2 contributed to the restriction of Lb cardiac expression by repressing its activity in neighboring Eve cells. As shown in Figure 3F (arrows), in bab2 mutant embryos Lbe expression is enlarged and can be detected in Eve-positive cells, suggesting that Bab2 acts as a repressor of Lb in Eve-positive cells. The repressive potential of Bab2 during the diversification of heart cells is confirmed by

Figure 3. bab2 expression and regulation during cardiac development. [A–D] Bab2 expression in cardiac precursors and the dorsal vessel. (A) At early stage 12 Bab2 is expressed in Eve-positive but not in Lb-positive cardiac precursors. (B) At the end of stage 12, a weak Bab2 expression starts to appear in cardioblasts including Lb-positive cells (arrows). (C) At stage 13, Bab2 is expressed in all cardiac cells. (D) After fusion of cardiac primordia, at stage 15, Bab2 protein is particularly well seen in the heart proper. (E–L) bab2 GOF and LOF influences cell fate specification within the heart and leads to the altered positioning of cardiac precursors [shown in E,J]. Wild-type stage 12 and stage 15 embryos stained for Tin, Eve, and Lbe. Note that Lbe and Eve are coexpressed with Tin but mark distinct subset of cardioblasts and pericardial cells. Arrowheads in I point to the two Lbe/Tin cardioblasts present in each hemisegment. (E,J) In bab2 mutant embryos Lbe expression is enlarged (arrowheads in J) and appears in Eve-positive cells (arrows). (G,K) In contrast, bab2 GOF leads to the repression of Lbe within the cardiac primordium (arrows in G, arrowheads in K). (K) Asterisks indicate lacking cardioblasts and arrows point to a supernumerary Eve-positive pericardial cell. (H,L) Triple (β3-Tubulin, Lbe, Eve) staining of stage 15 wild-type (H) and bab2 mutant (L) embryos. Note irregular β3-Tubulin pattern (arrowheads in L) and abnormal, cardioblast-like position (arrow in L) of some Eve-positive cells in bab2 mutants.
the overexpression of bab2 using the 24B-GAL4 driver [Fig. 3G]. In embryos with forced Bab2 expression the number of Lbe-positive cells in cardiac primordia was dramatically reduced [Fig. 3G], strongly suggesting that during early heart development Bab2 contributes to cell fate specification decisions by acting as an Lb repressor. The cell identity function of bab2 and its repressive influence on cardiac lb expression is also observed in stage 15 GOF and LOF bab2 mutant embryos [Fig. 3 H–L]. Notably, loss of bab2 leads to the abnormal positioning of Eve-positive cells some of which coexpress Lbe and are present more dorsally at cardioblasts positions [Fig. 3L, arrow], whereas GOF of bab2 provokes loss of some Tin-positive cardioblasts [Fig. 3K, asterisks]. Alterations in cardiac cell fates and in their spatial arrangement result in affected cardiac tube morphology [Fig. 3L].

Genes downstream from lb control founder migration, myoblast fusion, and muscle attachment processes

The acquisition of cell identity is expected to involve the action of genes controlling cell movements, the establishment of contacts with neighboring cells and the development of cell type-specific functions. In Drosophila embryos, lb-expressing muscle founders undergo directed cell movements, fuse with a defined number of feeder myoblasts, and attach to the stereotypic insertion sites. lb is expressed during all these steps of SBM development but its role has only been reported for the initial SBM founder specification step [Jagla et al. 1998]. Here we find that lb regulates the expression of a broad range of genes encoding extracellular matrix (ECM) components and actin-binding proteins known to control cell-cell contact and cell motility, suggesting potential role of lb and its targets in migration, fusion, and muscle attachment processes. To better characterize the role of lb in all steps of SBM formation we used muscle-targeted attenuation of lb function via RNAi. This allowed us to generate a range of RNAi-induced hypomorph-like SBM phenotypes [Fig. 4], which are seen in ~41% of 24B > lbRNAi embryos [48 out of 118 embryos analyzed]. Embryos presenting phenotypes can be subdivided on three categories: (1) embryos with founder migration phenotype characterized by lack of large red nucleus seen in the wild type at position indicated by the arrowhead in Figure 4C (37%), 18 out of 48 embryos), (2) embryos with affected myoblast fusion characterized by rounded cells close to SBM indicated by arrowheads in Figure 4E (50%, 24 out of 48 embryos), and (3) embryos with affected SBM attachment characterized by the abnormal SBM shape and abnormal attachment to adjacent muscles indicated by arrowheads in Figure 4G (13%, 6 out of 48 embryos). These data demonstrate that lb function is required for all these aspects of SBM development. Interestingly, muscle targeted attenuation of several Lb targets [Fig. 4; data not shown] leads to similar defects in SBM formation. For example, SBM founder migration was strongly affected after 24B-targeted attenuation of Msp300, a microarray-identified Lb target known to be involved in actin cytoskeleton remodeling [Fig. 4D]. Lb targets appear also to contribute to myoblast fusion processes. Among them, muscle-targeted attenuation of Msp20 encoding actin-binding protein phenocopies lbRNAi-induced SBM fusion phenotypes [Fig. 4, cf. E and F]. The performed functional tests also suggest that lb targets encoding the ECM components are involved in the recognition of proper muscle attachment sites. This is particularly obvious when testing RNAi-induced attenuation of vkg function [Fig. 4H]. However, in contrast to SBM-specific phenotypes seen in lbRNAi embryos, the attenuation of pan-muscular Lb target genes such as Msp300 leads to the phenotypes affecting larger subsets of muscles. Thus, we propose that in the SBM context Lb ensures a fiber-type-specific transcriptional regulation of this class of genes [see Fig. 2S].

Functional properties of muscle and heart cells are controlled by lb

We know that lb expression persists in fully formed SBM muscle and can also be seen at a low level in cardiac cells after fusion of heart primordia. The functional significance of this late lb activity is not well understood. Here among the microarray candidates we find a large number of genes encoding muscle structure proteins (e.g., sls, Ptm, Up, Mhc), ion channels (Sh, CG12061, or CG4370), or proteins involved in muscle contraction (Mlc1, Mlc2, Tm, TnpC73F, bt, CG8698). These findings strongly suggest that during muscle differentiation lb regulates the expression of genes that determine functional properties of muscle fibers raising the possibility that the composition of the sarcomeric cytoskeleton may differ from muscle to muscle. Differences in the expression of muscle structure proteins can then be associated with the identity gene-governed regulation of genes controlling muscle contraction, leading to the formation of functionally distinct muscle fibers. To test whether lb can control muscle contraction we analyzed expression and function of ChEST-identified direct Lb target, the CG8698/Unc93 gene [Fig. 5; Supplementary Movie Fig. S3] encoding an evolutionarily conserved component of a two-pore potassium channel complex and known to be critically required for muscle contraction in Caenorhabditis elegans [de la Cruz et al. 2003]. The Drosophila CG8698 gene is specifically expressed in differentiating somatic muscles starting from stage 14, and its transcripts are particularly abundant in late embryonic stages in functional, contracting muscles [Fig. 5A–D]. To determine whether CG8698 is involved in controlling muscle contraction we used direct injection of double-stranded RNA [dsRNA] into Mhc-tauGFP embryos [see Materials and Methods, Estrada et al. 2006]. The age-matched Mhc-tauGFP embryos injected with dsRNA against the lacZ gene [Fig. 5E, corresponding Supplementary Movies Fig. S3A,C] were compared with those injected with dsRNA against the CG8698 gene [Fig. 5E, corresponding Supplementary Movies Fig. S3B,D]. To avoid the off-target effects, two dsRNAs of ~500 nt targeting two distinct gene regions were used and effects on muscle contraction analyzed by confocal time-lapse im-
aging followed by recording of muscle size changes (Fig. 5E). In total, contractions of six SBM muscles from control injected embryos and six from embryos injected with CG8698 dsRNA were recorded and used to obtain the mean changes in the SBM size (Fig. 5E, bottom panel). As a positive control we injected dsRNA against the CG2708 gene previously found to block muscle contraction (Estrada et al. 2006) and observed loss of muscle contractility phenotype similar to that reported by Estrada et al. (2006) (data not shown).

This analysis clearly showed that the attenuation of CG8698 leads to the significantly reduced muscle contractility indicating that \( \text{lb} \) by regulating transcription of genes such as CG8698, can modulate properties of SBM contraction.

**Discussion**

Uncovering how the cell fate-specifying genes exert their functions and determine unique properties of cells in a tissue is central to understanding the basic rules governing normal and pathological development. To approach the cell fate determination process at a whole genome level we performed a search for transcriptional targets of the homeobox transcription factor \( \text{lb} \) known to be evolutionarily conserved and required for specification of a subset of cardiac and muscular precursors. To this end we combined the targeted expression profiling and the novel ChIP-on-chip method ChEST (Junion et al. 2005). Our data revealed an unexpectedly complex gene network operating downstream from \( \text{lb} \), which appears to act not only by regulating components of the cell identity code but also as a modulator of pan-muscular gene expression at fiber-type level. Of note, the role of *Drosophila* \( \text{lb} \) in regulating SBM founder motility appears reminiscent of the role of its vertebrate ortholog *Lbx1*, known to control the migration of leg myoblasts in mouse embryos (Vasyutina et al. 2005).

The \( \text{lb} \)-regulated components of the cell identity code

Earlier genetic studies revealed that within the same competence domain the cell fate specifying factors acted as repressors to down-regulate genes determining the identity of neighboring cells (Jagla et al. 2002). Consistent with this finding, \( \text{lb} \) was found to repress *msh* and *kruppel* (*kr*) during diversification of lateral muscle precursors (Jagla et al. 1998, 1999) and *even skipped* (*eve*) within the heart primordium (Jagla et al. 1997, 2002). Here we find that additional identity code components are regulated negatively by \( \text{lb} \). In the lateral muscle domain \( \text{lb} \) acts as a repressor of the *MyoD* ortholog *nau* and the NK homeobox gene *slou* (Fig. 2C–F), both known to be required for the specification of a subset of somatic myoblast fusion. Nonfused myoblasts (arrowheads) are detected around the SBM, and the narrowed part of the muscle fiber (arrow, see also scheme). (G,H) In a portion of \( \text{lb} \) RNAi and in *vkg* RNAi embryos SBM fibers display abnormal shapes and attachments (arrowheads, see corresponding scheme). Attachment phenotypes are often accompanied by the presence of ectopic fibers located close to SBM (arrows).
versification of muscle fibers, our global approach identified a few novel potential players in the muscle identity network. Among them are expressed in somatic muscle precursors the Pdp1 gene encoding Par domain factor (Reddy et al. 2000) and the CG32611 gene containing a zinc finger motif (see Table 1).

Interestingly, in the cardiac domain our data demonstrate that Lb is able to positively regulate the expression of tin and the effector of RTK pathway pointed (pnt), both involved in cardiac cell fate specification (Alvares et al. 2003; Zaffran et al. 2006). These findings are consistent with earlier observations that the forced Lb expression leads to the ectopic tin-positive cells within the dorsal vessel (Jagla et al. 2002). Also, during early cardiogenesis Lb directly represses bric a brac 2 (bab2), which emerges as a novel component of the genetic cascade controlling the diversification of cardiac cells. Thus, the ability of Lb to act either as repressor or as activator suggests a context-dependent interaction with cofactors. Of note, several microarray identified Lb targets have also been found in the RNAi-based screen for genes involved in heart morphogenesis (see Supplementary Table S5; Kim et al. 2004).

Identity gene-dependent modulation of muscle gene expression

Our data indicate that Lb exerts its muscle identity functions via regulation of pan-muscular genes that control cell movements, cell shapes and cell–cell interactions including myoblast fusion, myotube growth, and attachment events.

Regulation of if [see Fig. 2S] provides an example of how the cell-type-specific fine-tuning of expression can be achieved for genes expressed in all muscle precursors. We reported previously (Junion et al. 2005) that if is directly regulated by dMef2 via two intronic CRMs (Fig. 2S), which most probably ensure a generic muscle-specific if expression. Here, we identify a distinct, Lb-dependent CRM able to drive expression in a restricted subset of muscles including the SBM, thus indicating that the if transcription is regulated by a coordinated action of generic and muscle-type-specific CRMs [Fig. 2S]. Identification of Lb-dependent CRM close to if also indicates that Lb may contribute to the dMef2-regulated feed-forward loops (Sandmann et al. 2006). Such a modular regulation of transcription levels would provide an efficient way for precise, fiber-type-specific setting of muscle genes activity. As shown for if in the SBM context, this regulation is expected to be direct via the identified here Lb-binding module [Fig. 2S]. Our data also suggest that a similar muscle-type-specific transcriptional regulation can be achieved in an indirect mode. For example, the pan-muscular kettin/sls gene encoding a giant protein required for the formation and maintenance of normal sarcomere structure is regulated by three distinct dMef2-binding CRMs [Fig. 2S] able to drive expression in a muscle-type-specific manner (Junion et al. 2005). Among them, CRM 1 located upstream the gene was found to drive expression in a subset of ventral and lateral
muscles including SBM [Junion et al. 2005]. Here we identify sls as an Lb target in the microarray screen, and demonstrate that in embryos overexpressing Lbe the sls expression is up-regulated [Fig. 2H]. These findings suggest that Lb can act via an as-yet-unknown factor to modulate sls expression in the SBM [Fig. 2S].

A model for Lb-governed multistep acquisition of muscular and cardiac fates

The reported here Lb targets fit into three main categories: (1) genes encoding transcription factors contributing to cell-type-specific identity gene code; (2) genes controlling cell shapes, adhesion, and cell motility; and (3) late-acting genes required for functional properties of cardiac and muscle cells [Fig. 6A]. Importantly, Lb is able to bind to enhancers of genes from all these categories [Fig. 6B, red-labeled genes]. Also, the functional properties of identified Lb targets strongly suggest that Lb-dependent multistep acquisition of cell identity is executed in a similar way in cardiac and in muscle cells [Fig. 6B]. With respect to the Lb targets encoding transcription factors we identify bab2 as a new component of cardiac cell identity code. Considering the regulation of cell shape, cell adhesion, and motility, a large number of Lb targets code for proteins involved in remodeling of the actin cytoskeleton, for the ECM components and for integrins. Down-regulation of the selected microarray candidates from this category [Fig. 6B, pink-labeled genes] leads to dramatic muscle phenotypes affecting founder migration, myoblast fusion or muscle fiber attachment. Among the late-acting genes, we analyzed expression and function of a direct, ChEST-identified Lb target gene, CG8698, and demonstrated that it is specifically expressed in differentiated muscles and required for muscle contraction.

Thus, we conclude that the Lb-governed acquisition of cell identity is a long-term process [1] initiated by the spatially restricted expression of a set of transcription factors and [2] executed by precise regulation of genes determining the morphological, dynamic and functional properties specific to a given muscle or heart cell type.

Materials and methods

Targeted expression profiling of Lb GOF and LOF embryos

As presented in Figure 1, two different drivers [24B-GAL4 and tin-GAL4] were used for crosses with UAS-lbe and UAS-lbRNAi lines to obtain heart (H) and muscle plus heart (M + H)-specific targeting conditions. The 24B-Gal4 (P[GawB]how248) was provided by Bloomington Stock Center (BL 1767) and tin-GAL4 [gift from R. Bodmer, Burnham Institute, San Diego, CA] was described in Zikova et al. [2003]. Aged [5-16-h-AEL-old] embryos were collected. One-hundred embryos were homogenized, and RNA was isolated using Trizol Reagent [Gibco] and purified using MessageClean [GenHunter] following the manufacturer’s protocols. Three independent RNA isolations were performed for each of five genetic contexts and used to produce probes for microarrays. The Affymetrix cDNA Drosophila Genome Arrays were used.

Analysis of microarray data

Primary data analysis was performed using the GeneChip Microarray Suite [MAS] 5.0 algorithms [Affymetrix] with default parameters and global scaling as a normalization method [target intensity = 100]. Reproducibility was ensured by using probes corresponding to three independent RNA extractions for each genetic context.

The Significance Analysis of Microarray Data [SAM] implemented in the TIGR MeV version 3.1 software [http://www.tm4.org/scgi-bin/getprogram.cgi?program=mev_old] was used to identify those genes that had statistically significant differences in expression between Ladybird GOF and LOF embryos and wild-type embryos. Parameters were set as specified below to analyze two-class, unpaired data: 1000 permutations, k-nearest neighbor imputer with 10 neighbours and newly initialized random number seeds for each analysis. FDR was varied.
between 0.05 and 0.1 for evaluation of changes in the number of significant genes.

In lβ GOF conditions, genes with a q value < 0.05 and a fold change of more than +1.5 or less than −1.5 with respect to the wild type were considered to be regulated by lβ. In the context of RNAi-based lβ attenuation, we applied a higher q value of < 0.01 (H + M conditions) and < 0.05 (H conditions). This was due to the incomplete RNAi-mediated attenuation of lβ [Maqbool et al. 2006]. The fold change cutoff values (+1.5, −1.5) were as for GOF experiments. The candidate genes were classified according to their GO annotations using DAVID software [http://david.abcc.ncifcrf.gov] and according to the embryonic expression profile. To validate microarray data the expression of some of the candidates were tested by RT–PCR, real-time RT–PCR, or by antibody staining.

Validation of microarray candidates by RT–PCR and by real-time RT–PCR

Mutant and wild-type embryos from stages 11–15 were collected. Total RNA was extracted with the Trizol Reagent [Gibco], and reverse transcription was performed on 0.5 µg of total RNA with the SuperScript III kit [Invitrogen]. RT–PCR conditions were 30 sec, 94°C, hybridization temperature: according to Tm, 60 sec, 72°C, 60 sec, repeated 40 times. RT–PCR product aliquots were taken for semiquantitative analysis at cycles 30, 35, and 40. Real-time RT–PCR was performed on 2 µL of cDNA using the LC-FastStart DNA Master SGI PCR amplification kit [Roche Diagnostics] and the LightCycler Carousel-Based system [Roche Diagnostics]. The real-time PCR conditions were as follows: denaturation 10 min, 95°C; 45 cycles; denaturation 10 sec, 95°C; hybridization: Tm primer 5 sec, 5°C, elongation 30 sec, 72°C. Fusion curve: denaturation 95°C, 0 sec, hybridization 65°C, 15 sec, melting 95°C, 0 sec. As an internal control the ribosomal protein 49 gene, rp49, was amplified. List of RT–PCR and real-time RT–PCR primers used to test candidate gene expression is provided in the Supplemental Material. The quantification of RT–PCR and real-time RT–PCR validation tests is provided in Supplementary Table S8. To compare relative gene expression between the wild-type and the Lb-GOF or Lb-LOF conditions from microarrays screen: 

ChEST approach

The ChEST approach was used as described previously [Junion et al. 2005]. Various combinations of binding sites of Lb, Twist [basic helix–loop–helix [bHLH]], Yan (ETS), dMef2, dTCF, and Tin were used to identify potential Lb-dependent CRMs [Supplemental Material]. The Lb consensus binding site was determined using the SELEX approach and then tested by gel shift assay [see below, Supplemental Material]. The set of binding sites including Tin and dMef2 was defined to select heart- and muscle-specific Lb target genes. The applied genome scanning conditions and the subsequent filtering procedure using SCAff software [Junion et al. 2005] are presented in the Supplemental Material. For each of the in silico-identified regulatory sequences, a pair of primers was defined and the sequence was amplified by PCR, purified, and spotted on a nylon membrane to obtain a computed LB CRM array. The DNA immunoprecipitated with anti-Lb antibody and the control DNA (ChIP with nonimmune serum) were used as templates for the synthesis of the probes, which were then hybridized to the generated CRM array.

Purification of His fusion proteins and identification of Lb consensus binding site by SELEX

Both the lbe and lbl homeodomain encoding DNA fragments as well as full protein-coding sequences have been cloned in His-Tag-pET expression vector [pET-45b+, Novagen]. DNA fragments were amplified by PCR using primers carrying appropriate restriction sites and then directionally cloned into the linearized His-Tag-pET vector. Protein purification was performed according to the manufacturer’s protocols. Isolated His fusion proteins were affinity-purified on agarose NTa beads [Novagen] and used for the oligonucleotide selection procedure [SELEX] as described in the Supplemental Material and according to Chitenden et al. [1991]. For the gel retardation assay, double-stranded oligonucleotides containing the Lb-binding sites were PAGE-purified prior to radioactive labeling. Five picoomoles of oligonucleotides were 5’-phosphorylated with T4 polynucleotide kinase [New England Biolabs] using 2.5 µL of [α-32P]-ATP (6000 Ci/mmol, Amersham) in a total volume of 20 µL. Probes were PAGE-purified, extracted, and precipitated with ethanol. Pellets were dissolved in 5 mM Tris-HCl [pH 8.0], 50 mM NaCl, and 0.5 mM EDTA. His-Tag fusion proteins were diluted in PBS containing 200 µg/mL BSA, 1 mM EDTA, and 1 mM DTT. Unless indicated otherwise, 100 ng of protein were incubated with the labeled probe (25,000 counts per minute [cpm]) in binding buffer [20 mM HEPES at pH 7.0, 50 mM KCl, 4 mM MgCl2, 4% Ficoll, 5% glycerol, 0.2 mM EDTA, 0.5 mM DTT, 500 ng poly[dI-dC]] for 20 min at room temperature. Five-microliter aliquots were loaded on 5% polyacrylamide gels and electrophoresed at 4°C in 0.5× TBE. Gels were dried and visualized by autoradiography.

Generation of transgenic lines and validation of ChEST-identified targets

ChEST-identified Lb-dependent CRMs were cloned upstream of lacZ or GFP reporter genes and then used for germline transformations to generate transgenic Drosophila lines. Reporter gene expression was revealed in whole-mount embryos double-stained with rabbit anti-LacZ, 1:1000 [Sigma], goat anti-GFP [1:300] [Biogenesis], and mouse anti-Lbe, 1:5000. To check whether the CRM-driven expression of reporter gene was Lb dependent, its expression was tested in an lβ GOF context. To determine the embryonic expression of bab2 we used rat anti-Bab2 antibodies (1:2000) [kindly provided by J.-L. Courtin]. We also used guinea pig anti-Eve (1:200) and rabbit anti-Slou (1:500) [gift of M. Frasch], rabbit anti-Nau (1:50) [from B. Patterson], guinea pig anti-Msp300 [1:1000] [provided by T. Volk], rat anti-if [1:2] [gift of N. Brown], rabbit anti-Tin [1:800] [from M. Frasch], rabbit anti-dMef2 [1:2000] [from H. Nguyen], and rabbit anti-b3-Tubulin [1:5000] [from R. Renkawitz-Pohl]. The bab2 mutant alleles bab2+ and bab20 and the uas-bab2 line [Couedic et al. 2002] were used to analyze the role of bab2 in cardiac development.

Testing functions of Lb microarray targets by muscle-targeted RNAi

The following uas-RNAi lines from the NIG-Fly collection [http://www.shigen.nig.ac.jp/fly/nigfly] were used to test functions of Lb target genes from microarrays screen: UAS-Msp300RNAi Ref. 31916R-1, UAS-Mp20RNAI Ref. 4696R-4, and UAS-vkgRNAi Ref. 16858R-1. The uas-RNAi lines for candidate genes as well as the UAS-lbeRNAi line, targeting both lbe and msg [Maqbool et al. 2006], were crossed with a 2AB-CALD driver, and the F1 embryos were double-stained for β3-Tubulin and Lbe and analyzed for muscle phenotypes.
Analysis of CG8698 expression pattern and its role in muscle contraction

CG8698 cDNA (GH09628, DGRC gold collection) clone was used to prepare a Dig-labeled antisense RNA probe according to standard procedure [Tautz and Pfeifle 1989]. Fluorescent in situ hybridization with TSA amplification system [Perkin-Elmer] was done as described previously [Junion et al. 2005] and followed by anti-dMeF2 (1:2000) and anti-Lbc (1:5000) stainings to reveal embryonic muscle cells. Triple-labeled embryos were analyzed using an LSM510 Meta [Zeiss] confocal microscope. To test the role of CG8698 in muscle contraction, we used a previously described approach [Estrada et al. 2006] based on direct dsRNA injection into Mhc-tauGFP embryos. To avoid off-target effects for each gene tested, two distinct gene segments of 400–600 base pairs [bp] in length were selected for dsRNA synthesis. These sequences were PCR-amplified from genomic DNA using primers flanked by T7 promoters on both ends [see primer sequences listed below]. PCR products were purified and transcribed in vitro using the Ribomax T7 In vitro transcription system [Promega]. dsRNA was diluted to the concentration of 2 mg/mL in water for embryo transfer [Sigma] and injected midventrally into dechorionated, syncytrial blastoderm-stage Mhc-tauGFP embryos. About 50 injected embryos were analyzed for each dsRNA probe. Injection of lacZ dsRNA was used as a negative control and the CG2708 dsRNA [Estrada et al. 2006] was used as a positive control. Muscle contraction was monitored using an Olympus FV300 confocal microscope [frame capture every 5 sec] and processed to make movies by Volocity software [Improvement]. In total, the muscle contractions of 10 lacZ dsRNA embryos and 10 CG8698 RNAi embryos were recorded. To define the mean changes in the SBM size due to the muscle contractions, three embryos for each context were selected [by criteria of their orientation to visualize SBMs] and two distinct SBM muscles per selected embryo were analyzed. A Velocity Measurement module was used to analyze time-lapse movies and to plot size changes of individual muscle fibers during contraction.

The microarray data sets are available via http://www.ebi.ac.uk/arayexpress. ArrayExpress accession number E-TABM-297.

Acknowledgments

We are grateful to C. Girardot for help in microarray data submission, S. Basset for help in setting real-time RT-PCR conditions, and Soluscience SA for designing bioinformatics tools for CRM scanning and data set comparisons. This work was supported by the Institut National de la Santé et de la Recherche Médicale, the Association Francaise contre les Myopathies, the Association pour la Recherche sur le Cancer, and European Grant LSHG-CT-2004-511978 to the MYORES Network of Excellence.

References

Alvares, L.E., Schubert, F.R., Thorpe, C., Mootoo, R.C., Cheng, L., Parkyn, G., Lumsden, A., and Dietrich, S. 2003. Intrinsic, Hox-dependent cues determine the fate of skeletal muscle precursors. Dev. Cell 3: 379–390.

Baylies, M.K. and Michelson, A.M. 2001. Invertebrate myogenesis: Looking back to the future of muscle development. Curr. Opin. Genet. Dev. 11: 431–439.

Bloor, J.W. and Brown, N.H. 1998. Genetic analysis of the Drosophila aPS2 integrin subunit reveals discrete adhesive, morphogenetic and sarcomeric functions. Genetics 148: 1127–1142.

Briscoe, J., Pierani, A., Jessell, T.M., and Ericson, J. 2000. A homeodomain protein code specifies progenitor cell identity and neuronal fate in the ventral neural tube. Cell 101: 435–445.

Brohmann, H., Jagla, K., and Birchmeier, C. 2000. The role of Lbx1 in migration of muscle precursor cells. Development 127: 437–445.

Brown, N.H., Gregory, S.L., and Martin-Bermudo, M.D. 2000. Integrins as mediators of morphogenesis in Drosophila. Dev. Biol. 223: 1–16.

Buckingham, M. 2003. How the community effect orchestrates muscle differentiation. Bioessays 25: 13–16.

Carmena, A., Gisselbrecht, S., Harrison, J., Jimenez, F., and Michelson, A.M. 1998. Combinatorial signaling codes for the progressive determination of cell fates in the Drosophila embryonic mesoderm. Genes & Dev. 12: 3910–3922.

Cerlet, S.J. and Thor, S. 2004. Specification of Drosophila motoneuron identity by the combinatorial action of POU and LIM-HD factors. Development 131: 5429–5439.

Cheng, L., Samad, O.A., Xu, Y., Mizuguchi, R., Luo, P., Shirasawa, S., Goulding, M., and Ma, Q. 2005. Lbx1 and Tlx3 are opposing switches in determining GABAergic versus glutamatergic transmitter phenotypes. Nat. Neurosci. 11: 1510–1515.

Chittenden, T., Livingston, D.M., and Kaelin Jr., W.G. 1991. The T/E1A-binding domain of the retinoblastoma product can interact selectively with a sequence-specific DNA-binding protein. Cell 65: 1073–1082.

Couderc, J.L., Godt, D., Zollman, S., Chen, J., Li, M., Tingon, S., Cranton, S.E., Sahut-Barnola, I., and Laski, F.A. 2002. The bric a brac locus consists of two paralogous genes encoding BTB/POZ domain proteins and acts as a homeotic and morphogenetic regulator of imaginal development in Drosophila. Development 129: 2419–2433.

De Graeve, F., Jagla, T., Daponte, J.P., Rickert, C., Dastugue, B., Urban, J., and Jagla, K. 2004. The ladybird homeobox genes are essential for the specification of a subpopulation of neural cells. Dev. Biol. 270: 122–134.

de la Cruz, I.P., Levin, J.Z., Cummins, C., Anderson, P., and Horvitz, H.R. 2003. sup-9, sup-10, and unc-93 may encode components of a two-pore K+ channel that coordinates muscle contraction in Caenorhabditis elegans. J. Neurosci. 23: 9133–9145.

Estrada, B., Choc, S.E., Gisselbrecht, S.S., Michaud, S., Raj, L., Busser, B.W., Hallon, M.S., Church, G., and Michelson, A.M. 2006. An integrated strategy for analyzing the unique developmental programs of different myoblast subtypes. PLoS Genet. 2: e16. doi: 10.1371/journal.pgen.0020016.

Fichelson, P., Audibert, A., Simon, F., and Gho, M. 2005. Cell cycle and cell-fate determination in Drosophila neural cell lineages. Trends Genet. 21: 413–420.

Frash, M. 1999a. Intersecting signalling and transcriptional pathways in Drosophila heart specification. Semin. Cell Dev. Biol. 10: 61–71.

Frash, M. 1999b. Controls in patterning and diversification of somatic muscles during Drosophila embryogenesis. Curr. Opin. Genet. Dev. 9: 522–529.

Gompel, N. and Carroll, S.B. 2003. Genetic mechanisms and constraints governing the evolution of correlated traits in drosophilid flies. Nature 424: 931–935.

Gross, M.K., Moran-Rivard, L., Velasquez, T., Nakatsu, M.N., Jagla, K., and Goulding, M. 2000. Lbx1 is required for muscle precursor migration along a lateral pathway into the limb. Development 127: 413–424.
Halton, M.S., Carmina, A., Gisselbrecht, S., Sackerson, C.M., Jimenez, F., Baylies, M.K., and Michelson, A.M. 2000. Ras pathway specificity is determined by the integration of multiple signal-activated and tissue-restricted transcription factors. Cell 103: 63–74.

Jagla, K., Frisch, M., Jagla, T., Dretzen, G., Bellard, F., and Bellard, M. 1997. ladybird, a new component of the cardiogenic pathway in Drosophila required for diversification of heart precursors. Development 124: 3471–3479.

Jagla, T., Bellard, F., Lutz, Y., Dretzen, G., Bellard, M., and Jagla, K. 1998. ladybird determines cell fate decisions during diversification of Drosophila somatic muscles. Development 125: 3699–3708.

Jagla, T., Bellard, F., Vonesch, J.L., Bellard, M., Dastugue, B., and Jagla, K. 1999. Plasticity within the lateral somatic mesoderm of Drosophila embryos. Int. J. Dev. Biol. 43: 571–573.

Jagla, T., Bidet, Y., Da Ponte, J.P., Dastugue, B., and Jagla, K. 2002. Cross-repressive interactions of identity genes are essential for proper specification of cardiac and muscular fates in Drosophila. Development 129: 1037–1047.

Junion, G., Jagla, T., Duplant, S., Tapin, R., Da Ponte, J.P., and Jagla, K. 2005. Mapping Dmef2-binding regulatory modules by using a ChIP-enriched in silico targets approach. Proc. Natl. Acad. Sci. 102: 18479–18484.

Kim, Y.O., Park, S.J., Balaban, R.S., Nirenberg, M., and Kim, Y. 2004. A functional genomic screen for cardiac regulatory genes using RNA interference in developing Drosophila embryos. Proc. Natl. Acad. Sci. 101: 159–164.

Kniir, S., Azpiazu, N., and Frisch, M. 1999. The role of the NK-homeobox gene slouch (S59) in somatic muscle patterning. Development 126: 4525–4535.

Lee, H.H. and Frisch, M. 2000. Wingless effects mesoderm patterning and ectoderm segmentation events via induction of its downstream target sloppy paired. Development 127: 5497–5508.

Liu, J., Qian, L., Wessells, R.J., Bidet, Y., Jagla, K., and Bodmer, R. 2006. Hedgehog and RAS pathways cooperate in the anterior–posterior specification and positioning of cardiac progenitor cells. Dev. Biol. 290: 373–385.

Livak, K.J. and Schmittgen, T.D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(−ΔΔCT) method. Methods 25: 402–408.

Maqbool, T., Soler, C., Jagla, T., Daczewska, M., Lodha, N., Palliyil, S., Vijayaraghavan, K., and Jagla, K. 2006. Shaping leg muscles in Drosophila: Role of ladybird, a conserved regulator of appendicular myogenesis. PLoS ONE 1: e122. doi: 10.1371/journal.pone.0000122.

Mootooosamy, R.C. and Dietrich, S. 2002. Distinct regulatory cascades for head and trunk myogenesis. Development 129: 573–583.

Neyt, C., Jagla, K., Thisse, C., Thisse, B., Haines, L., and Currie, P.D. 2000. Evolutionary origins of vertebrate appendicular muscle. Nature 408: 82–86.

Olson, E.N. 2006. Gene regulatory networks in the evolution and development of the heart. Science 313: 1922–1927.

Paterson, B.M., Walldorf, U., Eldridge, J., Dubendorfer, A., Frisch, M., and Gehring, W.J. 1991. The Drosophila homologue of vertebrate myogenic-determination genes encodes a transiently expressed nuclear protein marking primary myogenic cells. Proc. Natl. Acad. Sci. 88: 3782–3786.

Philippakis, A.A., Busser, B.W., Gisselbrecht, S.S., He, F.S., Estrada, B., Michelson, A.M., and Bulyk, M.L. 2006. Expression-guided in silico evaluation of candidate cis regulatory codes for Drosophila muscle founder cells. PLoS Comput. Biol. 2: e53. doi: 10.1371/journal.pcbi.0020053.

Rebeiz, M., Reeves, N.L., and Posakony, J.W. 2002. SCORE: A computational approach to the identification of cis-regulatory modules and target genes in whole-genome sequence data. Site clustering over random expectation. Proc. Natl. Acad. Sci. 99: 9888–9893.

Reddy, K.L., Weihlwill, A., Dzitoeva, S., Lin, M.H., Holbrook, S., and Storti, R.V. 2000. The Drosophila PAR domain protein 1 (Pdp1) gene encodes multiple differentially expressed mRNAs and proteins through the use of multiple enhancers and promoters. Dev. Biol. 224: 401–414.

Riechmann, V. and Ephrussi, A. 2001. Axis formation during Drosophila oogenesis. Curr. Opin. Genet. Dev. 11: 374–383.

Roulet, E., Busso, S., Camargo, A.A., Simpson, A.J., Mermod, N., and Bucher, P. 2002. High-throughput SELEX SAGE method for quantitative modeling of transcription-factor binding sites. Nat. Biotechnol. 20: 831–835.

Sandmann, T., Jensen, L.J., Jakobsen, J.S., Karzynski, M.M., Eichenlaub, M.P., Bork, P., and Furlong, E.E. 2006. A temporal map of transcription factor activity: mef2 directly regulates target genes at all stages of muscle development. Dev. Cell 16: 797–807.

Skeath, J.B. 1999. At the nexus between pattern formation and cell-type specification: The generation of individual neuroblast fates in the Drosophila embryonic central nervous system. Bioessays 21: 922–931.

Tautz, D., and Pfeifle, C. 1989. A non-radioactive in situ hybridization method for the localization of specific RNAs in Drosophila embryos reveals translational control of the segmentation gene hunchback. Chromosoma 98: 81–85.

Vasyutina, E., Stebler, J., Brand-Saberi, B., Schulz, S., Raz, E., and Birchmeier, C. 2005. CXCR4 and Gab1 cooperate to control the development of migrating muscle progenitor cells. Genes & Dev. 19: 2187–2198.

Xu, X., Yin, Z., Hudson, J.B., Ferguson, E.L., and Frisch, M. 1998. Smad proteins act in combination with synergistic and antagonistic regulators to target Dpp responses to the Drosophila mesoderm. Genes & Dev. 12: 2354–2370.

Zaffran, S. and Frisch, M. 2002. Early signals in cardiac development. Circ. Res. 91: 457–469.

Zaffran, S., Reim, I., Qian, L., Lo, P.C., Bodmer, R., and Frisch, M. 2006. Cardioblast-intrinsic Tinman activity controls proper diversification and differentiation of myocardial cells in Drosophila. Development 133: 4073–4083.

Zhou, Q. and Wong, W.H. 2004. CisModule: De novo discovery of cis-regulatory modules by hierarchical mixture modeling. Proc. Natl. Acad. Sci. 101: 12114–12119.

Zikova, M., Dastugue, B., and Jagla, K. 2003. Patterning of the cardiac outflow region in Drosophila. Proc. Natl. Acad. Sci. 100: 12189–12194.