

ショウジョウバエ原腸胚における
1細胞トランスクリプトームと
その細胞間差異

坂口 峻太

目次

要旨	5
略語表	7
第1章 序論	8
1.1 位置情報に基づいて細胞運命が決定する	8
1.2 ショウジョウバエ胚は発生生物学における重要なモデルである	9
1.2.1 ショウジョウバエ胚のパターニング	9
1.2.2 位置情報の欠失した胚における表現型	10
1.2.3 ショウジョウバエ原腸胚における位置に応じた細胞の振る舞い	11
1.3 ショウジョウバエ胚の1細胞トランスクリプトーム情報は不十分	14
1.4 本研究で行なったこと	15
第2章 結果	17
2.1 scRNA-seq データの取得	17
2.2 トリプシン処理による Notch ターゲット遺伝子のアーティフィシャルな活性化	18
2.3 Set 3 では遺伝子発現に基づき 77 のサブクラスタが検出できた	20
2.4 異なる細胞種の中間的な遺伝子発現を示すサブクラスタの検出	22
2.5 細胞膜関連遺伝子の発現は三胚葉の情報を強く反映する	23
2.6 ストライプ状遺伝子発現のトランスクリプトームレベルでの再構成	25
2.7 <i>bcd</i> -RNAi によって胚前方の細胞はトランスクリプトームレベルで後方化する	27
2.8 遺伝子発現のゲノムワイドな空間再構成	30
第3章 考察	33
3.1 ショウジョウバエ原腸胚1細胞トランスクリプトームアトラス	33
3.2 トリプシンによる細胞解離処理の影響	33

3.3 細胞分化の中間的な発現を示す細胞.....	34
3.4 <i>bcd-RNAi</i> 胚における遺伝子発現の収斂.....	35
3.5 細胞膜関連遺伝子のトランスクリプトームへの寄与.....	36
3.6 位置情報から細胞トランスクリプトームへの非線形変換.....	37
3.7 今後の展望.....	38
第4章 材料と手法.....	40
4.1 使用したショウジョウバエ系統.....	40
4.2 <i>scRNA-seq</i> の実行.....	40
4.2.1 細胞懸濁液の準備.....	40
4.2.2 C1HT による <i>scRNA-seq</i>	41
4.2.3 10x Chromium による <i>scRNA-seq</i>	42
4.3 <i>scRNA-seq</i> データの解析.....	42
4.3.1 10x Chromium データの前処理および Seurat クラスタリング.....	42
4.3.2 C1HT データの前処理および Seurat クラスタリング.....	44
4.3.3 NK-data の前処理および Seurat クラスタリング.....	44
4.3.4 トリプシンデータからの高 <i>tsr</i> 発現細胞の除去.....	45
4.3.5 それぞれの <i>scRNA-seq</i> データのサブクラスタリング.....	46
4.3.6 Harmony による <i>scRNA-seq</i> データの統合.....	46
4.3.7 統合データセットのサブクラスタリング.....	47
4.3.8 Gene Ontology term enrichment 解析.....	47
4.3.9 GLAD を用いた階層的クラスタリング.....	48
4.3.10 細胞列への割り当てと細胞列間の DEG 解析.....	48
4.3.11 Set 2 データと <i>bcd-RNAi</i> データのマージ.....	49
4.3.12 Set 2 データと <i>bcd-RNAi</i> データの間の DEG 解析.....	50
4.3.13 遺伝子発現の空間再構成.....	50

4.4 バルク RNA-seq.....	53
4.4.1 バルク RNA-seq の実行.....	53
4.4.2 バルク RNA-seq データの解析.....	54
4.5 SABER-FISH.....	55
4.6 プロット.....	56
図.....	57
表.....	127
引用文献.....	143
注釈.....	152
謝辞.....	152

要旨

多くの動物の胚発生ではモルフォゲンの勾配に基づく位置情報に依存して、細胞は特定の振る舞いや機能を示す。この過程では、転写因子の組み合わせが細胞の転写制御を介してトランスクリプトームを確立し、それに基づいて細胞の振る舞いが決定されると考えられている。しかし、転写因子がトランスクリプトームを介して細胞の振る舞いを制御するルールの理解は不十分である。その原因として、トランスクリプトームの細胞間差異が十分な空間解像度で明らかになっていないことが挙げられる。

そこで、ショウジョウバエ原腸胚をモデルに、1細胞トランスクリプトームの細胞間差異を明らかにすることを本研究の目的とする。本研究では第一に、クオリティーの高い1細胞トランスクリプトームデータを取得するため、独自に1細胞RNA-seqを行った。第二に、このデータに対して、クラスタリングとアノテーションを行い、胚における位置に対応する77のクラスタに細胞を分類できた。これにより高い空間解像度で細胞のトランスクリプトームが明らかになった。第三に、どのような遺伝子が細胞間の差異に寄与するかを調べるため、細胞間で発現の差が大きい遺伝子にどのカテゴリの遺伝子が多く含まれているかを調べたところ、転写因子に加え、細胞膜関連遺伝子が多く含まれていた。このことから、細胞間の差異に細胞膜関連遺伝子の発現の差異が強く寄与することが明らかになった。また、遺伝子はその種類ごとにどのような情報を持っているかを調べるため、特定のカテゴリの遺伝子のみを用いてクラスタリング解析を行ったところ、転写因子の発現パターンは空間に対応したクラスタを生じて事前知識なしに三胚葉を識別できないのに対し、細胞膜関連遺伝子の発現パターンは三胚葉を識別できた。この結果は転写因子の発現と細胞膜関連遺伝子の発現の間の非線形な変換の存在を示唆する。第四に、胚帯伸長を制御することが知られている8細胞列からなる遺伝子発現の空間的繰り返し単位内における遺伝子発現の差異を明らかにするため、この遺伝子発現の繰り返し単位をゲノムワイドに再構成した。これを用いて細胞列間で発現の異なる遺伝子を検出したところ、多くの細胞膜関連遺伝子が検出された。このことは細胞列間のトランスクリプトームの差異にも細胞膜関連因子の

寄与が大きいことを示す。第五に、位置情報の異常がトランスクリプトームをどのように変化させるかを明らかにするため、*bicoid* (*bcd*) 遺伝子機能阻害胚について 1 細胞 RNA-seq を行い、コントロール胚と比較した。その結果、*bcd* 遺伝子機能阻害胚における胚前方の細胞を後方の細胞から識別できなかった。この結果は胚前方細胞がトランスクリプトームレベルで後方化していることを示す。最後に、独自に開発した遺伝子発現の空間再構成手法を野生型胚に適用し、遺伝子発現の空間パターンをゲノムワイドかつこれまでより正確に再現した。

以上のように、本研究ではショウジョウバエ原腸胚の 1 細胞遺伝子発現をトランスクリプトームレベルで明らかにし、その細胞間差異に転写因子に加えて細胞膜関連遺伝子の寄与が大きいことを見出した。この事実は、転写因子が非線形な変換を介して細胞膜関連遺伝子の発現パターンの差異を生み出し、その差異が局所的な細胞間コミュニケーションを調整し、トランスクリプトームを細胞の振る舞いに変換していることを示唆する。

略語表

BDGP	Berkeley Drosophila genome project
BDTNP	Berkeley Drosophila transcription network project
CAP	cold active protease
<i>E(spl)-C</i>	Enhancer of split complex
FISH	fluorescence <i>in situ</i> hybridization
LOOCV	leave-one-gene-out cross validation
PS	parasegment
PCA	principal component analysis
RNA-seq	RNA-sequencing
scRNA-seq	single cell RNA-sequencing
snRNA-seq	single nucleus RNA-sequencing
t-SNE	t-distributed stochastic neighbor embedding
UMAP	uniform manifold approximation and projection
UMI	unique molecular identifier

第1章 序論

1.1 位置情報に基づいて細胞運命が決定する

胚発生では胚における位置に応じて特定の細胞の分化と、形態形成を駆動する細胞の特定の振る舞いとが生じる必要がある。この過程がどのように制御されているかを明らかにすることが発生生物学の基本的な問いの1つである。Nüsslein-Volhard と Wieschaus によるキイロショウジョウバエ (*Drosophila melanogaster*、以下、ショウジョウバエ) のパターンニング遺伝子についての解析 (Nüsslein-Volhard and Wieschaus, 1980)をはじめ、現在までに多くの遺伝学的研究がなされ、数多くの遺伝子が胚発生の制御に寄与していることが知られている。しかし、これらの研究は少数の遺伝子ごとに個別に行われてきたため、多数の遺伝子がどのように協調して胚発生を制御しているのかについては十分に明らかではなく、現在における発生生物学の重要な課題となっている。

多くの動物の胚発生に共通するメカニズムとしてモルフォゲンによる制御が知られている。すなわち、特定の分子 (モルフォゲン) の勾配によって形成される位置情報に基づいて個々の細胞の運命が決定し、細胞は特定の振る舞いや機能を示すというモデルである (Briscoe and Small, 2015; Gilmour et al., 2017) (図 1)。この過程は濃度勾配という連続的な情報から、細胞運命という離散的な情報への非線形な変換である。この変換においては、転写因子の組み合わせが重要である。位置情報に応じて特定の転写因子の組み合わせが生じ、その転写因子の組み合わせは転写制御を介して領域ごとに異なるトランスクリプトームを確立する。そして、そのトランスクリプトームの違いが異なる細胞の振る舞いを生み出すと考えられている。ゲノムワイドな研究の進展により遺伝子発現の制御や細胞の分化についての理解は深まりつつある一方で (Long et al., 2016; Stricker et al., 2016)、転写因子がトランスクリプトームを介して細胞の振る舞いを制御するルールの理解は不十分である。このルールを明らかにするためには、1) 胚の領域間での細胞の振る舞いの違いを明らかにすること、2) 胚の領域間での細胞のトランスクリプトームの違いを明らかにすること、3) それら

を対応づけることが必要である。後述するショウジョウバエ原腸胚の例のように、胚における細胞の振る舞いは詳細に観察されている。その一方で、胚の各細胞のトランスクリプトームは十分な空間解像度で明らかになっていない。

1.2 ショウジョウバエ胚は発生生物学における重要なモデルである

ショウジョウバエは豊富な遺伝学的リソースがあることに加え、多数の先行研究により、その胚発生における多様な現象が詳細に観察され、それらを制御する遺伝子や分子機構が明らかにされている。また、ショウジョウバエ胚は、発生過程において位置情報が確立され、転写因子によって読み出される過程が詳細に研究されている。ショウジョウバエ原腸胚では三胚葉が出現する（図 2）のと並行して、細胞はその領域ごとに決まった多様な振る舞いを示して（図 3）、形態形成を駆動する。したがって、特にショウジョウバエ原腸胚は転写因子がトランスクリプトームを確立し、そのトランスクリプトームに基づいて細胞が振る舞いを決定するルールを明らかにするために有用なモデルである。以下ではまず、ショウジョウバエ胚のパターニングについて述べ、そのあとショウジョウバエ原腸胚における細胞の振る舞いについて説明する。

1.2.1 ショウジョウバエ胚のパターニング

ショウジョウバエの胚発生では、背腹軸あるいは前後軸に沿った細胞の運命が、それぞれ別のモルフォゲンの濃度に依存した転写の制御により決定される。背腹軸に沿って働くモルフォゲンは転写因子 Dorsal タンパク質であり、その核内での濃度は腹側で高く、背側では低い。背腹軸に沿って、核内の Dorsal の濃度に応じて異なる遺伝子が発現し、腹側から順に中胚葉、正中線細胞、神経外胚葉、背側外胚葉、羊漿膜が形成される(Hong et al., 2008; Reeves and Stathopoulos, 2009; Steward and Govind, 1993) (図 4A)。

一方、ショウジョウバエ胚の前後軸形成は、卵細胞の段階で前方には転写因子をコードする *bicoid* mRNA (Berleth et al., 1988)が、後方には RNA 結合タンパク質をコードする *nanos*

mRNA (Gavis and Lehmann, 1992)が局在することから始まる (図 4B)。この mRNA から翻訳された Bicoid タンパク質と Nanos タンパク質はそれぞれ胚の前端あるいは後端をピークとする濃度勾配を形成し、胚に前後軸の情報を与える(Driever and Nüsslein-Volhard, 1988; Gavis and Lehmann, 1992)。また、卵形成の段階で母性因子の受容体型チロシンキナーゼである Torso (Tor)が胚末端部で活性化する(Casanova and Struhl, 1989)。

Bicoid と Nanos の濃度勾配および Tor 活性は、ギャップ遺伝子と呼ばれる転写因子をコードする遺伝子群の発現を制御し、ギャップ遺伝子は胚の前後軸の中で特定の位置にのみ発現するようになる(Driever and Nüsslein-Volhard, 1988; Gavis and Lehmann, 1992; Paroush et al., 1997)。ギャップ遺伝子の働きはさらに次の階層に属する、ペアルール遺伝子の発現につながる (図 4B)。ペアルール遺伝子前後軸に沿って 7 本のストライプ状に発現する。このストライプ状の発現がパラセグメントと呼ばれる 14 の繰り返り単位を規定する。そしてその制御下で *wingless (wg)* や *engrailed (en)* といったセグメントポラリティー遺伝子が発現する。*wg* と *en* は隣接する重なり合わないストライプを形成し、パラセグメント境界を安定化させることが知られている(Ingham and Arias, 1992)。*en* あるいは *wg* 発現細胞はそれぞれパラセグメントの前端と後端でシグナルセンターとして機能し、パラセグメント内の極性を形成する(Heemskerk and DiNardo, 1994)。その結果、パラセグメント内の各細胞列はそれぞれ異なったペアルール遺伝子とセグメントポラリティー遺伝子の組み合わせを持つようになる (Clark and Akam, 2016) (図 5)。

1.2.2 位置情報の欠失した胚における表現型

以上のように、ショウジョウバエ原腸胚では位置情報を読み出す仕組みが詳細に明らかにされている。それに加え、位置情報を欠失した胚は深刻な表現型を示すことが知られている。例として、*bcd* 変異体の母親から生まれた胚は胸部から前の部位が欠失し、胚の後方の構造に置き換わることが知られている (Frohnhofer and Nüsslein-Volhard, 1986、図 6)。遺伝子発現の面でも、*bcd* 欠失胚では胚前方の特徴が消失することが知られている。例えば、胚前方で発現する転写因子である *Deformed (Dfd)* と *ocelliless (oc)* の発現は *bcd* 欠失胚では

消失する(Finklstien and Perrimon, 1990; Jack and McGinnis, 1990)。また、*bcd* 欠失胚において、前後軸上の位置ごとに、その位置におけるギャップ遺伝子の発現が野生型のどの位置における発現パターンに相当するかを調べた解析(Petkova et al., 2019)により、6番目のパラセグメント付近を境に前方の細胞の遺伝子発現が後方の細胞の遺伝子発現に置き換わっていることが示唆されている。6番目のパラセグメントは胸部と腹部の境界に相当するため(Martinez-Arias and Lawrence, 1985)、このことは胸部から前が欠失することと合致している。

しかし、これらの後方化についての研究は少数の遺伝子の発現にのみ着目しており、遺伝子発現がトランスクリプトームレベルで後方化しているかは明らかではない。また、*bcd* 機能阻害胚においては、胚前方の細胞の遺伝子発現の履歴は後方領域とは異なっている。例えば、*bcd* 機能阻害胚において、ギャップ遺伝子の一つである *hunchback* (*hb*) 遺伝子の発現の開始は胚前方領域において、胚後方領域より遅い(Staller et al., 2015)。このような遺伝子発現の履歴の違いが最終的なトランスクリプトームに影響を与える。このため、前方細胞の後方細胞への転換をトランスクリプトームレベルで見ると、前方細胞と後方細胞の特徴が混在した細胞や、野生型原腸胚には全く存在しない特徴を持つ細胞が存在する可能性がある。

1.2.3 ショウジョウバエ原腸胚における位置に応じた細胞の振る舞い

ここまでにショウジョウバエ胚がモルフォゲン勾配から位置情報を読み出して領域特異的な転写因子を発現させる過程について述べた。ショウジョウバエ原腸胚では、領域特異的に発現する遺伝子の制御下において、細胞が多様な振る舞いを示すことで、組織変形が駆動され、単純なラグビーボール状の構造から複雑な身体が作られ始める(図3)。以下では、ショウジョウバエ原腸胚における細胞の振る舞いを、それが駆動する組織変形ごとに説明する。

腹溝形成

原腸形成期のショウジョウバエ胚における形態形成現象の中で、最もよく研究されてい

るものの一つが、腹溝形成である。腹溝形成は胚の腹側の18細胞幅、60細胞の長さを持つ領域で各細胞の表面が平坦化することから始まる(Sweeton et al., 1991)。この領域の細胞のうち、正中線付近の12細胞幅の細胞群(Sweeton et al., 1991)では、頂端面中央部のアクチンネットワークがパルス様に活性化する。これにより頂端面の収縮が引き起こされ、それが安定化されることが繰り返されることによって、頂端収縮が進行していく (Martin et al., 2009、図 7A)。その後、これらの細胞は細胞の基底側が広がり、頂端基底軸方向に短縮を起こし、組織が陥入を始める(Sweeton et al., 1991)。この制御には細胞の基底側でミオシンが減少することが必要である(Krueger et al., 2018)。Dorsalの標的遺伝子である転写因子の *twist (twi)* や *snail (sna)* の変異体では腹溝形成は起こらない(Leptin and Grunewald, 1990)。

後部内胚葉の陥入

腹溝形成と同様に頂端面中央部でのミオシン活性による頂端収縮によって駆動される陥入現象として、後部内胚葉の陥入がある。後部内胚葉の陥入では、後端部から背側前方にかけて、7細胞分の領域で細胞が頂端面中央部のミオシンの活性化によって頂端収縮を起こし、陥入を開始する。このとき、細胞は基底側で少し広がり、前方に隣接する細胞の基底側を前方かつ頂端側に押し出す。隣接細胞の頂端側は胚を覆うビテリン膜に押し広げられる。この力学的ストレスが隣接細胞でのミオシンの活性化を誘導し、その細胞の頂端収縮を引き起こす。これが繰り返されることにより、ミオシン活性と頂端収縮が胚の背側を前方に向かって伝播する (図 7B)。この伝播が後部内胚葉の陥入の特徴であり、固定された領域でミオシンの活性化と頂端収縮を起こす中胚葉陥入と大きく異なる点である(Bailles et al., 2019)。この後部内胚葉の陥入には Tor の制御下で発現するギャップ遺伝子 *tailless (tll)* と *huckebein (hkb)* が必要であり、その変異体では後部内胚葉の陥入が異常となる(Paroush et al., 1997; Weigel et al., 1990)。

頂端面のミオシン活性化を伴わない組織の陥入

組織の陥入には頂端面のミオシン活性による収縮を伴わないものも存在する。その例の

一つが胚の背側で起きる Dorsal fold の形成であり、もう一例は胚前方で、頭部と胸部とを分ける cephalic furrow の形成である。Dorsal fold の形成は胚の背側の二か所の細胞列 (initiator 細胞) において、隣接細胞との間の接着結合が基底方向へシフトを起こすことから始まる (図 7C)。これに続いて頂端面が狭まり、細胞が頂端基底軸方向に短縮する。このとき、頂端面中央部のミオシンの活性化は起こらない。initiator 細胞における接着結合の基底方向へのシフトにより、隣接する細胞における対称性が崩れ、隣接細胞は initiator 細胞のほうへ屈曲する(Wang et al., 2012)。形成された二つの Dorsal fold のうち、前方の fold よりも後方の fold のほうが深く陥入する。Dorsal fold はペアルール遺伝子である *runt (run)* の特定のストライプの位置に形成される(Wang et al., 2012)。Cephalic furrow の陥入はペアルール遺伝子である *even-skipped (eve)* の 1 本目のストライプ上において、細胞の高さが短縮することから始まる (Vincent et al., 1997、図 7D)。

胚帯の収斂伸長

原腸形成期には、これらの陥入だけでなく、組織の収斂伸長も起きる。胚側面 (胚帯) の細胞は背腹軸方向にインターカレーションを起こすことによって組織を前後軸方向に伸長させる (Irvine and Wieschaus, 1994、図 8)。インターカレーションは、前後軸に垂直な細胞境界 (AP 境界) が短縮し、代わりに前後軸に並行な境界 (DV 境界) が形成されることによって起きる (Bertet et al., 2004)。このとき、AP 境界にはミオシンが局在し、DV 境界には *Bazooka / Par-3* が局在することが知られている (Zallen and Wieschaus, 2004)。AP 境界の短縮を駆動するのは頂端面中央部のミオシン活性であり、AP 境界のミオシンは AP 境界の短縮を安定化させると考えられている (Rauzi et al., 2010)。この収斂伸長は前後軸のパターニングによって制御されていることが知られており、特に、ペアルール遺伝子である *eve* や *run* の変異体では組織の伸長幅と細胞のインターカレーションが減少する (Irvine and Wieschaus, 1994; Paré et al., 2014)。また、*eve*、*run* の一様発現、異所的発現は平面内細胞極性を変化させる (Zallen and Wieschaus, 2004)。

1.3 ショウジョウバエ胚の1細胞トランスクリプトーム情報は不十分

以上のように、ショウジョウバエ原腸胚における多様な細胞の振る舞いは詳細に観察されている。また、それを制御する転写因子やその上流の位置情報の読み出し機構が明らかにされている。一方で、以下で述べるように、ショウジョウバエ原腸胚におけるトランスクリプトーム情報は不十分である。

ショウジョウバエ原腸胚において、これまでに、多くの遺伝子の mRNA 発現は *in situ* hybridization (ISH)によって調べられている。しかし、それらのデータから1細胞トランスクリプトームを得ることは困難である。多くのデータは定量的でなく、また、定量的に行われている場合(Alberga et al., 1991; Fowlkes et al., 2008; Luengo Hendriks et al., 2006)でもその遺伝子数は限られる。

近年、1細胞 RNA シーケンシング (scRNA-seq) が一般化したことにより、1細胞レベルのトランスクリプトーム情報を取得することが可能になっている(Kolodziejczyk et al., 2015; Tanay and Regev, 2017)。さらに、scRNA-seq では、細胞を組織から1細胞レベルに解離する必要があり、その際に細胞が組織のどの位置に由来するかの情報が失われるが、遺伝子発現を元に scRNA-seq データの空間情報を復元し、遺伝子発現の空間パターンを再構成するための計算手法もいくつか開発されている(Karaiskos et al., 2017; Moriel et al., 2021; Satija et al., 2015)。scRNA-seq およびそれを元にした遺伝子発現の空間再構成はショウジョウバエ原腸胚に対してもすでに適用されている(Karaiskos et al., 2017)。しかし、このデータには改善の余地がある。第一に、このデータが含んでいる細胞数は1,297細胞であり、原腸胚全体の細胞数であるおよそ6,000細胞と比較して大幅に少ない。第二に、クラスタリング解析において13のクラスタしか検出できておらず、ショウジョウバエ原腸胚全体を詳細に表現できているとは言えない。第三に、多くの遺伝子で、scRNA-seq データから再構成された空間パターンが、*in situ* hybridization で明らかにされた実際のパターンと一致しない。この中にはセグメントポラリティー遺伝子 (*wg*, *en*) の特徴的な14本のストライプからなる発現も含まれる (図9)。これとは別に、1細胞核 RNA-seq (snRNA-seq) によるショウジ

ショウバエ胚の時系列トランスクリプトーム解析も行われている(Calderon et al., 2022)。しかし、この時系列データから原腸胚のステージだけを厳密に分離することは困難である。

1.4 本研究で行なったこと

ここまでで述べたように、ショウジョウバエ原腸胚は発生生物学における有用なモデルであるにもかかわらず、正確な空間情報をもつ 1 細胞トランスクリプトームデータが不足している。そのため、細胞間でのトランスクリプトームの差異が十分に明らかではなく、細胞のトランスクリプトームと細胞の振る舞いを対応づけることができていない。そこで、ショウジョウバエ原腸胚の十分な空間解像度を持つ 1 細胞トランスクリプトーム情報を取得し、トランスクリプトームが細胞間でどのような差異を示すかを明らかにすることを本研究の目的とする。

本研究では、第一に、既存のものよりもクオリティーの高い 1 細胞トランスクリプトームデータを取得するため、独自にショウジョウバエ原腸胚の scRNA-seq を行なった。その際、cold active protease (CAP)を用いた細胞解離手法を導入し、アーティフィシヤルな遺伝子発現変動を抑制した。第二に、細胞に空間情報を付与するために、このデータにクラスタリング解析を行って 77 個のクラスタを検出した。これにより高い空間解像度で細胞のトランスクリプトームが明らかになった。第三に、遺伝子カテゴリごとのトランスクリプトームへの寄与のしかたを調べ、細胞膜関連遺伝子がトランスクリプトームの差異に大きく寄与し、転写因子の発現パターンよりも細胞膜関連遺伝子の発現パターンのほうがより三胚葉の分化状態を反映していることを示した。第四に、胚帯伸長を制御することが知られている遺伝子発現の 8 細胞列からなる空間的繰り返し単位内における遺伝子発現の差異を明らかにするため、この遺伝子発現の繰り返し単位をゲノムワイドに再構成した。これを用いて、細胞列境界で発現に違いのある遺伝子を検出し、転写因子に加え細胞膜関連遺伝子が多く検出された。このことは細胞列間のトランスクリプトーム差異にも細胞膜関連因子の寄与が大きいことを示唆する。第五に、*bcd* 遺伝子の機能阻害胚の scRNA-seq データを

用いて、*bcd* 遺伝子の機能阻害胚の前方細胞がトランスクリプトームレベルで後方化していることを示した。最後に、scRNA-seq データの空間再構成を行い、遺伝子発現の空間パターンをゲノムワイドかつこれまでより正確に再現した。

第2章 結果

2.1 scRNA-seq データの取得

先行研究(Karaiskos et al., 2017)のデータよりもクオリティーの高い1細胞トランスクリプトームデータを取得するため、細胞の解離プロセスの検討をおこなった(図 10)。まず、先行研究と同じホモジェナイザーによる手法を試したが、十分な量の細胞数が得られなかった。そこで、ビテリン膜を壊したあとに酵素処理によって細胞を解離する手法を試した。使用する酵素にはまずトリプシンを選択した。しかし、先行研究により常温での酵素処理は細胞の遺伝子発現に影響を与える可能性が示されていた(Liu et al., 2014)。この問題に対する解決策として、Cold active protease (CAP) を用いた低温での細胞解離が提唱されている(Adam et al., 2017; O’Flanagan et al., 2019)。そこで、トリプシンに加えて、CAPである細菌 *Bacillus licheniformis* 由来のサブチリシン A (以下、単に CAP と書く) を用いた低温での細胞の解離も試した。加えて、細胞の溶解プロセス時に細胞の遺伝子発現が変化することを防ぐため、解離した細胞を非架橋の固定試薬である CellCover を用いて固定した。解離後固定した細胞と解離直後の細胞の間でバルク RNA-seq を行ったところ、両者は強く相関し(図 11)、固定細胞は解離直後の細胞の転写プロファイルを維持していることが示された。

先行研究によると、Fluidigm C1 HT IFC は 10x Genomics Chromium と比較して、一度に取得できる細胞数が少ない代わりにより多くの遺伝子を検出できる(Ashton et al., 2021)。いずれのプロトコールがトランスクリプトームの細胞間差異を知るのに適しているかを判断するため、scRNA-seq において両者のプロトコールを試し、以下の三種類のデータセットを作成した。Set 1 はトリプシンで細胞を解離し、プラットフォームに Fluidigm C1 HT IFC を用いた。Set 2 はトリプシンで細胞を解離し、10x Genomics Chromium V3.1 を用いた。Set 3 は CAP で細胞を解離し、10x Genomics Chromium V3.1 を用いた(図 10)。全てのデータについて、各転写産物には逆転写時に unique molecular identifier (UMI) を導入し、各細胞の各遺伝子について UMI の重複を除いて転写産物をカウントした(Islam et al., 2014)。フィルタリングを行なって低クオリティー細胞を除去したところ(「材料と手法」を参照)、Set 1 で 1,243

細胞、Set 2で7,314細胞、Set 3で6,180細胞が高クオリティー細胞（HQC）として残った。HQCについて細胞ごとの検出遺伝子数を数えたところ、その中央値はSet 1で4,480、Set 2で3,222、Set 3で4,053であった。細胞ごとの合計 UMI 数はSet 1が152,429と他のデータ（Set 2: 22,506、Set 3: 37,610）より高い値であった。Set 1は4つのバッチからなるが、バッチエフェクトは確認できなかった（図 12）。

各データが偏りなく胚の細胞を取得できていることを確かめるため、scRNA-seq解析用Rパッケージ Seurat v.3 (Stuart et al., 2019)を用いたグラフベースクラスタリングを行った(図 13A-B、図 14A-B)。以下では全細胞のデータに対して Seurat によるクラスタリングを用いて検出されたクラスタを Seurat クラスタ、または単にクラスタと呼ぶ。各クラスタのマーカ遺伝子を検出すると、各データセットにおいて、中胚葉マーカ (*sna*, *twi*)、内胚葉マーカ (*fork head (fkh)*)、背側外胚葉マーカ (*decapentaplegic (dpp)*)、神経外胚葉マーカ (*short gastrulation (sog)*)、頭部外胚葉マーカ (*Optix*, *oc*)、羊漿膜マーカ (*pebbled (peb)*) をそれぞれ発現するクラスタが検出された(図 13A-B、図 14A-B)。このことは、それぞれのデータセットが主要な細胞種を全て含んでいることを示す。

2.2 トリプシン処理による Notch ターゲット遺伝子のアーティフィシヤルな活性化

解離手法が細胞のトランスクリプトームに与える影響を調べるため、各クラスタのマーカ遺伝子の発現を比較した。Set 3 データにおいて Notch ターゲット遺伝子である *single-minded (sim)* と *Enhancer of split* コンプレックス (*E(spl)-C*) の遺伝子が Seurat クラスタ 16 のマーカ遺伝子として検出された。実際、*sim* および *Enhancer of split m8, helix-loop-helix (E(spl)m8-HLH)* の発現はクラスタ 16 に局在していた (図 13C-E)。*sim* および *E(spl)-C* 遺伝子の発現は正中線細胞 (midline cells) に局在することが知られていることから(Cowden and Levine, 2002; Morel and Schweisguth, 2000; Zinzen et al., 2006)、クラスタ 16 が正中線細胞として識別された。しかし、Set 2 細胞において、正中線細胞は *sim* の発現によって識別可

能であるものの（クラスタ 15）、*E(spl)m8-HLH* の発現は正中線細胞だけではなく他のクラスタでも高くなっていた（図 13G-I）。この *E(spl)-C* の発現上昇がトリプシン処理によるものであることを確かめるため、トリプシンによる細胞解離前の胚と細胞解離後の細胞（非固定）とでバルク RNA-seq を行い、両者の遺伝子発現を比較した。その結果、両者の遺伝子発現は強く相関したが、*E(spl)-C* の発現は解離細胞で有意に上昇していた（図 15A、B）。このことは、トリプシン処理が細胞種によらず *E(spl)-C* の発現上昇を引き起こすことを示す。以上の結果により、ショウジョウバエ原腸胚に対するトリプシン処理が Notch ターゲット遺伝子の発現を上昇させることが明らかになった

さらに、Set 2 は *twinstar* (*tsr*) の発現の高い Seurat クラスタを含んだ。たとえば、体幹部の中胚葉（trunk mesoderm）はクラスタ 1 と 10 の二つに分かれ、クラスタ 10 はクラスタ 1 と比較して高い *tsr* の発現を示した（図 13B、J）。また、クラスタ 6 は体幹部の外胚葉に属するクラスタであるが、同じく体幹部の外胚葉に属すると考えられるクラスタ 0、2、3 と比較して高い *tsr* の発現を示した（図 13B、J）。また、クラスタ 10 および 6 は他のクラスタと比較して低い *E(spl)m8-HLH* の発現を示した（図 13B、I）。これらの *tsr* 高発現細胞の特徴を調べるため、クラスタ 10 においてクラスタ 1 と比較して有意に高く発現している遺伝子を抽出し、それらについて Gene Ontology (GO) term enrichment 解析を行ったところ、GO ターム “oxidative phosphorylation” の濃縮が検出された（図 15C）。このことは *tsr* 高発現細胞が代謝的なストレス応答を起こした可能性を示唆する。

以上の結果は、Set 2 に二種類の細胞がいることを示唆する。一つはトリプシン処理中に Notch ターゲット遺伝子の発現が上昇した細胞であり、もう一つはトリプシン処理中に何らかのストレス応答を示したと考えられる細胞である。Set 2 と同様に、Set 1 でも広い *E(spl)-C* 遺伝子の強い発現と、*tsr* 高発現細胞のクラスタが検出された（図 14D、E）。一方、このような応答を示す細胞は Set 3 では検出されず（図 13E、F）、この応答がトリプシン処理に特異的であることが示された。また、この結果は CAP による低温での処理が細胞解離時の遺伝子発現の変動を抑えるのに有効であることを示す。なお、以降の解析では、トリプシン処理したデータについては、ストレス応答を示したと考えられる細胞は除去して用いた

(「材料と手法」を参照)。

2.3 Set 3 では遺伝子発現に基づき 77 のサブクラスタが検出できた

ショウジョウバエ原腸胚における 1 細胞トランスクリプトームを詳細な空間解像度で明らかにするため、サブクラスタリング解析と空間的アノテーションによって各細胞に位置の情報を割り当てた (「材料と手法」を参照)。アノテーションには Berkeley Drosophila Genome Project (BDGP) in situ データベース(Hammonds et al., 2013; Tomancak et al., 2007, 2002)、Fly-FISH データベース(<https://fly-fish.cabr.utoronto.ca>)および文献情報による既知の遺伝子発現パターンを用いた。前述したように、CAP による処理を行った Set 3 は他のデータと比較して細胞解離時の遺伝子発現の変動が抑えられていると考えられたため、まず、Set 3 に対するサブクラスタリングを行った。

体幹部 (Trunk) の領域では背腹軸に沿って、背側から羊漿膜 (amnioserosa)、背側外胚葉 (Dorsal ectoderm、DE)、神経外胚葉 (Neuroectoderm、NE)、正中線細胞 (midline cells)、中胚葉 (mesoderm) が出現している (図 4A) (Reeves and Stathopoulos, 2009)。一方、前後軸に沿って、細胞は 14 個のパラセグメント (PS) に分けられ、特に偶数番目の PS は *tartan* (*trn*) および *ftz* の発現によって特徴づけられる(Clark and Akam, 2016; Graham et al., 2019)。そこでまず、Set 3 の体幹領域の外胚葉細胞 (PS2-13) に関して (図 16A)、サブクラスタリングとアノテーションを前後軸、背腹軸について個別に行うことを試みた。前後軸のアノテーションでは、これらの細胞を PS2、腹部偶数 PS (PS4,6,8,10,12; abdominal_even)、腹部奇数 PS (PS3,5,7,9,11; abdominal_even)、PS13 の 4 つに分類することが可能であった (図 16B、C)。また、背腹軸を特徴づける 35 の遺伝子 (DV 遺伝子、「材料と手法」参照) を用いた k-means クラスタリングにより背腹軸に沿った羊漿膜、背側外胚葉 3 領域、神経外胚葉 3 領域の合計 7 領域に当てはめることが可能であった (図 16D)。これらのアノテーションの組み合わせにより、25 の領域に細胞を当てはめることに成功した (図 16E、F)。Set 3 の他のクラスタについてもサブクラスタリングを行ったところ(図 17)、全体で、胚における位置に

対応する 77 個のサブクラスタに分割することができた(図 18、表 2)。このサブクラスタリング結果は、細胞の転写情報が胚における位置を反映し、細胞を細かな領域に割り当てられるだけの情報を含んでいることを示す。特に、頭部 (PS1 より前) の領域は 15 のサブクラスタに分割することが可能であった(図 17A、図 18)。このことは頭部がトランスクリプトームレベルで細分化されることを示しており、後の頭部発生の複雑さを反映していると考えられる。なお、この過程で、ある胚葉に属する細胞でかつ他の胚葉のマーカー遺伝子の発現を示す細胞集団はダブレット細胞として除去し (図 19)、6,118 細胞が残った。

Set 3 と同様に、Set 1、Set 2 および先行研究(Karaiskos et al., 2017)によって取得されたデータである NK-data についてクラスタリングを行った (図 20、図 21、表 3-表 5)。Set 2 の Seurat クラスタリングにおいて、Set 3 と大きく異なっていた点は、Set 2 では腹側神経外胚葉 (ventral neuroectoderm) をパラセグメントの偶数番目/奇数番目に対応したクラスタに分割することができなかった点である (図 20A、B)。サブクラスタリングによって Set 2 は 55 のサブクラスタに分割できたが、ここでも腹側神経外胚葉を偶数番目/奇数番目のパラセグメントに背腹軸全体にわたって分離することはできなかった (図 20C)。これはトリプシン処理による遺伝子発現の変化が原因であると考えられる。

さらに Set 1、NK-data で検出できたサブクラスタ数はそれぞれ 32 および 28 であった (図 21、表 4、表 5)。この結果は細胞数が少ないために検出できたクラスタ数が少なくなったことを示唆し、シーケンス深度よりも細胞数の多さが、より少数の細胞種を検出するのに重要であるという先行研究(Heimberg et al., 2016; Zhang et al., 2020)と一致する。

同時により多くの細胞を解析することでクラスタの検出力が上がるならば、4つのデータセット全てを統合してクラスタリングすることで、より詳細な空間情報を付加したアトラスを作成することができる可能性がある。そこで、バッチコレクション手法 Harmony (Korsunsky et al., 2019)を用いて4つのデータセットを統合した。Harmonyによる補正なしでは4つのデータセットはUMAP上で部分的に混合しなかったが、Harmonyの適用により4つのデータセットはUMAP上で混合した(図 22A)。このデータについてSeuratによるクラスタリングおよびサブクラスタリングを行ったところ、検出されたサブクラスタ数は68で

あり、Set 3 単独の場合より少なかった (図 23、表 6)。特に、統合データのクラスタリングではパラセグメントに対応した分離が失敗する傾向にあった。まず、腹側神経外胚葉において *trn* が傾斜的な発現を示し、Seurat クラスタはパラセグメントの偶数番目、奇数番目に対応しなかった (図 22B、C)。さらに、サブクラスタリングを実行しても両者に対応するクラスタを検出することはできなかった (図 23、表 6)。この他、統合データのサブクラスタリングで検出できなかったクラスタは、PS13 の外胚葉、PS3 および PS14 の羊漿膜、PS4 と PS6 の中胚葉にそれぞれ対応するクラスタであった。いずれもパラセグメントに対応するクラスタであり、生物学的に意味のあるクラスタであった。この結果は、データの統合によって細胞数を増やすことが必ずしもクラスタリングの結果を向上させるとは限らず、特にトリプシン処理データのようなクオリティーの低いデータを加えることが、クラスタリングの性能を低下させうることを示す。

CAP 処理によって細胞を解離したほうがトリプシン処理の場合よりも本来の遺伝子発現、特に Notch ターゲット遺伝子の発現を維持していると考えられること、また、全てのデータセットを統合しても Set 3 単独の場合と比較してクラスタリングの解像度が上がらないことをふまえ、以下では原則として Set 3 を対象とした解析を行った。

2.4 異なる細胞種の中間的な遺伝子発現を示すサブクラスタの検出

サブクラスタリングにより、原腸胚には異なる細胞種の中間的な遺伝子発現を持つ細胞が存在することが明らかになった。1 つは Seurat クラスタ 18 のサブクラスタリングによって検出された "endoderm_antMG_wntD" である (図 17E)。Seurat クラスタ 18 は 3 つのサブクラスタに分割できた。そのうち "endoderm_antMG_wg" は *fkh*、*hkb*、*surpent (srp)* など内胚葉マーカー遺伝子を発現しており、"mesoderm_head" サブクラスタは中胚葉マーカー遺伝子である *sna*、*twi* などを発現していた。したがって、これらのクラスタはそれぞれ前方内胚葉と頭部中胚葉に分類できた。しかし、"endoderm_antMG_wntD" は *wnt inhibitor of Dorsal (wntD)* を特異的なマーカーとして発現するとともに、内胚葉マーカー遺伝子 *fkh*、*hkb*、*srp*

と中胚葉マーカー遺伝子 *twi*、*sna* との両方を発現していた。先行研究によると、内胚葉細胞でも *twi*、*sna* の発現は見られ(Alberga et al., 1991; Reuter et al., 1993; Thisse et al., 1988)、*wntD* は中胚葉への分化を抑制することが知られている(Ganguly et al., 2005; Rahimi et al., 2016)。また、“endoderm_antMG_wntD”は *sna*、*twi* の他の中胚葉マーカー遺伝子を発現していなかった。これらの理由から、このサブクラスタは暫定的に内胚葉に分類した。

もう1つは“ectoderm_PS14/hindgut”である (図 17C)。このサブクラスタは PS14 の外胚葉と後腸の中間的な細胞であると考えられた。なぜなら、このサブクラスタは PS14 のマーカー遺伝子である *Abdominal B (Abd-B)* と後腸マーカー遺伝子である *disconnected (disco)*、*wg* と *brachyenteron (byn)* を発現していたためである。このサブクラスタに対応する細胞が実際の胚に存在することを確認するため、*Abd-B* と *byn* の fluorescence *in situ* hybridization (FISH) による多重染色を行ったところ、将来の上皮細胞と後腸になる領域の間に *Abd-B* と *byn* の共発現細胞が存在した (図 24)。このことは、中間的な状態を取る細胞が scRNA-seq におけるダブレット等ではなく、原腸胚における将来の表皮細胞と後腸の間の領域に実際に存在していることを示す。

2.5 細胞膜関連遺伝子の発現は三胚葉の情報を強く反映する

サブクラスタリングによって細胞を領域ごとに分類できたことは、トランスクリプトームの細胞間差異が領域間の違いを強く反映していることを示す。このトランスクリプトームの細胞間差異にどのような遺伝子が寄与しているかを調べるため、細胞間分散の大きい遺伝子 (highly variable gene、HVG) の上位 3,000 遺伝子について GO term enrichment 解析を行った。濃縮していた GO タームには “nucleus” に加え、“plasma membrane” など細胞膜に関連するタームが多く含まれていた (図 25)。このことは、トランスクリプトームの細胞間差異に、転写因子に加え、細胞膜関連遺伝子が強く寄与していることを示唆する。そこで、Gene annotation list for *Drosophila* (GLAD) データベース (Hu et al., 2015) の情報をもとに 3,000 HVGs を 4 つのカテゴリに分類した。転写因子 (TF)、細胞膜関連遺伝子 (PM-related)、そ

のどちらでもないが GLAD には登録されている遺伝子 (Others)、そして、GLAD に登録されていない遺伝子 (Not in GLAD) である (図 26A)。この分類を用いた HVGs の内訳を見ると、転写因子と細胞膜関連遺伝子の濃縮が見られ、特に、分散の大きい上位 1,500 位までに転写因子が濃縮していた (図 26B)。この結果は、GO term enrichment 解析と同様に、転写因子と細胞膜関連遺伝子のトランスクリプトームへの寄与が大きいことを示す。

次に、転写因子の濃縮している上位 1,500 HVGs が細胞を分類する情報をどれだけ有しているかを調べるため、上位 1,500 HVGs を用いて、生殖細胞系列である極細胞 (pole_cells) を除く 76 サブクラスタについて階層的クラスタリングを行ったところ、サブクラスタは三胚葉に分類された (図 27)。これにより、原腸胚の時点でのトランスクリプトームの差異が三胚葉を区別するのに十分であることが明らかになった。

さらに、遺伝子はその種類ごとにどのような情報を持っているかを調べるため、特定のカテゴリの遺伝子のみを用いたサブクラスタの階層的クラスタリングを行った。1,500HVGs に含まれる 258 個の転写因子全てを用いた場合、大半のサブクラスタは三胚葉に分類される一方で、PS14 および後腸に相当するサブクラスタ ("ectoderm_PS14_ventral"、"mesoderm_PS14"など) が胚葉をまたいで一つのクラスタを形成した (図 28)。このことは、転写因子の発現パターンは細胞の分化状態のみでなく、位置の情報を含んでいることを示唆する。

対して、転写因子と同数の 258 個の細胞膜関連遺伝子を分散の大きいものから選択して用いた場合、階層的クラスタリングはサブクラスタを三胚葉に分類した (図 29)。また、転写因子でも細胞膜関連遺伝子でもない遺伝子群 258 個を用いた場合でも、サブクラスタは三胚葉に分類された (図 30)。これらの結果は、原腸胚の時点での転写因子の発現パターンは個々の遺伝子についての事前知識なしに将来の細胞運命を判別するのに不十分である一方で、その他のエフェクター遺伝子の発現パターンは細胞の分化状態をより強く反映していることを示す。

2.6 ストライプ状遺伝子発現のトランスクリプトームレベルでの再構成

序論でも述べたように、ショウジョウバエの初期発生ではパラセグメントという 14 の繰り返し単位が形成される。繰り返し単位内部の細胞列はそれぞれ異なった組み合わせのペアルール遺伝子とセグメントポラリティー遺伝子の発現を持ち、原腸形成開始時には 4 細胞列×2パラセグメントからなる 8細胞列の繰り返し構造が形成される(Clark and Akam, 2016) (図 31)。

ペアルール遺伝子の発現を乱すと、胚帯伸長が阻害されることが知られている(Irvine and Wieschaus, 1994)。ペアルール遺伝子は転写因子をコードすることから、ペアルール遺伝子は下流の遺伝子発現の制御を介してパラセグメント内の細胞間での遺伝子発現の違いを確立し、細胞の振る舞いを制御すると考えられる。実際、エフェクター遺伝子として、*18 wheeler* (*18w*、*Toll-2*)、*Toll-6*、*Tollo* (*Toll-8*)の三種類の Toll レセプターファミリーおよび *trn* が胚帯伸長を制御していることが知られている。これらの遺伝子の発現に隣接細胞間で差があることでその細胞境界にミオシンが集積し、細胞境界を縮小させ、細胞のインターカレーションを引き起こすと考えられている(Paré et al., 2019, 2014)。Tetley らはイメージング観察によってパラセグメント内の細胞列のうち 3 番目と 4 番目の細胞列が区別しにくいことを指摘している(Tetley et al., 2016)。このことから遺伝子発現においても 3 番目と 4 番目の細胞列の間では遺伝子発現の差が小さいことが予想される。さらに、実際の胚の 1 パラセグメントの長さは平均して 4 細胞列分に満たないため、“super-boundary” (“skipped boundary”)と呼ばれる、本来は隣接しない細胞列同士が接する細胞境界が生じる。この細胞境界ではより強く細胞境界が収縮することが、適切な組織変形に必要であるというモデルが示されている(Tetley et al., 2016)。前述の遺伝子発現の差異が細胞境界の収縮に必要であるというモデルをふまえると、super-boundary では通常の細胞境界より遺伝子発現の差異が大きいと考えられる。

以上のようなモデルが提案されているものの、胚帯伸長を制御するエフェクター遺伝子として Toll レセプターファミリー (*18w*、*Toll-6*、*Tollo*) と *trn* 以外の因子が存在するかは不

明である。さらに、パラセグメント内での細胞列間の遺伝子発現の差異はゲノムワイドかつ定量的には明らかになっていない。

そこで、パラセグメント内の細胞列ごとの遺伝子発現をゲノムワイドかつ定量的に明らかにするため、体幹部側面の外胚葉細胞に対して追加のクラスタリング解析を行い、細胞をパラセグメント内の各細胞列に分類した。まず、ストライプ状の遺伝子発現パターンを示す領域である体幹部側面の外胚葉細胞 1,717 細胞を抽出した (図 32 ステップ 1)。次に、これらの細胞を、ペアルール遺伝子 6 つとセグメントポラリティー遺伝子 3 つをランドマーク遺伝子 (図 31、stripe landmark genes) として用いて、パラセグメント内の 4 細胞列に分類した (図 32 ステップ 2)。さらに、7 本のストライプ状の発現を示す 2 つのランドマーク遺伝子 (図 31、even/odd landmark genes) を用いて、ストライプごとに細胞を偶数番目・奇数番目のパラセグメントに分類した (図 32 ステップ 3)。これらのプロセスにより、1,717 細胞を 8 つの細胞列に分類することができた (図 33)。

各ストライプについて遺伝子の平均発現量を計算することで、ストライプ状遺伝子発現の繰り返し単位を再構成したところ、ランドマーク遺伝子の発現は既知の遺伝子発現パターンと合致していた (図 34B)。さらに、ランドマーク遺伝子として使用していない遺伝子の再構成した遺伝子発現パターンも、既知の発現パターンを再現していた (図 34A、B)。特に、胚帯伸長の制御因子である *18w*、*Toll-6*、*Tollo* の再構成結果は、勾配をもつストライプ状の発現を示し (図 34C)、先行研究(Paré et al., 2014)と合致した。この結果は、ストライプ状遺伝子発現の繰り返し単位を 1 細胞列レベルで再構成できていることを示す。

次に、細胞列間の遺伝子発現の差異を明らかにするため、隣接細胞間 (細胞境界、boundary) で遺伝子発現変動解析を行った。1.75 倍以上の発現の違いと family-wise error rate (FWER) 閾値 0.01 の条件のもと、各細胞境界で 10-26 の遺伝子が発現に差のある遺伝子 (differentially expressed gene、DEG) として検出された (図 35A)。DEGs が最も少なかったのは各パラセグメントの 3 番目と 4 番目の間の細胞境界であった。このことは前述の先行研究による観察結果と合致する。また、パラセグメント内の各細胞列間では、少なくとも一つの Toll レセプターファミリー (*18w*、*Toll-6*、*Tollo*) が DEG として検出された (図 34C)。

この結果は細胞列間での Toll レセプターファミリーの発現の差異が胚帯伸長を制御するというモデルと整合的である。さらに、HVGs の場合と同様に、少なくとも一つの細胞境界で DEG として検出された遺伝子の内訳を GLAD カテゴリで見ると、その大半は転写因子または細胞膜関連遺伝子で占められていた (図 35B)。このことは転写因子に加え、細胞膜関連遺伝子がパラセグメント内の細胞列間のトランスクリプトームの差異に強く寄与していることを示す。加えて、隣接細胞間だけでなく、1 細胞列飛ばした細胞列間 (super-boundary) での遺伝子発現変動解析も実行した。各 super-boundary において 25-42 の DEGs が検出された (図 35A)。通常の細胞境界よりも DEGs の数が多いことは super-boundary において遺伝子発現の差異が大きいというモデルと合致する。

最後に、胚帯伸長を制御する未知のエフェクター遺伝子の候補を探索するため、個別の DEG に着目した。すると、Toll レセプターファミリーや *trn* に加えて、よく知られた膜貫通タンパク質をコードする遺伝子である *commissureless (comm)*、*comm2*、*Semaphorin 5c (Sema5c)* が DEGs に含まれていた (図 35C)。これらの遺伝子は細胞間の相互作用に寄与し、細胞のインターカレーションの制御因子となる可能性がある。

2.7 *bcd*-RNAi によって胚前方の細胞はトランスクリプトームレベルで後方化する

序論でも述べたように、前方の位置情報を与えるモルフォゲンをコードする遺伝子である *bcd* を欠失した胚では胚の前方の構造が欠失し、後方の構造に置き換わる。また、前方の細胞の遺伝子発現が後方の遺伝子発現に転換する (Petkova et al., 2019; Staller et al., 2015)。しかし、この現象は少数の遺伝子の発現にのみ着目して研究されたため、遺伝子発現がトランスクリプトームレベルで後方化しているかは明らかではない。また、*bcd* 機能阻害胚においては、胚前方の細胞の遺伝子発現の履歴は後方領域とは異なっており (Staller et al., 2015)、遺伝子発現の履歴の違いが最終的なトランスクリプトームに影響を与えうる。このため、前方細胞の後方細胞への転換をトランスクリプトームレベルで見ると、前方細胞と

後方細胞の特徴が混在した細胞や、野生型原腸胚には全く存在しない特徴を持つ細胞が存在しうる。

そこで、位置情報の異常がトランスクリプトームをどのように変化させるかを明らかにするため、RNAiによる *bcd* 機能阻害胚（以下、*bcd*-RNAi 胚）に対する scRNA-seq を行い（この結果を以下、*bcd*-RNAi データと呼ぶ）、コントロールと比較した。なお、*bcd*-RNAi 胚に対する scRNA-seq ではトリプシン処理による細胞の解離を用いたため、コントロールには同じトリプシン処理した細胞のデータである Set 2 を用いた。

まず、*bcd*-RNAi の胚前方細胞に特徴的な遺伝子発現への影響を調べるため、先行研究 (Finklstien and Perrimon, 1990; Jack and McGinnis, 1990)において *bcd* 欠失胚で発現が消失することが示されている *Dfd* および *oc* の遺伝子発現を調べた。すると、これらの遺伝子の発現は *bcd*-RNAi データにおいてほぼ消失していた (図 36)。この結果は胚前方に特徴的な遺伝子発現が消失していることを示唆する。

次に、遺伝子発現情報をもとに後方化した前方の細胞を識別できるかを明らかにするため、*bcd*-RNAi データについて、コントロールデータと同様にサブクラスタリングを行った。その結果、胚前方に相当するクラスタ（前方内胚葉、頭部および PS1-2 の外胚葉、および前方の中胚葉）が検出できず、また、全てのクラスタはコントロール胚に存在する細胞集団としてアノテーションすることが可能であった (図 37)。さらに、コントロールデータと *bcd*-RNAi データをバッチコレクションなしに結合して、Seurat によるクラスタリングを行った (図 38A)。その結果、胚前方に相当するクラスタすなわち前方内胚葉・中胚葉 (Anterior_midgut/mesoderm)、頭部外胚葉 (Head_ectoderm) および PS1-2 の外胚葉 (Ectoderm_PS1-2) はコントロールデータ由来の細胞のみからなっていた。また、*bcd*-RNAi データのみからなるクラスタは検出できなかった。これらの結果は、後方化した胚前方の細胞を本来の後方の細胞から識別することができないことを意味する。加えて、後方の細胞に相当するクラスタに割り当てられた *bcd*-RNAi 胚の細胞の全体に対する比率は、コントロール胚のそれよりも有意に大きく、およそ二倍であった (図 38B、表 8)。このことは後方化した前方の細胞が後方の細胞のクラスタに混ざっていることを示唆する。以上の

結果は、*bcd*-RNAi によって胚前方の細胞がトランスクリプトームレベルで後方化したことを支持する。

クラスタ情報だけでなく、個別の遺伝子発現も考慮して解析するため、各サブクラスタについて、*bcd*-RNA データとコントロールデータの間で遺伝子発現変動解析を行った (図 39)。その結果、161 遺伝子が *bcd*-RNAi データでコントロールと比較して高く発現する遺伝子 (高発現 DEG) として検出され、113 遺伝子が *bcd*-RNAi データでコントロールと比較して低く発現する遺伝子 (低発現 DEG) として検出された。もし、これらの遺伝子が *bcd*-RNAi の影響による DEGs であれば、本来の前方の細胞を含む後部領域特異的に検出されると考えられる。しかし、検出された遺伝子のうち、94 個の高発現 DEGs と 64 個の低発現 DEGs が二つ以上のクラスタで検出された DEG であり、さらにそれらのうち、68 個の高発現 DEGs と 37 個の低発現 DEGs は体幹部と後部の両方の領域にまたがって検出された。すなわち、多くの DEGs が領域特異性を示さなかった。このことは、この遺伝子発現変動解析の結果が *bcd*-RNAi による前方細胞の後方化よりもむしろ、遺伝的背景を反映している可能性が高いことを示唆する。このことは前方細胞がトランスクリプトームレベルで後方化していることを支持する結果である。

さらに、後方化した胚前方の細胞のわずかな特徴を捉えるため、Set2 と Set3 の双方で、胚前方のクラスタでのみマーカー遺伝子として検出される遺伝子をリスト化し、これと高発現 DEGs と共通する遺伝子を探索した。この基準に該当する遺伝子は *Distal-less (Dll)* の 1 つのみであった。*Dll* はコントロール胚において、*bcd*-RNAi 胚では欠失する領域である PS1-2 の背側外胚葉と羊漿膜で発現し、体幹部の羊漿膜では発現は見られなかった。一方、*bcd*-RNAi データにおいては *Dll* の発現は体幹部の羊漿膜に見られた (図 40)。このことから、*bcd*-RNAi データにおいては *Dll* の発現を示す細胞は PS1-2 の羊漿膜細胞から後方化した細胞であると考えられる。しかし、この *Dll* の発現は羊漿膜に限られ、背側外胚葉へは広がっていなかった。この結果は前方細胞がほぼ完全にトランスクリプトームレベルで後方の細胞に転換していることを支持する。

2.8 遺伝子発現のゲノムワイドな空間再構成

序論で述べたように、ショウジョウバエ原腸胚における遺伝子発現空間再構成の既存のデータベース(Karaiskos et al., 2017)は一部の遺伝子について不正確であった。この原因として空間再構成手法の性能が十分でない可能性と、空間再構成に用いるデータセットのクオリティーが不十分である可能性が考えられた。空間再構成手法の改善によって、より正確な空間再構成が可能であることは共同研究(Okochi et al., 2021)によって示した。これに加えて、空間再構成に用いるデータセットそのものを改善することで、さらに正確な空間再構成が可能であるかもしれない。そこで、本研究では、Set 3 に対して、共同研究によって開発した遺伝子発現の空間再構成手法 Perler (Okochi et al., 2021) (図 41) を適用し、ゲノムワイドな遺伝子発現の空間パターンデータベースを作成した。なお、レファレンスとなる *in situ hybridization* データベースには BDTNP データベース(Alberga et al., 1991; Fowlkes et al., 2008; Luengo Hendriks et al., 2006)を用いた。

まず、その結果を、Perler を NK-data に適用した場合と比較した。定量的には第一に、1 遺伝子抜き交差検証 (leave-one-gene-out cross-validation、LOOCV) のスコアは Set 3 を使用した場合 (相関係数の中央値 0.66) のほうが、NK-data を使用した場合 (相関係数の中央値 0.61) よりも有意に高かった (図 42A)。第二に、再構成した結果において元の scRNA-seq における遺伝子発現の相関関係がどの程度維持されているかを比較したところ、Set 3 を用いた場合のほうが NK-data を用いた場合よりも相関関係を有意に強く保存した (図 42B、「材料と手法」も参照)。第三に、Set 3 を用いた場合は元の scRNA-seq の発現スケールを維持していたが、NK-data を用いた場合はスケールを維持しなかった (図 43)。例えば、Set 3 を用いた場合、NK-data よりもバックグラウンドシグナルが小さかった。これらの結果は Set 3 を用いた空間再構成のほうが、NK-data を用いた場合よりも正確で、生物学的な解釈が容易であることを示す。

Set 3 データを用いた場合、NK-data を用いた場合と比較して定性的にも再構成結果の改善が見られた。例えば、Set 3 を用いた場合、NK-data よりもセグメントポラリティー遺伝

子 (*wg*) の 14 本ストライプがより明確であった (図 44)。NK-data を用いた場合はストライプの間における発現量がバックグラウンドよりも高い値を示すのに対し、Set 3 を用いた場合はストライプの間における発現量がバックグラウンドと同程度であった。さらに、*in situ hybridization* によって羊漿膜でしか発現しないことがわかっている遺伝子 *C15* の発現が、NK-data を用いた再構成では後部内胚葉まで伸びていたのに対し、Set 3 を用いた再構成では *in situ hybridization* の結果を再現していた (図 45)。なお、遺伝子 *egr* は *in situ hybridization* によって羊漿膜から後部内胚葉にかけて発現の見られる遺伝子であり、この遺伝子の再構成は Set 3 を用いた場合、NK-data を用いた場合いずれの場合でも *in situ hybridization* を再現した (図 45)。

近年、Perler 以外にも遺伝子発現の空間再構成手法が提案されている。そのうちの一つの *NovoSpaRc* (Moriel et al., 2021) であり、この手法は Perler と異なるアプローチである輸送最適化に基づいていること、また、細胞間の物理的な距離を考慮に入れることが特徴である。より正確な空間再構成を作成するため、*NovoSpaRc* を Set 3 に対して適用し、その結果を Perler の結果と比較した。第一に、両者の結果は強い相関を示した (図 46A)。第二に、LOOCV のスコアに大きな違いは見られなかった (図 46B)。第三に、Perler による再構成のほうが *NovoSpaRc* による再構成よりも有意に遺伝子間の相関関係を保存した (図 46C)。定性的な違いとして、*NovoSpaRc* による空間再構成は、Perler を用いた場合よりも滑らかな発現パターンを予測する傾向にあった。例えば、胚の腹側、中胚葉で特異的に発現する *twi* の発現を見ると、Perler による再構成では腹側領域内ではばらついた発現パターンが再構成されるのに対して、*NovoSpaRc* による再構成では腹側領域に均一な発現パターンが再構成された (図 47)。この違いは、*NovoSpaRc* が細胞間の物理的な距離を考慮することに起因すると考えられる。これらの結果は、Perler と *NovoSpaRc* は同程度の性能を示すものの、その再構成結果には質的な違いがあり、応用時にはそれを考慮する必要があることを示す。

以上のように、Set 3 を用いた Perler と *NovoSpaRc* は正確性の高い遺伝子発現空間再構成を行うことが可能である。特にこれらの再構成結果は前後軸に沿ったパターンについて高い性能を示した。しかし、背腹軸に沿ったパターンの再構成結果には不十分な点が見られ

た。例えば、元の scRNA-seq では *brk* 発現細胞において *vnd* と *ind* の発現は互いに排他的であるのに対して、空間再構成では Perler、NovoSpaRc どちらによる再構成の場合でも、*vnd* と *ind* の発現が重なった (図 48、図 49)。このことはレファレンスに用いた BDTNP データベースだけでは十分な背腹軸の位置情報が得られないことを示唆する。

第3章 考察

3.1 ショウジョウバエ原腸胚1細胞トランスクリプトームアトラス

本研究の目的はトランスクリプトームの細胞間での差異を明らかにすることであった。本研究ではまず、ショウジョウバエ原腸胚に対して scRNA-seq を適用し、細胞レベルのトランスクリプトーム情報を取得した。

本研究で得られたデータは最終的に 6,118 細胞を含んだ (Set 3)。この細胞数は先行研究 (Karaiskos et al., 2017) の 1,237 細胞と比較して 5 倍多く、ショウジョウバエ原腸胚 1 つが含む細胞数とほぼ同じである。ショウジョウバエ原腸胚が左右対称であることもふまえると、本研究で得られたデータはショウジョウバエ原腸胚の持つ細胞の種類の数のおよそ倍の細胞を含むことになり、ショウジョウバエ原腸胚の細胞をほぼ網羅していると考えられる。

このデータに対し、本研究ではまず、空間に対応する 77 のサブクラス情報を細胞に割り当てることで、高い空間解像度を付与した。これにより、胚の領域間での遺伝子発現の差異を詳細に比較することが可能となった。さらに、ショウジョウバエ原腸胚の持つ遺伝子発現の繰り返し単位を再構成し、1細胞列レベルでの遺伝子発現の比較を可能にした。最後に、遺伝子発現の空間パターンをゲノムワイドに再構成した。これらのデータは可視化ツールとともに公開しており (<https://github.com/TKondolab/flygastrula2>)、今後の発生生物学における重要なリソースとなることが期待される。

3.2 トリプシンによる細胞解離処理の影響

細胞の解離は scRNA-seq において不可欠なプロセスであり、また、人工的な遺伝子発現の変動を最小限にすることは scRNA-seq において重要である。本研究では常温でのトリプシン処理によって解離した細胞と低温下での CAP 処理によって解離した細胞の両方に scRNA-seq を適用し、トリプシン処理した細胞においてのみ Notch ターゲット遺伝子である *E(spl)-C* 遺伝子の発現が上昇することを見出した。この結果は、ショウジョウバエ原腸胚においてトリプシン処理が Notch ターゲット遺伝子を活性化することを示唆する。これと同

様の現象は哺乳類の細胞でも報告されている(Liu et al., 2014)。この現象の詳細なメカニズムは明らかでないが、考えられるメカニズムとして、トリプシンが直接 Notch を活性化する可能性がある。また、直接的な活性化ではなく、トリプシンが細胞表面のタンパク質を分解することで、Notch の cis 抑制(Del Álamo et al., 2011)を解除する可能性も考えられる。

本研究では酵素処理による細胞解離過程における遺伝子発現の変動の抑制に CAP を用いたが、他にも考えられる手段がある。1 つはアクチノマイシン D のような転写阻害薬を用いる方法である(Wu et al., 2017)。この方法は CAP を用いた低温での処理では組織を解離できないような場合に有効であると考えられる。ただし、この手法では転写を止めることができるが、mRNA の分解を止めることはできない。もう 1 つの手段は 1 細胞核 RNA-seq (snRNA-seq)を用いる方法である。この手法の利点は、複雑な組織から核のみを抽出することが細胞を解離するよりも容易であること、および、組織を瞬間冷凍することで遺伝子発現の変動を抑制できることである。また、核のみを用いるため、細胞化を生じる前のショウジョウバエ胚のような対象にも適用可能である(Albright et al., 2022)。しかし、snRNA-seq は検出できる転写産物数と遺伝子数が scRNA-seq よりも少ない傾向にある(Bakken et al., 2018; Basile et al., 2021)。本研究では、ショウジョウバエ原腸胚が細胞を解離しやすい単純な構造であることから、よりシーケンス深度の大きいトランスクリプトームデータを取得することを優先し、snRNA-seq でなく scRNA-seq を用いた。

3.3 細胞分化の中間的な発現を示す細胞

サブクラスタリングにおいて、PS14 外胚葉と後腸の中間的な遺伝子発現を示すサブクラスタ“ectoderm_PS14/hindgut”が検出された。このサブクラスタに属する細胞は PS14 外胚葉と後腸の分化における中間的な状態（両方に分化可能な状態）であると考えられる。同様の細胞は他の種における scRNA-seq 解析でも検出されており(Briggs et al., 2018; Farrell et al., 2018; Packer et al., 2019)、中間状態は細胞分化に広く見られる現象であることが示唆される。このような中間状態は一時的なものであり、やがていずれかの組織の細胞に分化すると考

えられるが、細胞の分化がいずれの方向へ進むかを制御するメカニズムは不明である。考えられるメカニズムとして、細胞分化と並行して起こる形態形成が細胞の分化を調整することが考えられる。実際、ショウジョウバエ胚における気管の陥入が、気管細胞（陥入した細胞）と表皮細胞（陥入しなかった細胞）の分化を調整する例がある(Kondo and Hayashi, 2019)。今回検出された中間状態”ectoderm_PS14/hindgut”において、後腸は後部内胚葉と共に胚の内部へ陥入する部位であり、もう一方の PS14 外胚葉は陥入せずに胚の表面に留まる領域である。胚内部への陥入の有無が”ectoderm_PS14/hindgut”の細胞の最終的な分化先を決定する可能性がある。これを検証するためには、形態形成と遺伝子発現の双方の時系列データが必要である。

3.4 *bcd*-RNAi 胚における遺伝子発現の収斂

限られた転写因子の組み合わせから多様な細胞種を生み出すメカニズムとして、遺伝子発現の履歴がのちの遺伝子発現を決定するというモデルが考案されている(Letsou and Cai, 2016)。序論でも述べたが、ショウジョウバエの *bcd* 機能阻害胚においては胚前方細胞の後方が知られていたものの、ギャップ遺伝子 *hb* を始め、遺伝子発現の履歴が胚の前方と後方で異なる(Staller et al., 2015)。この履歴の違いが最終的なトランスクリプトームに影響する可能性が残っていた。本研究では、位置情報を与えるモルフォゲンをコードする *bcd* 遺伝子をノックダウンした場合に、胚前方の細胞がトランスクリプトームレベルで後方化することを明らかにした。この結果は、*bcd*-RNAi 胚における胚前方と後方の間の遺伝子発現の履歴の違いは、少なくとも原腸形成開始時のトランスクリプトームには影響しないことを示している。また、転写状態が野生型に存在する状態へ収斂する現象はゼブラフィッシュの scRNA-seq でも報告されている(Farrell et al., 2018)。

この背景には、位置情報が乱された場合に転写状態を野生型に存在する状態に収斂させる遺伝子制御ネットワークがあると考えられる。本研究の結果は、そうしたネットワークが特定の遺伝子に関してだけでなく、ゲノム全体にわたって存在していることを示唆し

ている。この遺伝子制御ネットワークがどのように機能しているかを調べるためには、遺伝子制御ネットワークを推定し、その活性を解析に含める(Aibar et al., 2017)ことが有効であると考えられる。しかしながら、遺伝子制御ネットワークを推定するにはトランスクリプトーム情報だけでは不十分であることが指摘されている(Chen and Mar, 2018)。遺伝子制御ネットワークの正確な推定には ATAC-seq によるクロマチンアクセシビリティの情報を組み合わせるなどの方法(Janssens et al., 2022) が考えられるが、現時点では、ショウジョウバエ原腸胚、特に *bcd*-RNAi 胚に関する 1 細胞 ATAC-seq データは存在せず、これは今後の課題である。

3.5 細胞膜関連遺伝子のトランスクリプトームへの寄与

本研究において、ショウジョウバエ原腸胚では、細胞膜関連遺伝子がトランスクリプトームの差異に強く寄与していることを明らかにした。また、細胞膜関連遺伝子の発現パターンが転写因子の発現パターンよりも細胞の分化状態を反映していることも明らかにした。さらに、ペアルール遺伝子によって特徴づけられるパラセグメント内の細胞列間で発現の異なる遺伝子にも細胞膜関連遺伝子が多く含まれることも明らかにした。これらの結果は細胞膜関連遺伝子が他の細胞質で機能する遺伝子よりも、転写状態の細胞の挙動への変換において強い寄与を持っていることを示唆する。このことは、膜貫通タンパク質に制御される細胞間シグナルが細胞と組織の振る舞いを制御するという先行研究に合致する(Manning et al., 2013; Paré et al., 2019, 2014)。HVGs のうち、転写因子でも細胞膜関連遺伝子でもない遺伝子の発現パターンも細胞をある程度、細胞を三胚葉に分類できたことをふまえると、これらの遺伝子は細胞の基本的な性質を決定し、細胞の振る舞いとして可能な幅を規定している可能性がある。そのうえで、細胞間で細胞膜関連遺伝子の発現が異なることで、局所的な細胞間相互作用が調整され、それに基づいて、可能な細胞の振る舞いの中から領域特異的な細胞の振る舞いが選択されることが考えられる。

ペアルール遺伝子によって特徴づけられるパラセグメント内の細胞列間において、胚帯

伸長を制御する既知の膜貫通タンパク質をコードする遺伝子(Paré et al., 2019, 2014; Schaerlinger et al., 2007)である Toll レセプターファミリー (*18w*、*Toll-6*、*Tollo*) と *trn*、*5-HT2A* が DEGs として検出されたが、これ以外に *comm*、*comm2*、*Sema5c* のように神経細胞の軸索誘導に関係することが知られている膜貫通タンパク質遺伝子(Keleman et al., 2005, 2002; Yazdani and Terman, 2006)も検出された。先行研究により、神経ネットワーク形成に関係する多くの遺伝子が上皮の形態形成や恒常性に関与することがわかってきている(Cammarota et al., 2020; Hinck, 2004; Vaughn and Igaki, 2016; Yoo et al., 2016)。また、*Sema5c* は上皮性の濾胞細胞の形態形成を制御することが報告されている(Stedden et al., 2019)。したがって、これらの DEGs は胚帯伸長を制御する遺伝子の有力な候補である。また、本研究で検出された DEGs の数は限られる。解析における閾値の問題は残るものの、この限られた膜貫通タンパク質が上皮形態形成、特に胚帯伸長を制御する因子として十分であることが期待される。

3.6 位置情報から細胞トランスクリプトームへの非線形変換

胚発生においては、連続的な位置情報から局所的な転写因子の発現が生じ、その下流で細胞の分化状態が決定すると考えられている(Briscoe and Small, 2015)。しかし、この非線形な変換は限られた遺伝子について研究され、ゲノムワイドな視点からの研究は不十分であった。本研究では、ショウジョウバエ原腸胚においては、転写因子の発現パターンよりも、細胞膜関連遺伝子のようなエフェクター遺伝子の発現パターンのほうが細胞の分化状態をより反映していることが明らかになった。加えて、転写因子の発現パターンは位置についての情報を反映する傾向にあることも明らかになった。これらの結果は転写因子の発現パターンとエフェクター遺伝子の発現パターンの間の非線形な変換を示すものである。細胞の機能や挙動を実際に担うのがエフェクター遺伝子であることをふまえると、この非線形な変換は、位置情報が転写因子のレイヤーを介して細胞の分化状態へと非線形に変換される過程の一部を、限られた遺伝子に関してのみではなく、トランスクリプトームレベルで

捉えたものと解釈できる。

また、このような非線形な変換の背景として、遺伝子制御ネットワークにおいて、特定の転写因子の組み合わせが強い寄与を持つマスター制御因子として選択され、その下流で細胞の分化に対応した遺伝子発現が形成される可能性が考えられる。他の可能性として、一部の転写因子の mRNA 量がタンパク質量と相関していない可能性が考えられる。また、mRNA 量がタンパク質量と相関していても、転写因子の活性を反映していない可能性も考えられる。例えば、一部の転写因子が翻訳後修飾によってより強い活性を示す場合である。

3.7 今後の展望

本研究では、ショウジョウバエ原腸胚をモデルに、1細胞トランスクリプトームの細胞間差異を解析し、トランスクリプトームの細胞間差異に細胞膜関連遺伝子の寄与が大きいことを明らかにした。さらに、前述したように、細胞膜関連遺伝子の寄与の大きさは、細胞膜関連遺伝子の発現が異なることで、局所的な細胞間相互作用が調整され、その結果として細胞と組織の振る舞いが決定される可能性を示唆する。したがって、細胞膜関連遺伝子の発現パターンと細胞の振る舞いを結びつけるルールを明らかにすることで、どのように細胞が位置情報からトランスクリプトームを介して固有の振る舞いを決定するかを理解できると考えられる。このルールを明らかにするためのアプローチとして、細胞の振る舞いを領域ごとに観察・定量し、対応する領域の遺伝子発現と統合して解析することが必要である。この解析には序論で述べた定性的な細胞の振る舞いの情報のほか、原腸胚全体の細胞形状を網羅的に観察したデータ(Stern et al., 2022)を用いることが可能である。また、近接する細胞間の相互作用を考慮に入れるため、タンパク質間のインタラクトーム(Ding et al., 2020)と各細胞の遺伝子発現を組み合わせることも有効であると考えられる。いずれのアプローチにおいても、遺伝子発現の空間パターンがわかっていることが重要である。本研究において、ショウジョウバエ原腸胚のトランスクリプトーム情報を取得し、その空間パターンを明らかにしたことは、発生生物学分野の今後の研究の進展における重要な基

盤となることが期待される。

第4章 材料と手法

4.1 使用したショウジョウバエ系統

使用した全てのショウジョウバエ系統はトウモロコシ粉、コーングリッツ、ドライイースト、グルコース、寒天、プロピオン酸、ブチル p-ヒドロキシベンゾールを含む実験室標準餌を用いて飼育された。コントロール系統として、Set 1 および Set 2 には *y w* 系統、Set 3 には *R14E10-GAL4[attP2] UAS-mCD8.chRFP (III)* 系統、SABER-FISH には *w* 系統を用いた。母性 *bcd* ノックダウンは以下のようにして行った。まず、*UAS-bcd RNAi* (TRiP.GL00407) 系統のメスを *matalpha4-GAL-VP16[67]* と *matalpha4-GAL-VP16[15]* を持つオスと交配させ、*matalpha4-GAL-VP16[67]/+; matalpha4-GAL-VP16[15]/ UAS-bcd RNAi* (TRiP.GL00407) のメスを得た。このメスに *UAS-bcd RNAi* (TRiP.GL00407) 系統のオスを交配して得た胚を scRNA-seq に用いた。

4.2 scRNA-seq の実行

4.2.1 細胞懸濁液の準備

ピンセット、筆、ナイロンメッシュを含む、使用した全ての道具は RNase quiet (Nacalai) で処理したあと、RNase-free の水で洗浄した。ハエに 20-30 分間産卵させたあと、胚を 25°C で 90 分放置した。そのあと、胚を水で半分の濃度に薄めたハイター（花王株式会社）で処理してコリオンを除去し、RNase-free PBS で洗浄した。蛍光実体顕微鏡（Nikon SMZ18）下で胚を観察し、原腸形成開始直後のステージ 6-7 の胚を拾い上げた。拾い上げた胚は氷冷した 10 μ L の解離用バッファー（1x RNase-free PBS、5%トレハロース）入りの 1.5mL チューブ（Watson、PROKEEP protein low binding tube）の中に速やかに移した。150-300 の胚を回収したあと、ピペットチップ（Axygen, Maxymum Recovery 200 μ L Universal Fit Tip with Filter）の先とチューブの底の角ですり潰すようにして胚のビテリン膜を壊した。潰された胚を氷冷した 500 μ L の解離用バッファーで懸濁したあと、800 rcf、4°C で 2 分間遠心して沈殿させた。上清を除去し、再度、氷冷した 500 μ L の解離用バッファーで懸濁し、800 rcf、4°C で 2

分間遠心した。

トリプシン処理の場合は、ペレットを 1x trypsin-EDTA (Sigma, T3924)で懸濁した。25°Cで 10 分間静置したあと、500 μ L の氷冷した停止バッファー (1x PBS, 5% trehalose, 0.375 % BSA (WAKO, 012-23881), 0.1 mg/mL trypsin inhibitor (Sigma, T6522)) を加えて反応を停止させた。氷冷した 500 μ L の洗浄バッファー1 (1x PBS, 5% trehalose, 0.375 % BSA (WAKO, 012-23881)) で 2 回洗浄したあと、ペレットを氷冷した 200 μ L のローディングバッファー(1x PBS, 5% trehalose, 0.5 mg/mL ULTRAPURE BSA (Thermo Fisher, AM2616))で懸濁した。

CAP 処理の場合はペレットを 500 μ L の CAP 溶液 (5 mg/mL *Bacillus licheniformis* protease (Sigma P5380), 5% trehalose, in 1x PBS)で懸濁したあと、6°Cで 30 分間静置した。そのあと 500 μ L の洗浄バッファー2 (1x PBS, 5% trehalose, 0.5 mg/mL ULTRAPURE BSA (Thermo Fisher, AM2616)) を加えた。氷冷した 500 μ L の洗浄バッファー2 で 4 回洗浄したあと、ペレットを氷冷した 200 μ L のローディングバッファーで懸濁した。

トリプシン処理、CAP 処理いずれの場合も、細胞懸濁液をセルストレイナー (FLOWMI Cell Strainers for 1000 μ L Pipette Tip, 40 μ m Porosity) に通したあと、1mL の CellCover (Anacyte Laboratories) を加えて 25°Cで 1 時間、4°Cで一晩静置して細胞を固定した。固定した細胞は氷冷した 500 μ L のローディングバッファーで洗浄したあと、氷冷した 100 μ L のローディングバッファーで再懸濁した。細胞の濃度を血球計算盤で測定したあと、細胞濃度を Fluidigm C1HT の場合はおよそ 200 または 300 cells/ μ L に、10x genomics Chromium の場合はおよそ 300 cells/ μ L に調整した。

4.2.2 C1HT による scRNA-seq

Fluidigm C1 と C1 Single-Cell mRNA Seq HT IFC を用いた scRNA-seq ライブラリの作製は開発元のプロトコールに以下の修正を加えて行った。細胞を溶解する前に、AxioCam 105 color (Zeiss) と電動ステージを備えた Axio Observer.Z.1 (Zeiss) を用いて、全 800 ヶ所のキャプチャーサイトを自動的に画像化した。逆転写反応用に 8 塩基の UMI を挿入したカスタムプライマーを使用した。プライマー配列は表 9 に記載した。また、ERCC spike-in mix

(Thermo Fisher, 4456740)を Lysis Mix に加えた。さらに、ライブラリー増幅ステップで使用するプライマーをプロトコール指定の 10 倍低い濃度にして使用した。ライブラリー増幅のための PCR サイクルは 12 回とした。

Bioanalyzer と qPCR を用いたクオリティーチェックと定量化の後、ライブラリーは NextSeq 500 (Illumina)、75 cycles high-output kit v2 (Read1 : 15 サイクル、Read2 : 69 サイクル、Index1 : 8 サイクル、合計 92 サイクル) で配列決定した。

4.2.3 10x Chromium による scRNA-seq

10x Chromium with the Chromium Next GEM Single Cell 3' Reagent Kits v3.1 を用いた scRNA-seq ライブラリーの作製は開発元のプロトコールに沿って行った。cDNA 増幅のための PCR サイクルは 11 回、ライブラリー増幅のための PCR サイクルは 12 回とした。

Bioanalyzer と qPCR を用いたクオリティーチェックと定量化の後、NextSeq 500 (Read1 : 28 サイクル、Read2 : 56 サイクル)、NovaSeq 6000 (Illumina) (Read1 : 28 または 151 サイクル、Read2 : 91、98 または 151 サイクル)、または HiSeq X (Illumina) を用いて配列決定した。

4.3 scRNA-seq データの解析

4.3.1 10x Chromium データの前処理および Seurat クラスタリング

UMI と細胞バーコードを含む Read1 は、fastx_trimmer (FASTX-toolkit, version 0.0.14, http://hannonlab.cshl.edu/fastx_toolkit) を用いて 28 塩基長にトリミングした。fastp (version 0.20.1、Chen et al., 2018)を用いてアダプターのトリミングとクオリティーフィルタリングを行った。このとき、オプションは NextSeq 500 データについては -q 20 --cut_tail -l 28 --max_len1 28 --max_len2 55 --trim_poly_g --trim_poly_x、10x NovaSeq データについては -q 20 --cut_tail -l 28 --max_len1 28 --max_len2 97 --trim_poly_g --trim_poly_x、HiSeq X データについては -q 20 --cut_tail -l 28 --max_len1 28 --max_len2 97 --trim_poly_x を用いた。トリミングした

リードを STARsolo (version 2.7.7a, Kaminow et al., 2021) を用いて *Drosophila melanogaster* のゲノム配列 (BDGP6.22.98) にマッピングし、UMI をカウントした。この過程において、STARsolo (version 2.7.7a) は複数の遺伝子に割り当てられたリードをカウントすることに対応していなかったため、gtf アノテーションファイルに修正を加え、複数の遺伝子についてゲノム上に同じ向きで領域が重なっている場合にそれらの遺伝子を統合し、一つの遺伝子として取り扱うようにした (表 10)。細胞のフィルタリングでは、STARsolo のフィルタリング出力における細胞ごとの総 UMI 数の中央値を算出し、その二倍よりも大きい総 UMI 数を持つ細胞をダブレットの可能性のある細胞として除去した。また、検出遺伝子数、リボソーム RNA 遺伝子の UMI の割合、ミトコンドリアゲノム遺伝子の UMI の割合のいずれかが平均値から標準偏差の 2.5 倍以内の範囲から外れている細胞を低クオリティ細胞として除去した。

残った細胞を Seurat (version 3.2.3, Stuart et al., 2019) にロードして、SCTransform 関数を利用してノーマライズした。ノーマライズではオプションとして、vars.to.regress = c("percent.mt", "percent.rRNA") を指定した。ここで、"percent.mt" と "percent.rRNA" はそれぞれミトコンドリアゲノム由来の転写産物、核ゲノム rRNA の全転写産物に対する割合を細胞ごとに計算したメタデータのラベルである。SCTransform が返す residual_variance (ピアソン残差) が 1 であれば、その遺伝子の持つ分散はノイズであると判断される。(residual_variance - 1) を "biological variance" と定義し、その大きさが上位の遺伝子から累積和を計算した。その値が飽和する 3,000 遺伝子を Highly variable genes (HVGs) とし、以降の解析に用いた。Seurat の RunPCA 関数を用いて PCA にかけて、30 次元を以降の解析に用いた。なお、HVGs の数と PCA の次元数はこれ以降、特に言及しない限りは同じ値を用いた。UMAP による次元削減は RunUMAP 関数 (オプション: dims = 1:30, n.neighbors = 20L) を用いて行った。Seurat によるクラスタリングは FindNeighbors 関数 と FindClusters 関数を用いて行った。リボソームタンパク質遺伝子の高発現を示し、胚における位置に対応するマーカーを発現していない Seurat クラスターは、低クオリティ細胞としてフィルタリングされた。各クラスターは、FindAllMarkers 関数で同定されたマーカー遺伝子に基づいて手動でアノ

テーションを行った。

4.3.2 C1HT データの前処理および Seurat クラスタリング

fastp を用いてアダプターのトリミングとクオリティーフィルタリングを行った。このとき、オプションは-q 20 --cut_tail -l 14 --max_len1 14 --max_len2 68 --trim_poly_g --trim_poly_x を用いた。STARsolo を用いてトリミングしたリードを *Drosophila melanogaster* のゲノム配列 (BDGP6.22.98) にマッピングし、UMI をカウントした。このとき、gtf ファイルには上述の修正を加えたものを用いた。キャプチャーサイトの画像からシングレットであると判断したサイトのデータのみを選択し、Seurat にロードした。それぞれのバッチごとに、検出遺伝子数、リボソーム RNA 遺伝子の UMI の割合、ミトコンドリアゲノム遺伝子の UMI の割合、ERCC spike-in の UMI の割合のいずれかが平均値から標準偏差の 2.5 倍以内の範囲から外れている細胞を低クオリティー細胞として除去した。全てのバッチをマージしたあと、SCTransform でノーマライズした。オプションとして、vars.to.regress = c("percent.mt", "percent.rRNA", "percent.ERCC")を指定した。ここで、“percent.ERCC”は ERCC spike-in の UMI の全転写産物に対する割合を細胞ごとに計算したメタデータのラベルである。次元削減とクラスタリングは 10x Chromium データと同様に行った。

4.3.3 NK-data の前処理および Seurat クラスタリング

Karaiskos et al., 2017 で報告された fastq ファイル (GSM2494783 – GSM2494789) は SRA データベースから入手した。fastp を用いてアダプターのトリミングとクオリティーフィルタリングを行った。このとき、オプションは-q 20 --cut_tail -l 20 --max_len1 20 --max_len2 64 --trim_poly_g --trim_poly_x を用いた。また、Cutadapt (version 3.4, Martin, 2011) をオプション -m 20:20 -G AAGCAGTGGTATCAACGCAGAGTACATGGG 付きで用いて Read2 から SMART アダプターを除去した。トリミングしたリードを、*D. melanogaster* (BDGP6.22.98) のゲノム配列またはこの *D. melanogaster* レファレンスと *D. virilis* (GCF_003285735.1_DvirRS2) レファレンスを結合したレファレンスに、STARsolo を用いてマッピングした。この際、上述の

修正した `gtf` ファイルを用いた。その後、Drop-seq tools (version 2.5.1, Macosko et al., 2015) の `TagReadWithGeneFunction`, `DetectBeadSubstitutionErrors`, `DetectBeadSynthesisErrors` (オプション `--PRIMER_SEQUENCE AAGCAGTGGTATCAACGCAGTAC`) および `DigitalExpression` (オプション `--NUM_CORE_BARCODES 5000`) で BAM ファイルから UMI カウントテーブルを生成した。*D. melanogaster* と *D. virilis* の細胞を含むデータについては、*D. melanogaster* にマッピングされた UMI の総数が 90%以上の細胞のみを *D. melanogaster* の細胞とみなし、下流の解析のために抽出した。元論文と同様に、総 UMI 数が 12,500 以上の細胞のみを高クオリティ細胞として残した。また、それぞれのバッチごとに、検出遺伝子数、リボソーム RNA 遺伝子の UMI の割合、ミトコンドリアゲノム遺伝子の UMI の割合のいずれかが平均値から標準偏差の 2.5 倍以内の範囲から外れている細胞を低クオリティ細胞として除去した。全てのバッチをマージしたあと、`SCTransform` 関数を利用してノーマライズした。ノーマライズではオプションとして、`vars.to.regress = c("percent.mt", "percent.rRNA")` を指定した。次元削減とクラスタリング、プロットは 10x Chromium データと同様に行った。

4.3.4 トリプシンデータからの高 *tsr* 発現細胞の除去

Set 2 と *bcd*-RNAi データから高 *tsr* 発現細胞を除去するために、R の `corr` パッケージ (version 0.4.3) の `correlate` 関数で 2,000HVGs のすべてのペア間の相関係数を計算した。次に、*tsr* と相関を持つ遺伝子を抽出し、R の `mclust` パッケージ (version 5.4.7, <https://cran.r-project.org/package=mclust>) の `Mclust` 関数にオプション `pca=30`, `G=2`, `modelName="VVV"` をつけて主成分分析 (principal component analysis, PCA) とクラスタリングを実行した。最後に、*tsr* の発現量が高いクラスターをストレス細胞としてフィルタリングした。

Set 1 については、極細胞を除去した後、`SCTransform` によるノーマライズ、`RunPCA` 関数による次元削減、`FindNeighbors` および `FindClusters` 関数 (`resolution=2.0`) による 30 次元でのクラスタリングを行った。検出されたクラスターのうち、*tsr* の高発現を示す 2 つのクラスターを除去した。

4.3.5 それぞれの scRNA-seq データのサブクラスタリング

それぞれのデータセットのサブクラスタリングについては、表2-表5、表7に示す単位ごとに PCA による次元削減と FindNeighbours 関数と FindClusters 関数を用いての教師なしグラフベースのクラスタリングを適用した。適用単位の細胞数が 500 以上の場合のみ、サブクラスタリング前に SCTransform 関数による再ノーマライズを行った。細胞数が 500 未満の場合は、SCTransform が適切なノーマライズを行えなかったため、再度のノーマライズを省略した。細胞数が 31 以上の場合は RunUMAP、FindNeighbours と FindClusters の各関数に対して 30 次元を使用し、30 以下の場合は細胞数-1 を次元数とした。FindClusters 関数の resolution パラメータは、文献の知識に基づいてアノテーションできないサブクラスタが出現しない最大値を採用した。例外として、体幹部側面の外胚葉の背腹軸に沿ったサブクラスタリングでは、クラスタ数7の k-means クラスタリングで行った。このとき、35の DV 遺伝子 (*Ama*, *Ance*, *Atx-1*, *bbg*, *brk*, *C15*, *CG13653*, *cic*, *cv-2*, *dap*, *Doc1*, *Doc2*, *Doc3*, *dpp*, *Dr*, *Dtg*, *Egfr*, *egr*, *emc*, *ind*, *mirr*, *peb*, *pnt*, *pnt*, *rho*, *sog*, *SoxN*, *srp*, *stg*, *tup*, *ush*, *vn*, *vnd*, *Z600*, *zen*) のみを用いた。各サブクラスタは、表2-表5、表7に示すマーカー遺伝子に基づいて手動でアノテーションした。サブクラスタリングの際、外胚葉系遺伝子と中胚葉系遺伝子の両方の発現を示す細胞は、ダブルットとして除去した。また、胚における空間に対応するマーカーを発現していない細胞も低クオリティー細胞として除去された。残りのシングルットデータセットの細胞数は、Set 3、Set 1、Set 2、NK-data についてそれぞれ 6,118、1018、4,855、1,476 であった。

4.3.6 Harmony による scRNA-seq データの統合

個別にサブクラスタリングした後の Set 3、Set 2、Set 1、NK-data を用いた。それぞれのデータセットは個別に SCTransform 関数で再ノーマライズしたあと、統合に用いる 3,000 遺伝子を Seurat の FindIntegrationFeatures 関数を用いて決定した。4つのデータをマージしたものを RunPCA 関数で PCA にかけて、30 次元を用いて Harmony (version 0.1.0, Korsunsky et al., 2019) でバッチコレクションを行った。バッチコレクション後の 30 次元を用いて、

RunUMAP 関数による次元削減と、FindNeighbors 関数、FindClusters 関数によるクラスタリングを行った。個別のデータセットと同様に、クラスタは手動でアノテーションを行った。

4.3.7 統合データセットのサブクラスタリング

まず、クラスタリングの対象となる単位 (表6) を元の4つのデータに再分割した。再分割した単位において細胞数が500以上の場合、SCTransform で再ノーマライズを行った。細胞数が500未満の場合、再ノーマライズの代わりに Seurat オブジェクト内の SCT assay の scale.data slot の値を Seurat の ScaleData 関数 (オプション do.scale=FALSE) で中心化した。再度の統合に用いる3,000遺伝子を FindIntegrationFeatures 関数で決定し、RunPCA 関数で次元削減を行った。30次元を用いて Harmony でバッチコレクションを実行した。さらに、RunUMAP 関数を適用した。クラスタリングとアノテーションは上述の個別のデータセットの場合と同様に行った。UMAP とクラスタリングには、バッチコレクション後のデータ30次元を用いた。

4.3.8 Gene Ontology term enrichment 解析

HVGs に関する Gene Ontology term enrichment 解析では、極細胞のサブクラスタ (pole_cells) を除去したあと、Seurat を用いて分散の大きい上位3,000遺伝子を抜き出した。この遺伝子リストを g:Profiler (<https://biit.cs.ut.ee/gprofiler>) を用いて、g:SCS アルゴリズムで解析した。FDRの閾値を0.01、term_sizeを4,000未満のものを有意に濃縮したGOタームとして検出した。Set2において高tsr発現細胞で高発現する遺伝子についての解析では、クラスタ1 (Trunk_mesoderm) と比較してクラスタ10 (Mesoderm_tsr-high) で高発現する遺伝子328個を FindMarkers 関数で検出した。この際、検出手法には MAST (Finak et al., 2015) を指定し、FDRの閾値を0.01、logFCの閾値を0.25とした。このリストを g:Profiler で解析し、有意に濃縮された Kyoto Encyclopedia of Genes and Genomes (KEGG) terms を FDR 閾値0.01で検出した。

4.3.9 GLAD を用いた階層的クラスタリング

各 GLAD カテゴリの遺伝子リストは <https://www.flyrnai.org/tools/glad/web/> からダウンロードした。"Transcription factor/DNA binding" カテゴリに属する遺伝子には他のカテゴリにも含まれているものがあるため、これらの重複を除外した修正データベースを作成した。"Trans-membrane proteins"、"Receptors"、"Secreted proteins" および "Matrisome" のリストを統合して、細胞膜関連遺伝子 (PM-related genes) とした。"Transcription factor/DNA binding" カテゴリにも細胞膜関連遺伝子にも含まれない遺伝子をその他の遺伝子 ("Others") として扱った。なお、細胞膜関連遺伝子のリストはこの時点で細胞膜以外の細胞小器官で機能する遺伝子を含むと考えられたため、GO ターム "intracellular membrane-bounded organelle" (GO: 0043231) およびその child term ("mitochondrion"、"Golgi apparatus"、"endoplasmic reticulum" などを含む) の付与されている遺伝子を除去し、その他の遺伝子 (Others) に移した。転写因子は上位 1,500 HVGs に集中していたため、以下の階層的クラスタリング解析では上位 1,500 遺伝子のみを用いた。特定の遺伝子セットのみを用いた解析では、最も遺伝子数の少ない TF に遺伝子数を合わせるため、全てのセットで分散の上位から 258 個の遺伝子のみを選択した。次元削減解析は、上述のように Seurat を使用して行った。UMAP プロット上での点の色は事前のアノテーションに従って決定した (表 2 の "Super Cluster")。階層的クラスタリングのため、極細胞 (pole_cells) を除く 76 サブクラスタについて各遺伝子の平均発現量を計算した。平均発現量には SCTransform でノーマライズしたカウント値の対数をとったものを使用した。計算した平均発現量を用いてサブクラスタの全ペアのユークリッド距離を R の dist 関数で算出し、これに基づいて hclust 関数を用いて平均連結法による階層的クラスタリングを実行した。ヒートマップおよび樹形図は gplot パッケージ (version 3.1.1, <https://CRAN.R-project.org/package=gplots>) の heatmap.2 関数を用いて作成した。

4.3.10 細胞列への割り当てと細胞列間の DEG 解析

まず表 2 に示した "Trunk ectoderm 2" の細胞を抽出した。各パラセグメントは前後軸に沿った 4 つのストライプで構成され、各ストライプは 1 細胞幅のカラムであり、異なる遺伝

子発現プロファイルを持つと考えられている(Clark and Akam, 2016; Tetley et al., 2016)。そこで、抽出した細胞に対して、パラセグメント内のストライプ位置を推測するために、9つのストライプランドマーク遺伝子(図31)を用いたk-meansクラスタリング(クラスタ数4)を実行した。そのあと、割り当てられたストライプごとに、even/oddランドマーク遺伝子(図31)の発現に基づき、偶数番目、奇数番目のいずれのパラセグメントに由来する細胞かを分類した。even/oddランドマーク遺伝子にはストライプ1と2には *trn* を、ストライプ3と4には *pxb* を用い、ランドマーク遺伝子および上位1,000 HVGsのうちランドマーク遺伝子と相関のある遺伝子を用いたk-meansクラスタリング(クラスタ数2)を実行した。K-meansクラスタリングの実行には、RのClusterRパッケージ(version 1.2.2, <https://CRAN.R-project.org/package=ClusterR>)のk-means関数を用いた。ランドマーク遺伝子とHVGsの相関の計算にはRのcorrパッケージ(version 0.4.3)のcorrelate関数を用いた。各隣接細胞境界またはsuper-boundaryにおけるDEGsの検出はMASTを用いたFindMarkers関数で同定し、FWER閾値0.01、FC閾値1.75でフィルタリングした。FC閾値は、FWER閾値でフィルタリングしたあと、隣接細胞境界間のDEGのFC分布をプロットし経験的に決定した。

4.3.1.1 Set 2 データと bcd-RNAi データのマージ

サブクラスタリング後のSet 2とbcd-RNAiデータについて、両者を個別にSCTransformで再ノーマライズし、Seuratのmerge関数でマージした。解析に用いる3,000遺伝子はSeuratのFindIntegrationFeatures関数を用いて決定した。バッチコレクションは行わなかった。RunPCAで次元削減し、30次元をRunUMAP関数による次元削減とFindNeighbors関数、FindClusters関数によるクラスタリングに用いた。各クラスタのアノテーションは個別のデータセットの場合と同様に行った。各Seuratクラスターについて、クラスターに割り当てられた細胞の比率がbcd-RNAiとコントロールデータセットで異なるかどうかを調べるために、フィッシャーの正確検定を用いた。検定にはRのfisher.exact関数を用いた。

4.3.1.2 Set 2 データと *bcd*-RNAi データの間の DEG 解析

bcd-RNAi データと Set 2 のサブクラスタを手動で対応づけた。ただし、Set2 と *bcd*-RNAi をマージした Seurat クラスタで”Dorsal_lateral_ectoderm_PS13”、”Amnioserosa_PS13-14”とアノテーションされた細胞については、サブクラスタ情報の代わりにこのクラスタ情報を使用した。各サブクラスタについて、*bcd*-RNAi データと Set 2 データの間で FindMarkers 関数 (method = ‘MAST’) を用いて DEGs を検出した。このとき、FWER の閾値は 0.01、|logFC| の閾値は 0.5 とした。DEGs のうち、*bcd*-RNAi データで Set 2 より発現の高いものを高発現 DEGs、低いものを低発現 DEGs とした。

Set 2 の各サブクラスタのマーカー遺伝子の検出には FindAllMarkers 関数 (method = ‘MAST’) を用いた。このとき、FWER の閾値は 0.01 とした。検出された DEGs のうち、*bcd*-RNAi データで見つからなかった胚前方のクラスタでのみ、DEGs として検出される遺伝子をリスト化した。このリストと高発現 DEGs の内訳を比較し、共通するものを検出した。

4.3.1.3 遺伝子発現の空間再構成

scRNA-seq データの前処理

Set 3 データについては、レファレンスである BDTNP データが極細胞を含まないため、サブクラスタ”pole_cells”に含まれる細胞 123 細胞を除去した。残りの 5,995 細胞について SCTransform による再ノーマライズを行い、Seurat の RunPCA 関数で次元削減を行った。

NK-data については、UMI カウントテーブルとして *Drosophila* Virtual Expression eXplorer (<https://shiny.mdc-berlin.de/DVEX/>)から取得したテーブルを用いた。このカウントテーブルを Seurat にロードし、SCTransform で再度ノーマライズした。なお、このデータには極細胞は含まれず、ミトコンドリアゲノム遺伝子および rRNA は省略されていた。

いずれのデータセットについても、空間再構成には log スケールのデータ (Seurat オブジェクトの”SCT” assay の”data” slot の値) を用いた。

ランドマーク遺伝子の選択

in situ hybridization レファレンスは主に BDTNP データベース (<http://bdtnp.lbl.gov>) から取得したもの (D_mel_wt_atlas_r2.vpc) を用い、一部、DVEX (<https://shiny.mdc-berlin.de/DVEX/>) から取得したもの(bdtnp.txt)を用いた。DVEX のレファレンスは BDTNP のレファレンスから分岐したものであり、3 遺伝子 (*bowl*, *ems*, *exex*) は DVEX レファレンスには含まれない。scRNA-seq データはいずれもステージ 6-7 の胚から得られたもので、*in situ* hybridization レファレンスデータはステージ 5 の胚で確立されたものである。レファレンスデータには、ステージ 5 からステージ 6-7 にかけて発現パターンが動的に変化する遺伝子もあった。そこで、2つの時点の間で発現パターンが大きく変化し、再構成を悪化させる可能性のある遺伝子はレファレンスから除外した。その結果、67 個の遺伝子が空間再構成のランドマークとして残った (表 11)。また、DVEX レファレンスの 3,039 個の細胞のうち、 $y < 0$ の 8 個の細胞が取り除かれた。

Perler による空間再構成

Python パッケージ Perler (version 0.1.0) は GitHub (<https://github.com/yasokochi/Perler>) から取得した。scRNA-seq データとレファレンスデータを PERLER オブジェクトにロードして、`em_algorithm` メソッドをオプション `optimize_pi = False` を用いて EM アルゴリズムを実行した。次に、scRNA-seq データ点とレファレンスデータ点の間の距離を、`calc_dist` メソッドをデフォルトパラメータで用いて計算した。ハイパーパラメータの最適化は、デフォルトのパラメータを用いた `loocv` メソッドと、パラメータ `grids = ((0,1), (0.01,1))` を用いた `gridsearch` メソッドにより行った。最後に、パラメータ `mirror = False`, `_3d = True`, `z_scored = False` を用いて、空間的遺伝子発現パターンを `spatial_reconstruction` メソッドにより再構成した。

NovoSpaRc による空間再構成

NovoSpaRc (version 0.4.3) による再構成は、主に <https://github.com/rajewsky-lab/novosparc>

にしたがって行った。scRNA-seq データとレファレンスデータを NovoSpaRc の Tissue オブジェクトにロードした。30 個の主成分に基づく `set_up_smooth_costs` メソッドと、レファレンスデータとデフォルトパラメータを用いた `setup_linear_cost` メソッドにより、最適輸送フレームワークのコスト行列を算出した。そして、パラメータ `alpha_linear=0.3`, `epsilon=5e-3` を用いた `reconstruction` メソッドにより、空間再構成を行った。なお、`alpha_linear` パラメータと使用する主成分の数は、後述の LOOCV スコアが最大となるようにグリッドサーチで決定した。

Leave-one-gene-out cross-validation (LOOCV)

レファレンスに含まれる 67 個のランドマーク遺伝子のそれぞれを正解の発現としてレファレンスから除去し、残りの 66 個の遺伝子をレファレンスとして Perler または NovoSpaRc による空間再構成を行った。そして、除去された遺伝子の再構成された発現パターンと正解の発現との間のピアソン相関を計算した。

NovoSpaRc で用いるハイパーパラメータの決定には、以下のスコアを用いた。

$$J = -\frac{1}{2} \sum_i^{67} \ln(1 - \rho_i^2)$$

ここで、 ρ_i は遺伝子 i についての正解の発現と再構成された遺伝子発現の間のピアソン相関係数を示す。このスコアが最大となるようにハイパーパラメータと使用する主成分の数を決定した。

遺伝子発現の相関の維持の比較

まず、Set3 と NK-data の両 scRNA-seq データセットの上位 500 HVGs に含まれる共通の HVGs 372 個を選択した。各 HVG について、元の scRNA-seq データにおけるその遺伝子と他の 371 個の HVGs の間のピアソン相関係数と、Set3 ベースまたは NK-data ベースの再構成におけるそれらの相関係数を算出した。そして、scRNA-seq の相関係数セットと再構成の相関係数セットの間のピアソン相関係数を、各 HVG について算出し、元の scRNA-seq デー

タと再構成の間の遺伝子-遺伝子相関構造の保存を示すスコアとした。Perler と NovoSpaRc の比較には、Set3 データの上位 500 個の HVG を使用した。

再構成パターンのプロット

プロットは matplotlib パッケージ (version 3.3.4, Hunter, 2007) の scatter 関数によって行った。再構成された遺伝子発現の値は log スケールから線型スケールに変換してプロットした。側面から見た図では、レファレンスに含まれる全ての細胞をプロットした。背側または腹側から見た図ではそれぞれ $z > 0$ 、 $z \leq 0$ の座標を持つ細胞のみをプロットし、x-z 平面でミラーリングした。全ての図で左が胚の前方であり、側面から見た図では上が背側となるようにした。

vnd および ind の密度プロット

scRNA-seq データのプロットには、サブクラスタリング情報に基づき、Set3 データから腹部または PS13 の中間外胚葉細胞 (intermediate_NE) または内側神経外胚葉細胞 (medial_NE) と正中線 (midline_cells) の細胞を抽出した。再構成データプロットでは、 $|x| < 50$ 、 $-55^\circ < \theta < 0^\circ$ の細胞を抽出した。 θ は、細胞を含む y-z 平面に平行な断面で胚の中心から細胞まで引いた線分と y 軸がなす角度で、背腹軸上での細胞の位置を表す (図 49A)。密度推定は、Scipy パッケージ (version 1.6.0, Virtanen et al., 2020) の stats モジュールの gaussian_kde クラスを使用して行った。推定結果は、matplotlib パッケージの pcolormesh 関数を用いてプロットした。

4.4 バルク RNA-seq

4.4.1 バルク RNA-seq の実行

胚からの total RNA 調製では、6-7 期胚 80 個を採取し、RNeasy Lipid Tissue Mini Kit (Qiagen) を用いて total RNA を精製した。解離した細胞から Total RNA を調製するために、ステージ

6-7の胚 200-300個を、上記のようにトリプシン-EDTA処理により単細胞懸濁液に解離させた。洗浄後、細胞を40 μ mのストレーナーに通し、ペレット化した。RNeasy Mini Kit (Qiagen)を用いて、約40,000個の細胞からtotal RNAを精製した。固定細胞からの全RNA調製のために、ステージ6-7の200-300個の胚をトリプシン-EDTA処理によって単細胞懸濁液に解離し、上記のようにCellCoverによって固定した。細胞は4°Cで1日保存し、ペレット化した。RNeasy Mini Kit (Qiagen)を用いて、約40,000個のペレット化した細胞からTotal RNAを精製した。

SMART-Seq v4 Ultra Low Input RNA Kit (Clontech)を用いて、各Total RNA 250 ngからcDNAを合成した。次に、Illumina Nextera XT DNA Library Preparation Kitを用いて0.0625 ngのcDNAからイルミナシーケンサー用のライブラリーを構築した。このライブラリーをIllumina NextSeq 500でシーケンスし、76塩基の長さのシングルエンドリードを取得した。各サンプルは重複して解析した。各ライブラリーについて、36,577,021~41,844,986のリードが配列決定された。

4.4.2 バルク RNA-seq データの解析

配列決定されたリードについて、Trim Galore (version 0.6.4, https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)と Cutadapt (オプション: --nextseq 20)を用いてクオリティーコントロールとトリミングを行った。それぞれのリードから76番目の塩基を除去した後、残ったリードを *Drosophila melanogaster* のゲノム配列 (BDGP6.22.98) に STAR を用いてマッピングした。このとき、gtf ファイルには上述の修正を加えたものを用いた。遺伝子発現は RSEM (version 1.3.3, Li and Dewey, 2011) を用いて計算し、DEG解析は edgeR (version 3.32.1, Robinson et al., 2010) を用いた。ミトコンドリアとリボソーム RNA の遺伝子を除去したあと、6サンプルすべてで CPM が 0.1 未満の低発現遺伝子もフィルターで除去した。ノーマライズは calcNormFactors を使用して行った。スピアマン相関係数は R の cor 関数を用いて計算した。DEG は edgeR パッケージの glmQLFit および glmQLFTest 関数を用いて、FDR 閾値 0.01、logFC 閾値 2 の条件で同定した。

4.5 SABER-FISH

SABER-FISH は Kishi et al., 2019 のプロトコルに修正を加えて実行した。OligoMiner (Beliveau et al., 2018)を用いて予め設計されたプローブ配列の候補の”Balance” list は <https://oligopaints.hms.harvard.edu/genome-files> からダウンロードした。各遺伝子に対して、全てのアイソフォームで共通するエクソンを標的とするプローブをリストからランダムに抽出した。SABER-FISH プローブは、IDT 社から購入した DNA オリゴ (表 12) を用いて PER 増幅により作成した。

胚を水で半分の濃度に薄めたハイター (花王株式会社) で 2 分間処理してコリオンを除去したあと、CaCl₂ 2mM を含む 4% PFA とヘプタンを 1:1 で混合した液で 20 分間固定した。固定した胚はメタノールとヘプタンを 1:1 で混合した液中で振ってビテリン膜を除去した。胚をメタノールで 2 回洗浄したあと、NA LoBind Tubes (Eppendorf)中に回収した。さらに、胚を 0.2% Tween-20 と 0.2% Triton X-100 を含む PBS で洗浄した (2分ずつ 2回)。続けて、PBSTw (0.2% Tween-20 を含む PBS) で 5分ずつ 3回洗浄した。液を PBSTw と Whyb バッファー (1% Tween-20 と 40% formamide を含む 2× SSC pH 7.0) の 1:1 混合液に置き換え、5分間静置した。ハイブリダイゼーションの前に、胚を Whyb 中で 43°C で 10 分間インキュベートした。そのあと、胚をそして、胚をあらかじめ温めた Hyb1 バッファー (2.5% dextran sulfate 500 kDa (FUJIFILM Wako, 193-09981)を含む Whyb) 中で 1 μg/100 μL のプローブとともに 43°C で 16-48 時間インキュベートした。ハイブリダイゼーション後、胚を 43°C で Whyb で洗浄し (素早く 1回、30分ずつ 3回)、さらに 2×SSCT (0.1% Tween-20 を含む 2×SSC) で洗浄した (5分ずつ 3回)。蛍光オリゴをハイブリダイゼーションさせる前に、チューブを室温に戻し、胚を PBSTw で洗浄した (5分ずつ 3回)。その後、チューブを 37°C に移し、チューブが温まったら PBSTw を取り除き、Alexa-fluor と結合した 0.2 μM DNA オリゴ (Thermo Fisher 社合成、表 12) を含む、あらかじめ温めた Hyb2 バッファー (2.5% dextran sulfate 500 kDa を含む PBSTw) と交換した。37°C の暗所で 20 分間インキュベートしたあと、37°C で胚を PBSTw で洗浄した (素早く 1回、5分ずつ 3回)。その後、胚を PBSTw 中 1 μg/mL

DAPI (DOJINDO) と共に暗所室温で 30 分間インキュベートし、PBSTw で洗浄した (15 分ずつ 3 回)。胚は SlowFade Diamond Antifade Mountant (Invitrogen) でマウントした。画像は、40 倍の水浸対物レンズを備えた Zeiss LSM800 (Objective LD LCI Plan-Apochromat 40x/1.2 Imm Corr DIC M27、Zeiss) を用いて撮影した。

4.6 プロット

データのプロットは他に記載のないかぎり、Seurat もしくは ggplot2 (version 3.3.3) を用いて行った。

図

図1 モルフォゲンによる細胞運命の決定

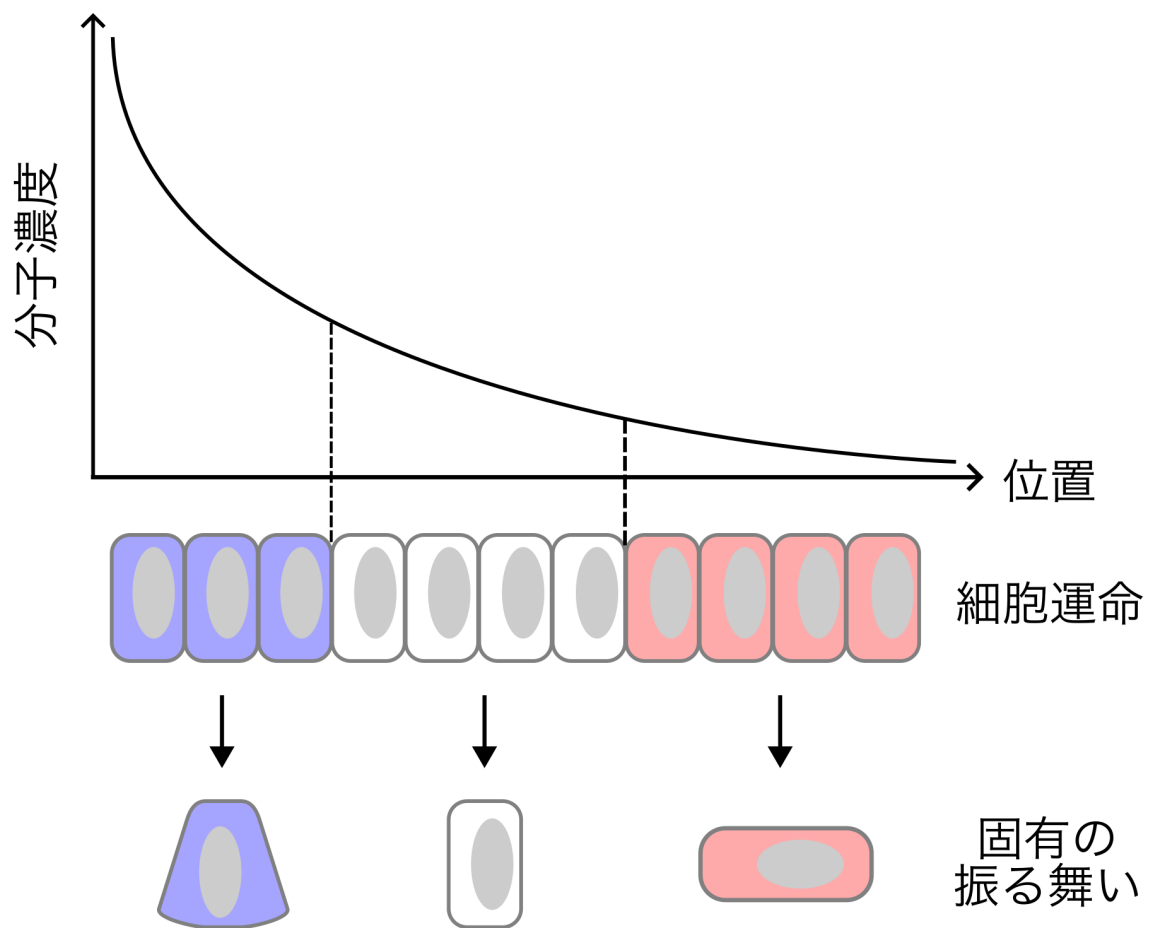


図1

モルフォゲンによる細胞の運命決定の模式図。細胞は分子濃度によって異なる細胞運命を獲得し、その下流で固有の振る舞いや機能を示す。

図2 ショウジョウバエ胚における三胚葉の出現

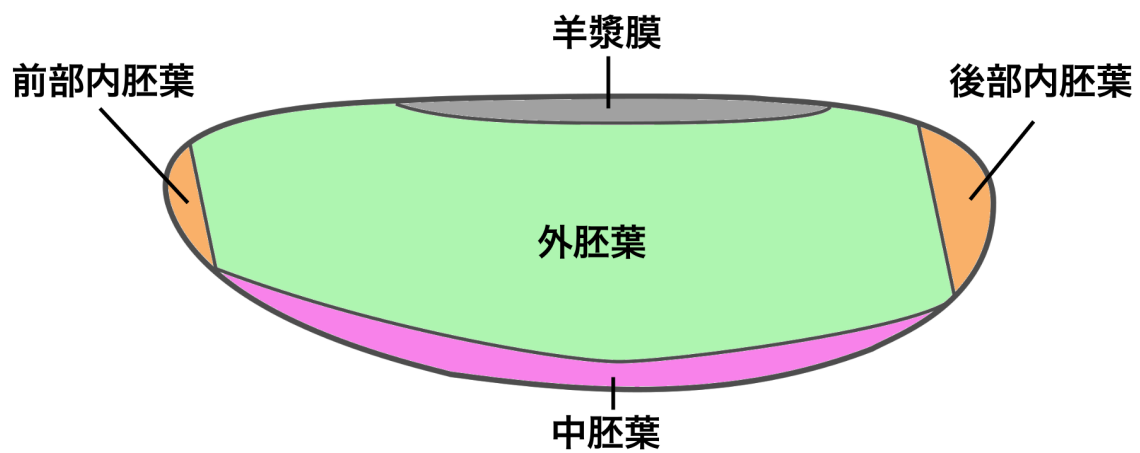


図2 ショウジョウバエ胚における三胚葉の出現。胚の腹側に中胚葉、側面に外胚葉が生じ、胚の前後の末端部には内胚葉が生じる。また、胚の背側には羊漿膜が生じる。

図3 ショウジョウバエ原腸胚で生じる多様な細胞の振舞い

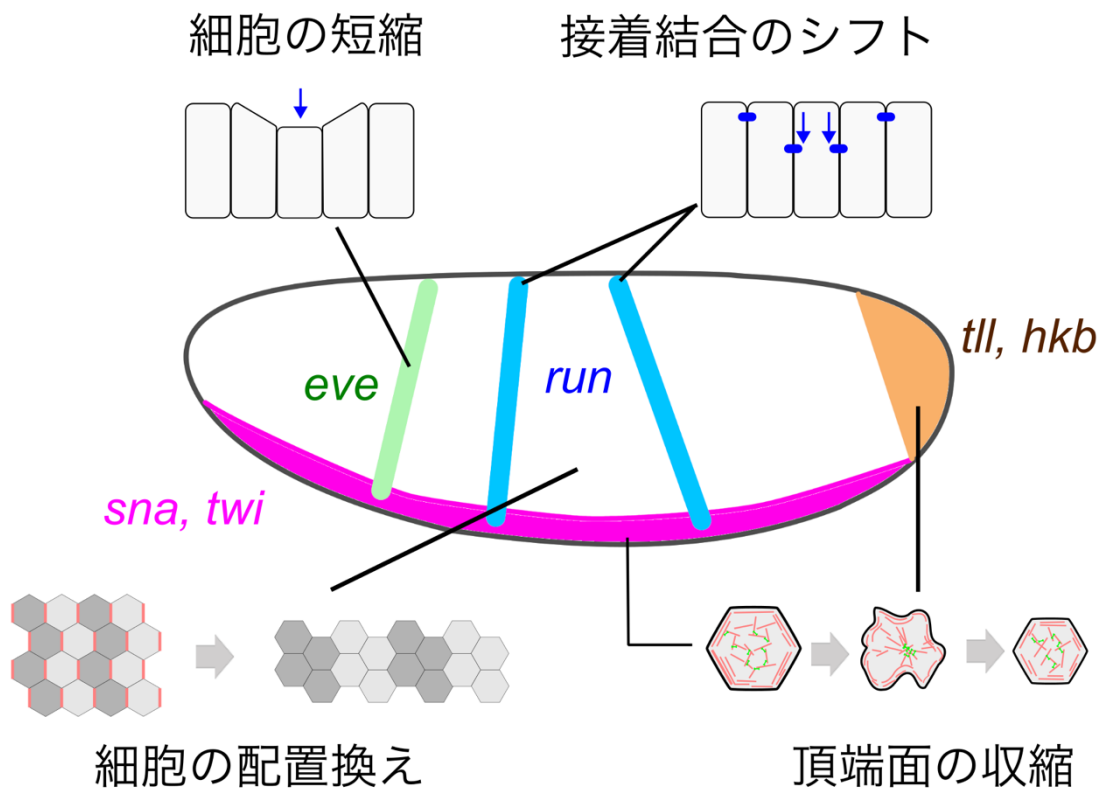


図3

ショウジョウバエ原腸胚で見られる多様な細胞の振る舞いを図示した。これらの振る舞いの制御に関与することが知られている遺伝子とその発現部位の一部もあわせて示してある。

図4 ショウジョウバエ胚における位置情報の読み出し

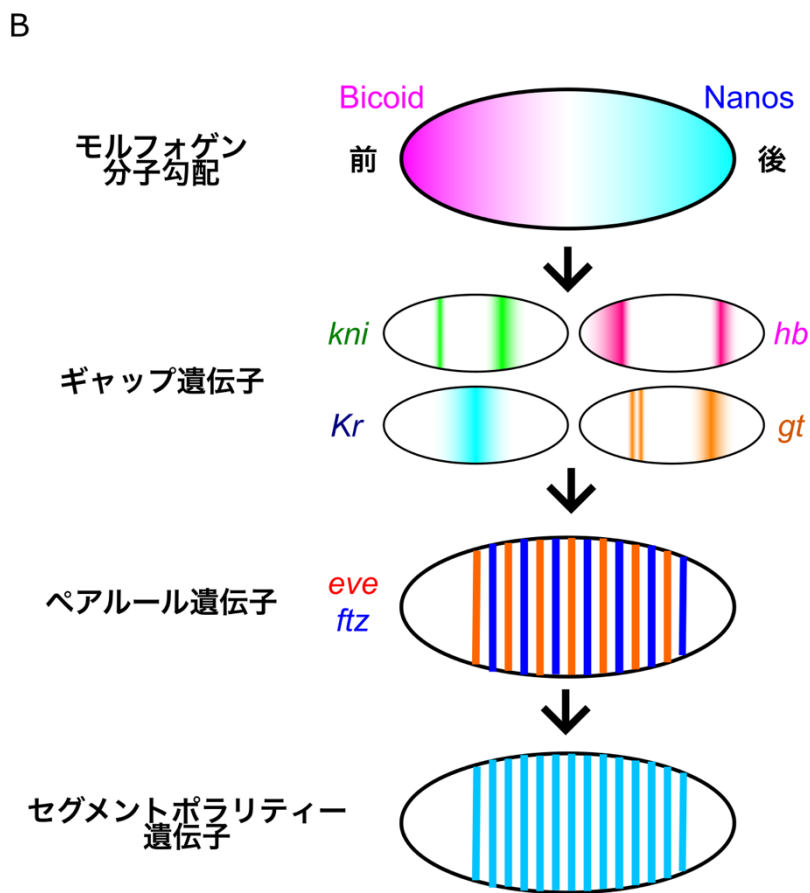
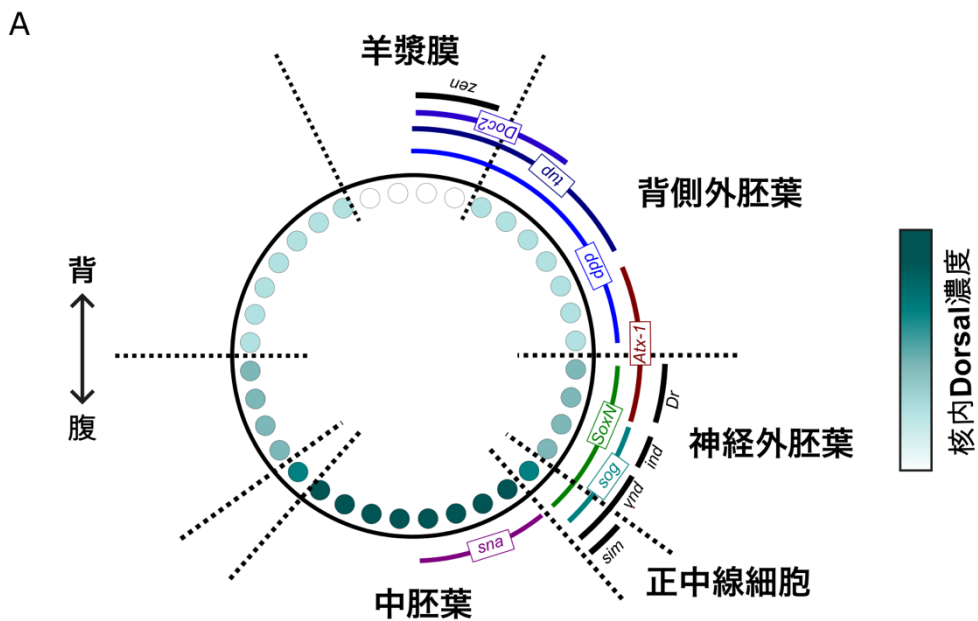


図 4

(A)Dorsal の核内濃度による背腹軸の細胞運命の決定の模式図。緑色の濃さが Dorsal 核内濃度を示す。図の上が背側、下が腹側。背腹軸に沿って異なる発現を示す遺伝子の発現範囲も合わせて示してある。(B) ショウジョウバエ胚の前後軸パターンニングの模式図。左が胚の前方、右が胚の後方である。

図5 パラセグメント内の細胞列は異なる遺伝子発現を持つ

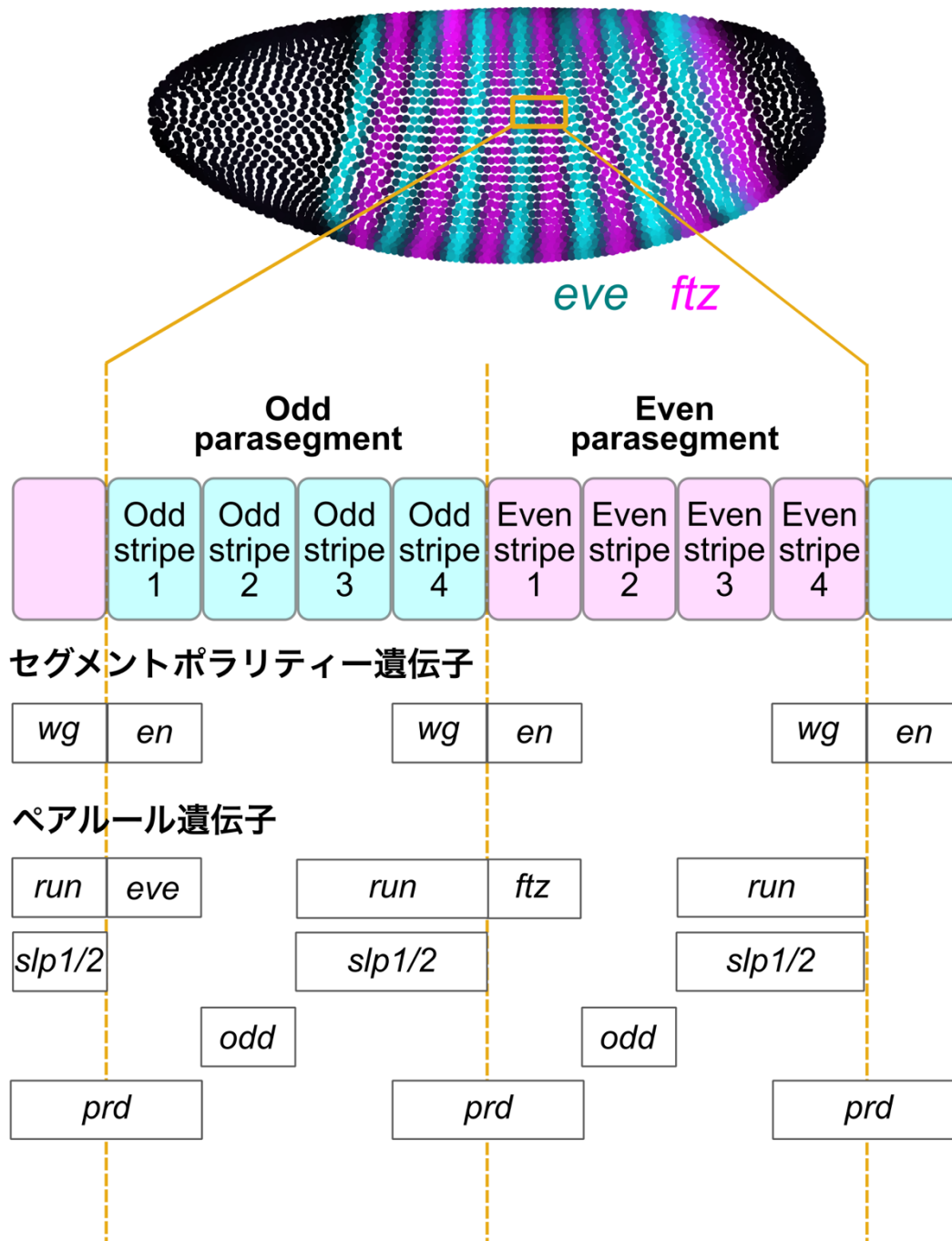
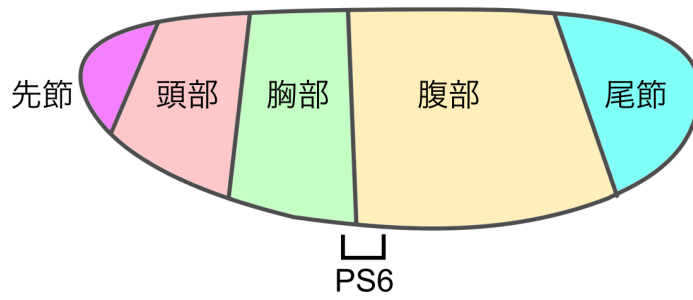


図5

(上) ショウジョウバエ胚ステージ5の終わりにおけるペアルール遺伝子 *eve* (青) と *ftz* (マゼンタ) の発現パターン。データは Berkeley Drosophila Transcription Network Project (<http://bdtnp.lbl.gov>) より取得した。(下) 原腸形成開始時のパラセグメント内の8細胞列におけるペアルール遺伝子とセグメントポラリティー遺伝子の発現パターン。Clark and Akam, 2016 を参考に作成した。

図 6 *bcd* 変異体の表現型

野生型



bcd 欠失胚

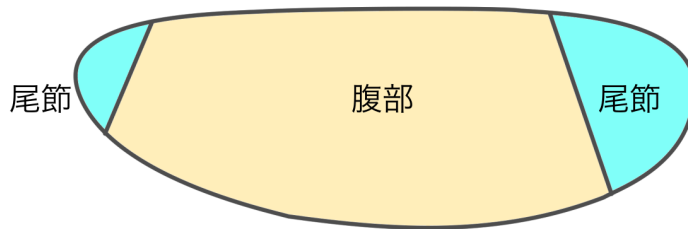


図 6

bcd 欠失胚の表現型の模式図。先節から胸部かけての構造が欠失し、胚の後部の構造と置き換わる。PS: パラセグメント

図7 陥入を引き起こす多様な細胞の挙動

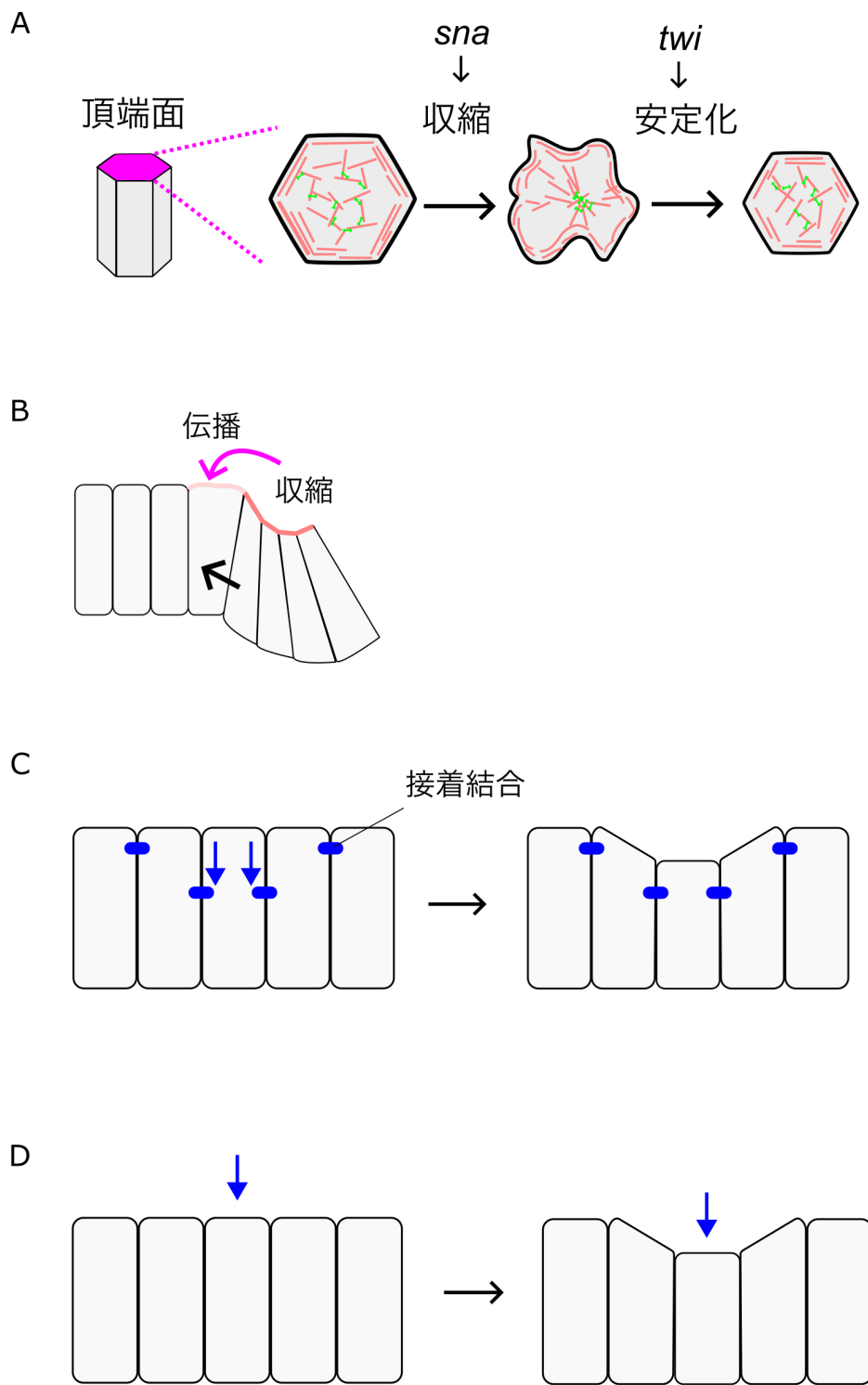


図7

(A) 頂端収縮のモデル図。ミオシン（緑）がアクチンネットワーク（赤）を収縮させ、その後安定化される。(B) 後部内胚葉陥入における力学的ストレスによる頂端収縮の伝播モデル (C) Dorsal fold の形成開始のモデル図。Initiator 細胞において接着結合が基底側へとシフトし、それによって陥入が開始される。(D) Cephalic furrow 形成開始のモデル図。最初に陥入する細胞が短縮することによって陥入が開始する。

図8 胚帯伸長を駆動する細胞のインターカレーション

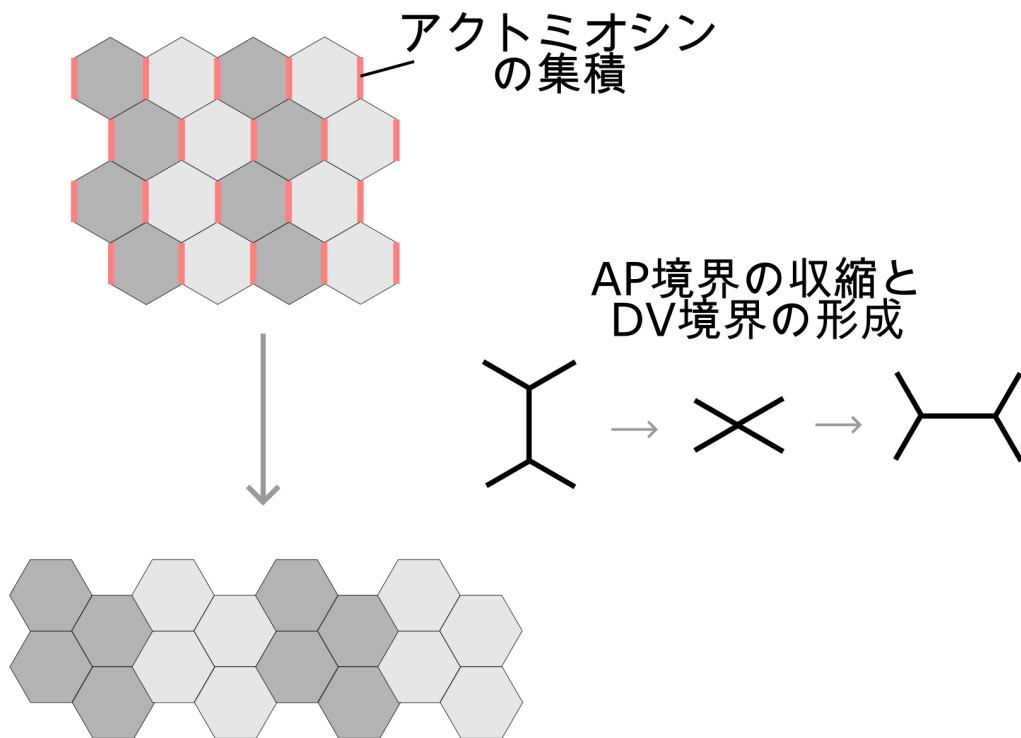


図8

ショウジョウバエ原腸胚で生じる組織の伸長のモデル図。細胞の配置が変化することによって伸長が生じる。その際、AP境界が短縮し、4細胞の接した状態を経て、DV境界が伸長する。

図 9 Karaiskos et al., 2017 による遺伝子発現の空間再構成

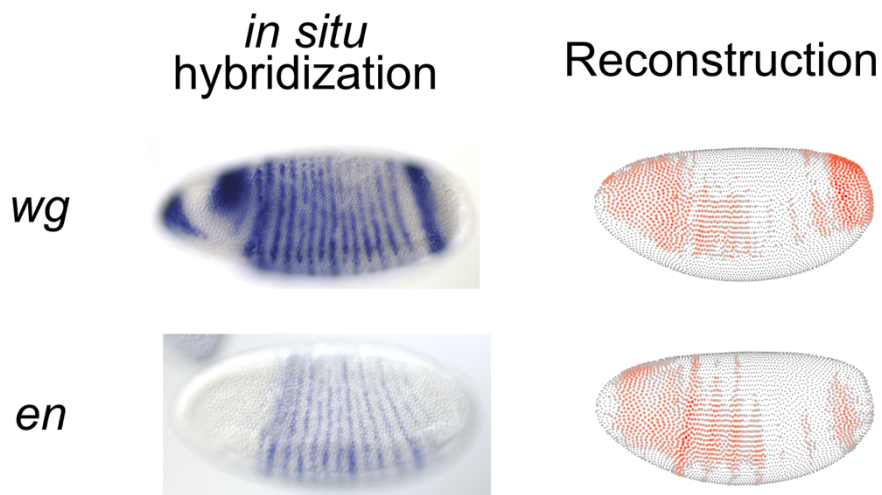


図 9

(左) *wg* および *en* の遺伝子発現パターン。画像は BDGP in situ データベース (<https://insitu.fruitfly.org/>)(Hammonds et al., 2013; Tomancak et al., 2007, 2002)から取得した。

(右) Karaiskos et al.による *wg* および *en* の遺伝子発現パターンの再構成結果。画像は DVEX (<https://shiny.mdc-berlin.de/DVEX/>) から取得した。ともに Threshold は 0.6 とした。

図 10 本実験で取得したデータの模式図

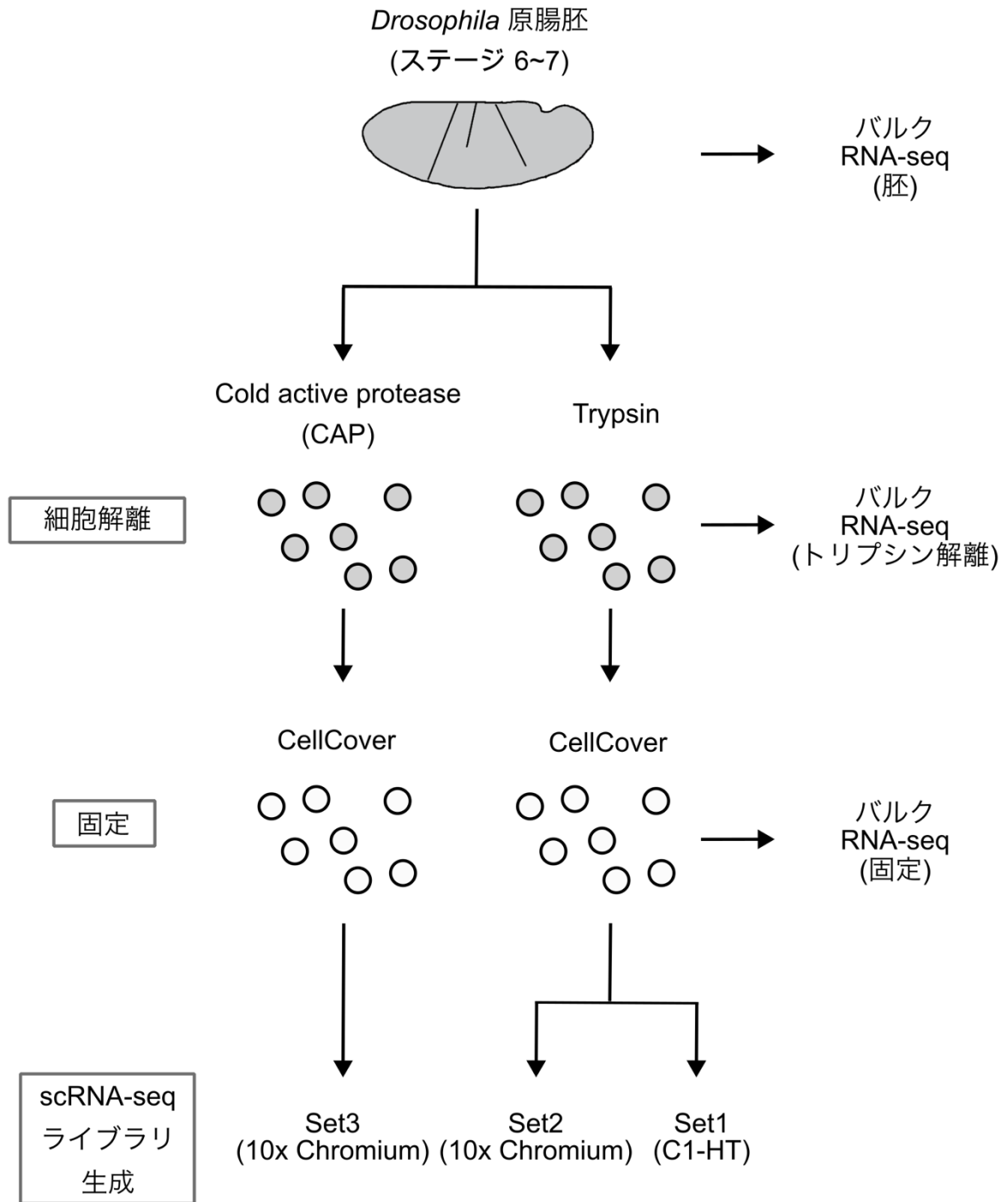


図 10

本実験で取得したデータの模式図

図 11 非固定細胞と固定細胞との間の bulk-RNA seq の比較

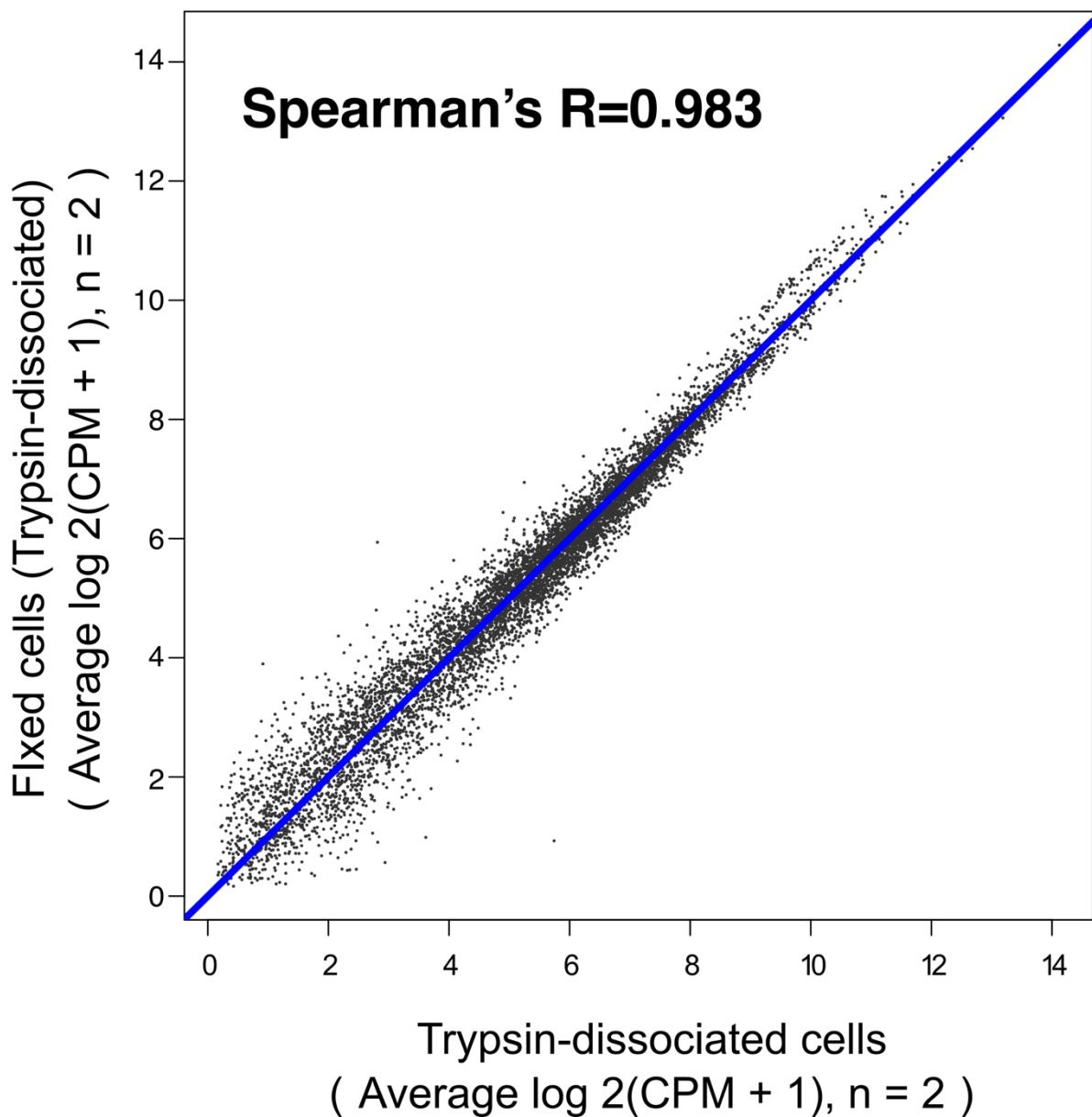


図 11

トリプシンによる細胞解離後の非固定細胞と固定細胞でのバルク RNA-seq の結果の散布図。各点が遺伝子を表し、横軸が細胞解離後の非固定細胞、縦軸が固定細胞での発現をそれぞれ表す。遺伝子発現量は(CMP + 1)の平均値の log₂ スケールで表示した。

図 12 C1HT データのバッチエフェクト

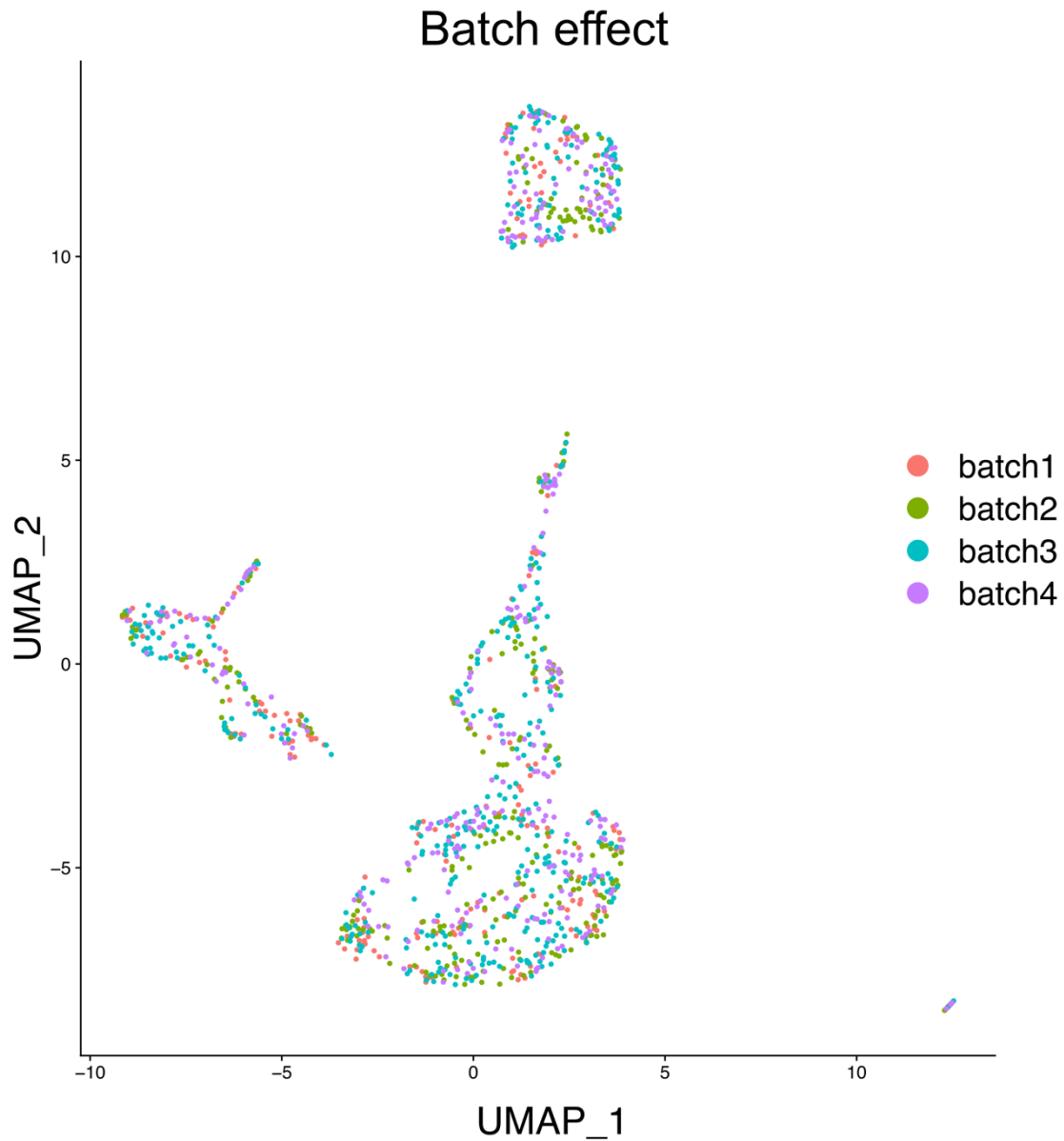


図 12

Set 1 の UMAP プロットに C1HT のバッチの情報を色で示してある。各バッチから得られたデータがプロット上で混合しており、バッチエフェクトがないことを示す。

図 13 トリプシンと CAP による細胞解離の比較

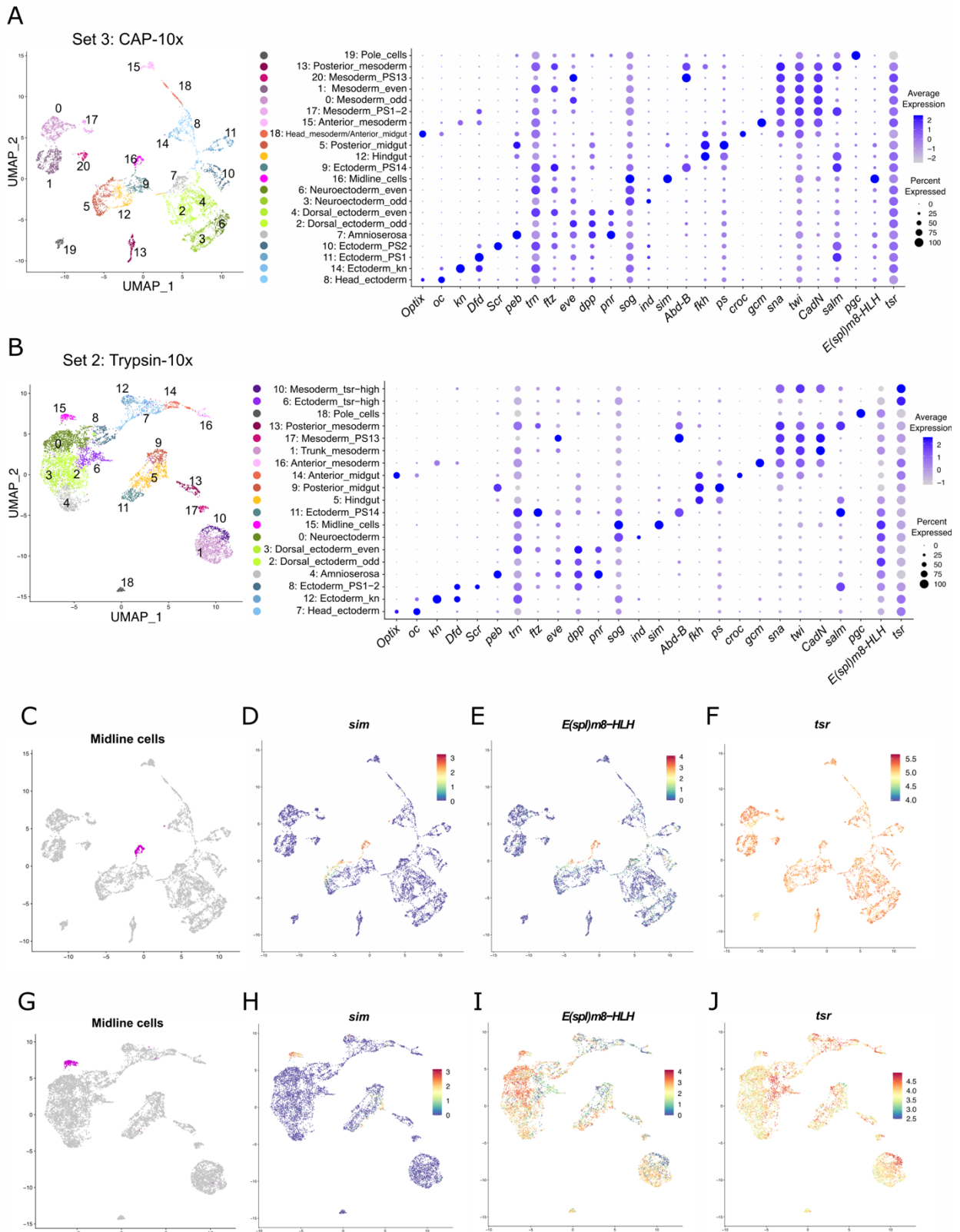


図 13

A-B. Set 3(A) および Set 2(B) の UMAP プロット。Seurat クラスタの情報を色で示してある。右のドットプロットには各クラスタの主要なマーカー遺伝子の発現パターンを示す。

C. Set 3 における正中線細胞のクラスタを UMAP プロット上でマゼンタ色で示した。

D-F. Set 3 における *sim* (D)、*E(spl)m8-HLH* (E)、*tsr* (F)の発現パターンを UMAP プロット上に示した。

G. Set 2 における正中線細胞のクラスタを UMAP プロット上でマゼンタ色で示した。

H-J. Set 2 における *sim* (H)、*E(spl)m8-HLH* (I)、*tsr* (J)の発現パターンを UMAP プロット上に示した。

図 14 C1HT データの Seurat クラスタ

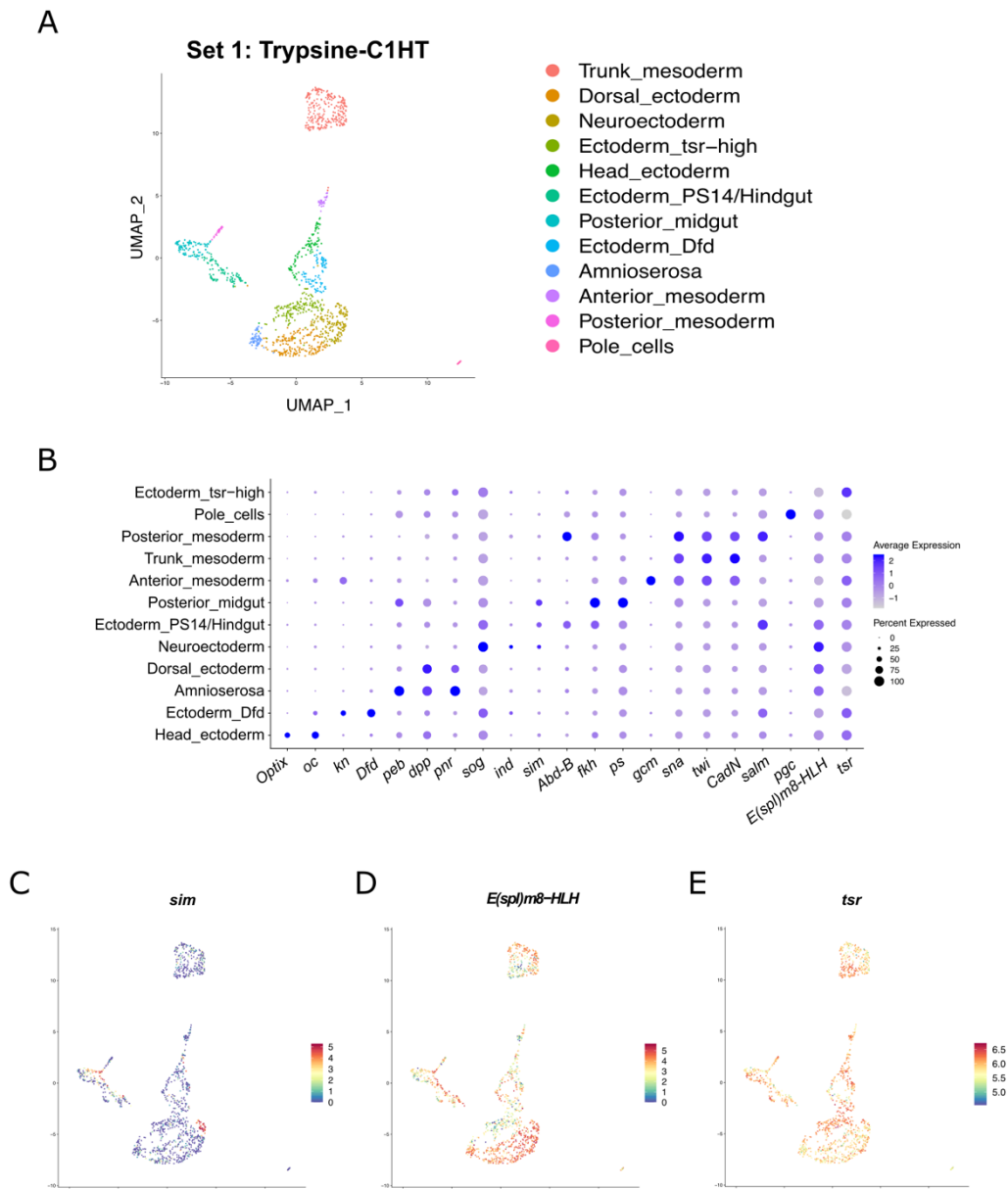


図 14

A. Set 1 の UMAP プロット。Seurat クラスタの情報を色で示してある。

B. Set 1 の各クラスタの主要なマーカー遺伝子の発現パターンを示したドットプロット。

C-E. Set 1 における *sim* (C)、*E(spl)m8-HLH* (D)、*tsr* (E) の発現パターンを UMAP 上に示した。

図 15 トリプシン処理の影響

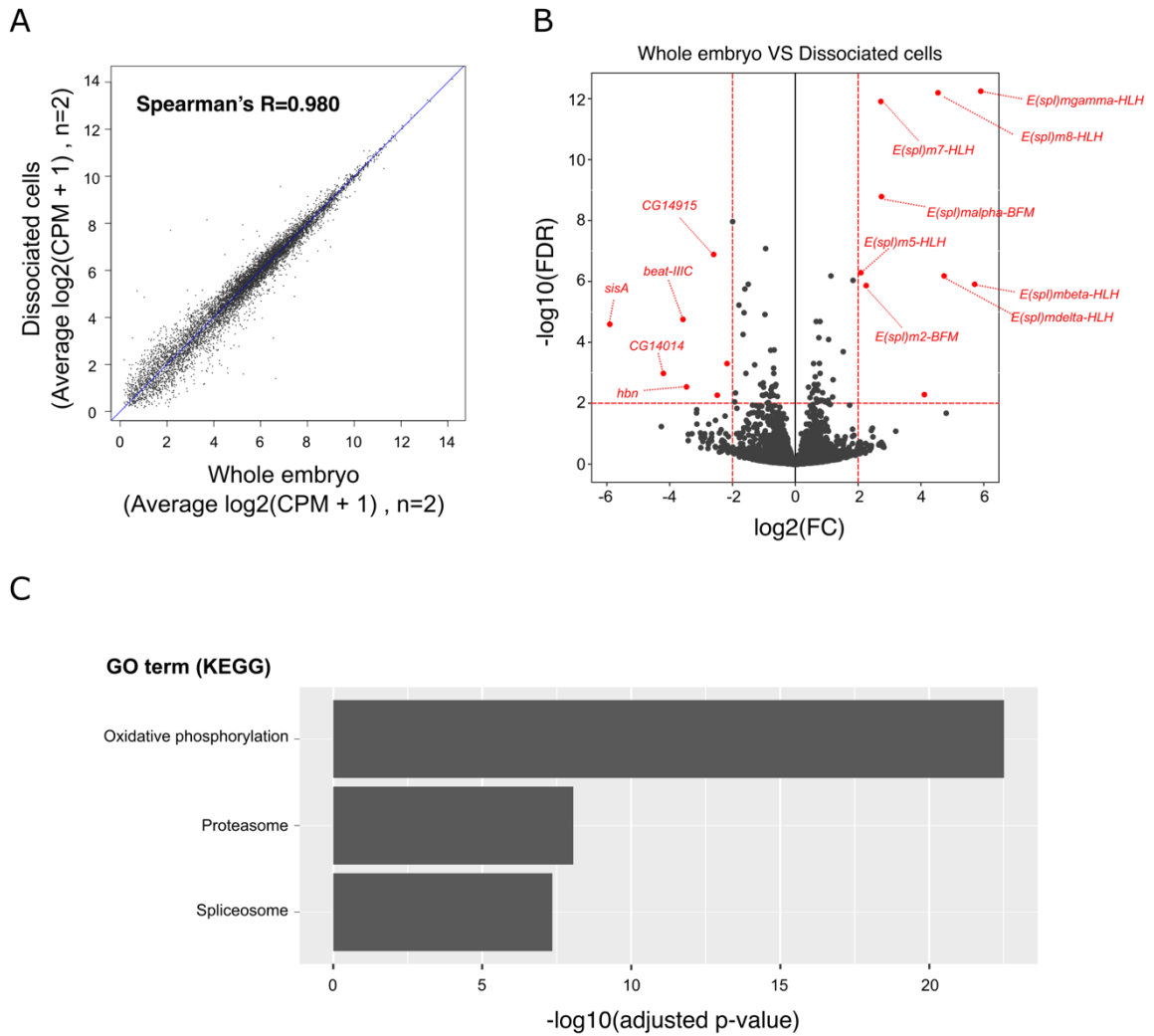


図 15

A. 胚と解離後の細胞でのバルク RNA-seq の結果の散布図。各点が遺伝子を表し、横軸が胚、縦軸が細胞解離後の非固定細胞の発現をそれぞれ表す。遺伝子発現量は(CMP + 1)の平均値の log₂ スケールで表示した。

B. 胚と解離後の細胞でのバルク RNA-seq との間のヴォルケーノプロット。各点が遺伝子を表し、横軸が発現量変化 (log₂ スケール)、縦軸が -log₁₀(FDR) の値を示す。赤い点は |logFC| > 2 and FDR < 0.01 を満たす DEG である。

C. Set 2 のクラスタ 10 でクラスタ 1 よりも有意に高く発現していた遺伝子に濃縮していた GO ターム。タームは KEGG のものだけを示した。

図 16 体幹部外胚葉の細胞のアノテーション

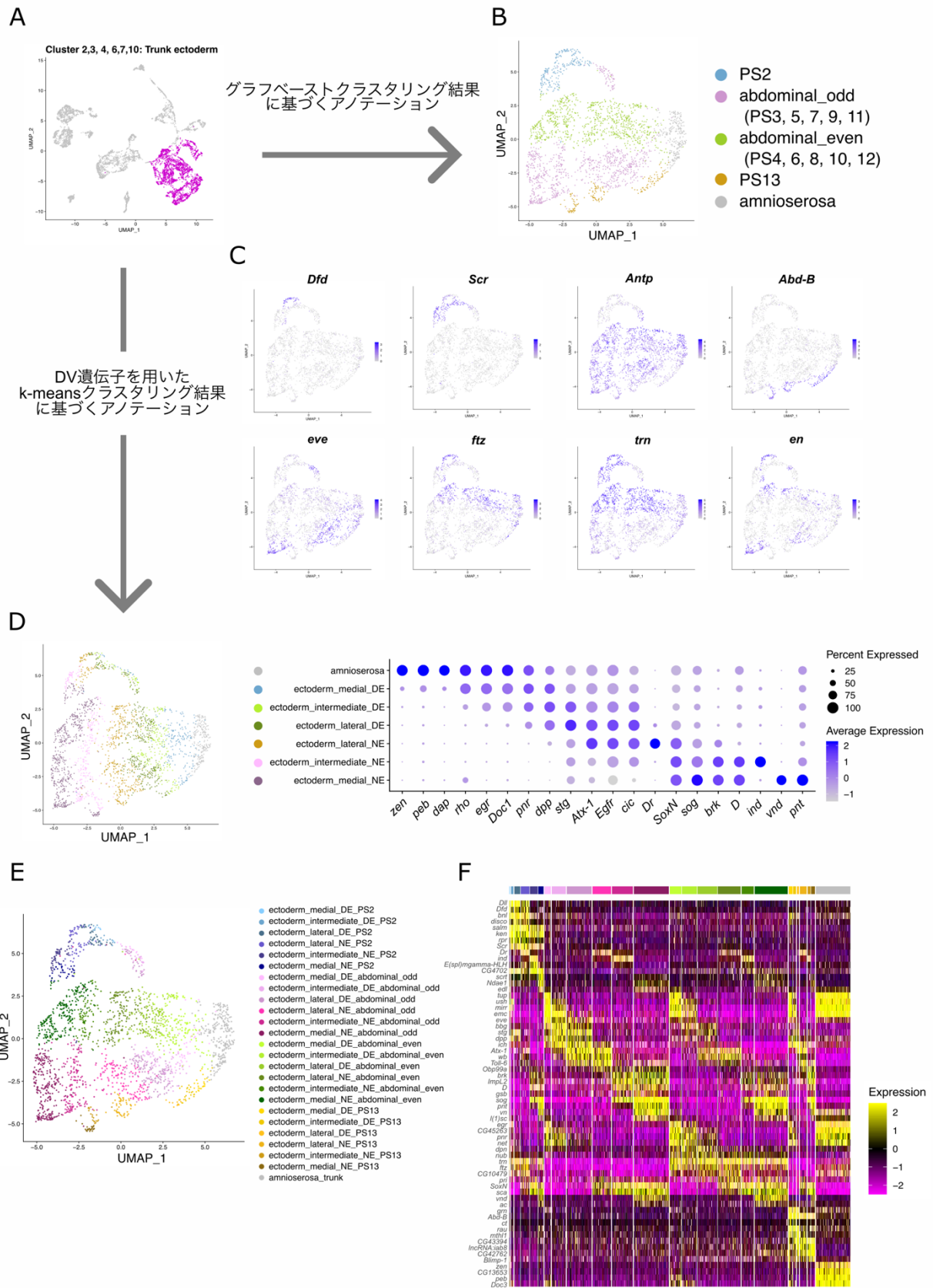


図 16

- A. 体幹部の外胚葉を UMAP 上にマゼンタで示した。
- B. 体幹部の外胚葉の前後軸に関するアノテーション結果。
- C. 前後軸で特異的な発現を示す遺伝子の発現パターン。
- D. 体幹部の外胚葉の背腹軸に関するアノテーション結果。右のドットプロットは主要なマーカー遺伝子の発現パターンを示す。
- E. 前後軸、背腹軸のアノテーションを組み合わせた体幹部の外胚葉のアノテーション結果。
- F. 各クラスターの主要なマーカー遺伝子の発現パターンを示すヒートマップ。

図 17 各クラスターのサブクラスタリング結果 (1/5)

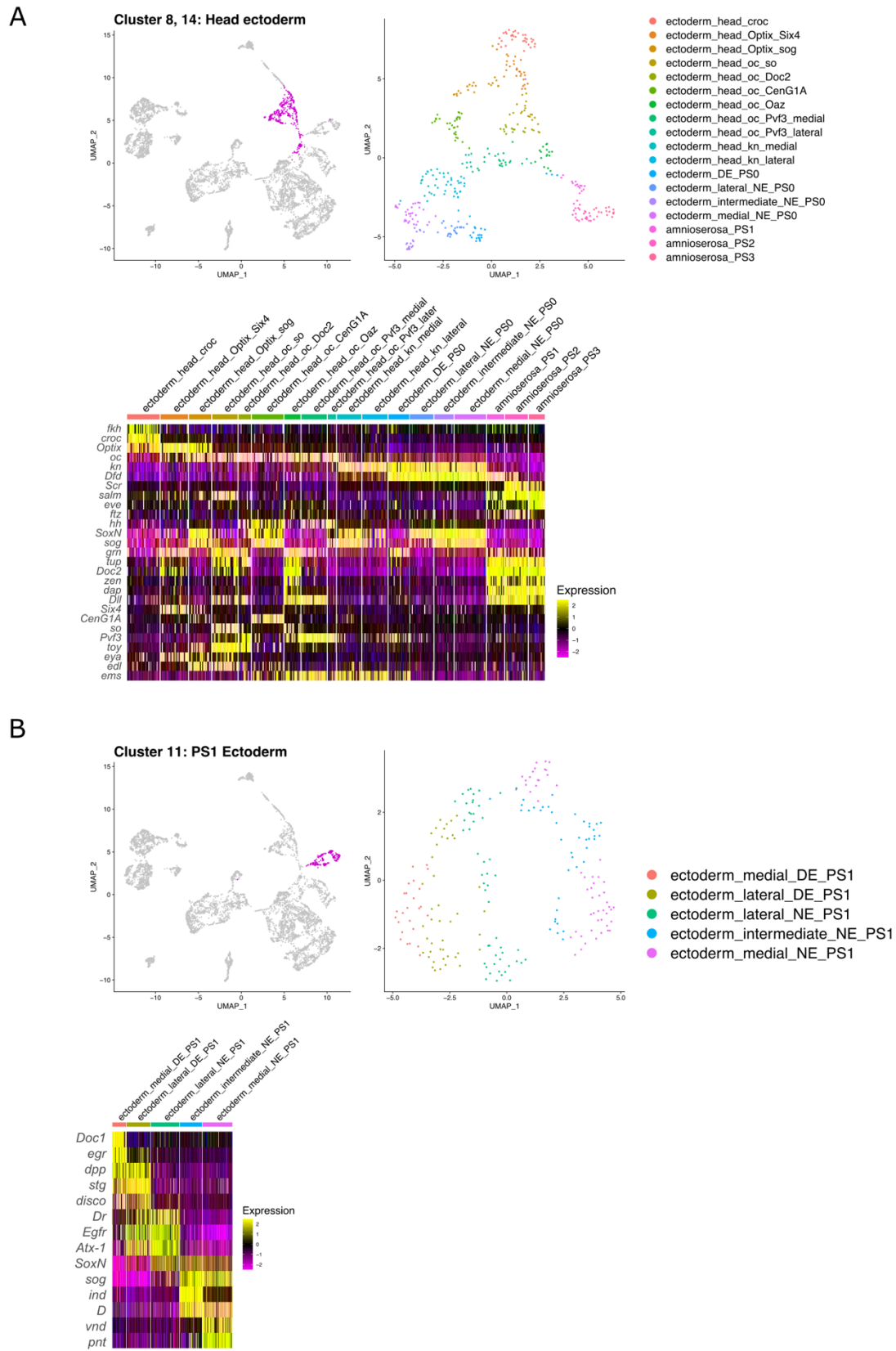
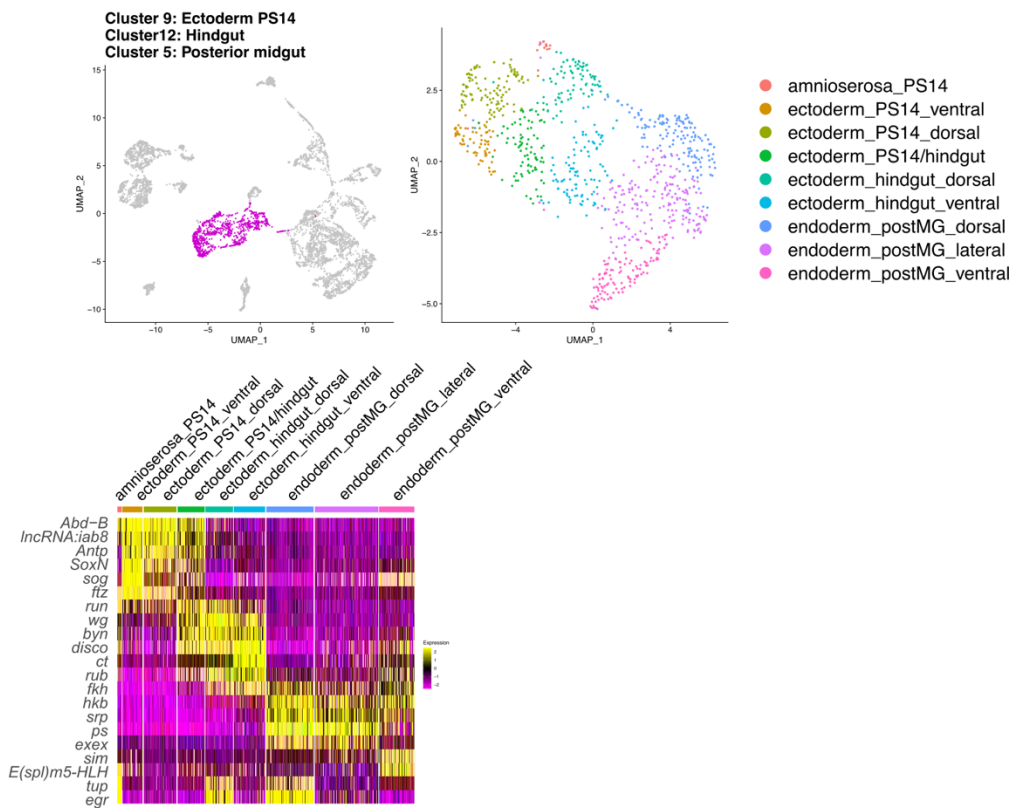


図 17 続き 各クラスターのサブクラスタリング結果 (2/5)

C



D

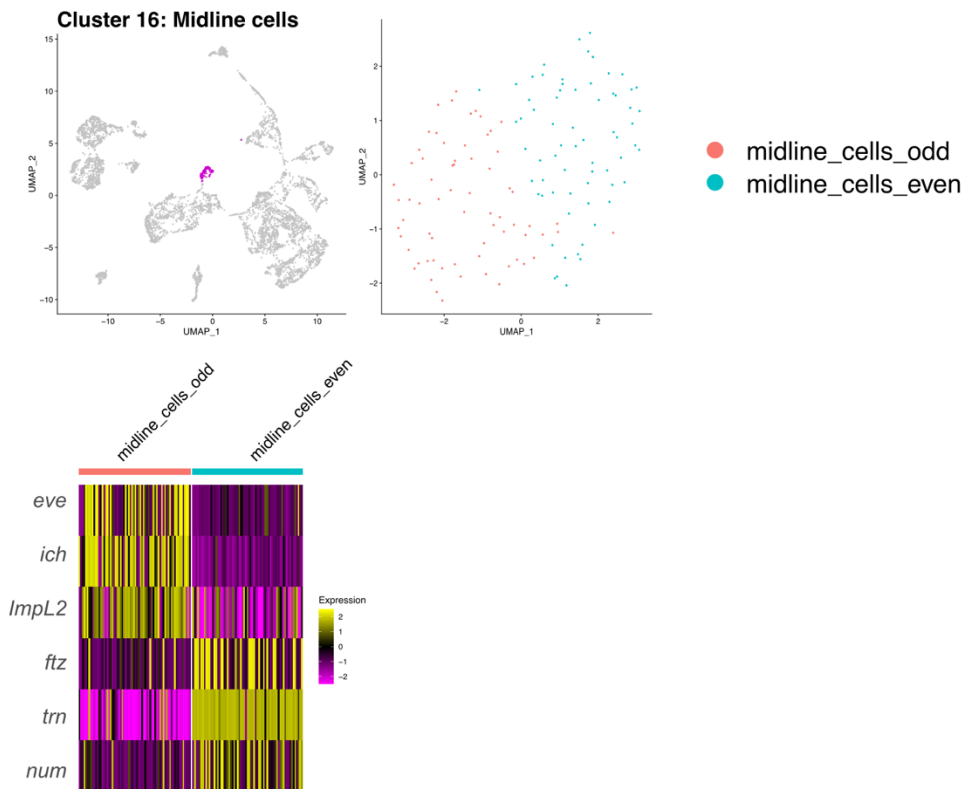


図 17 続き 各クラスターのサブクラスタリング結果 (3/5)

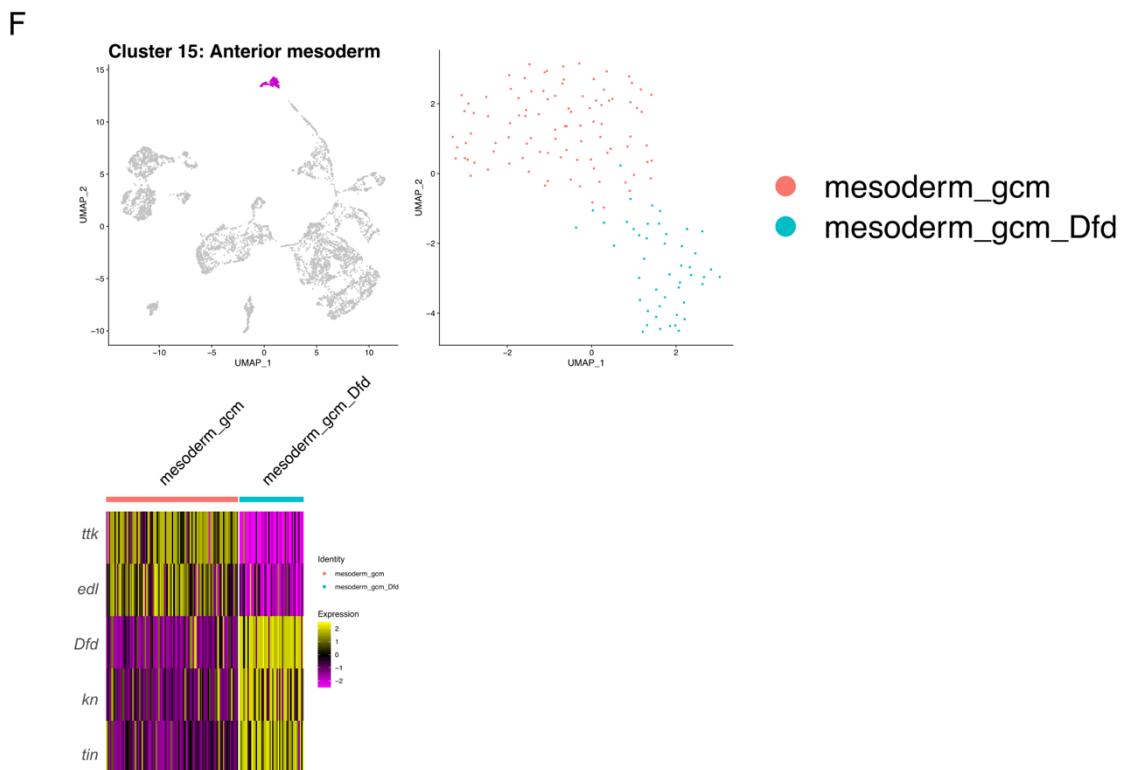
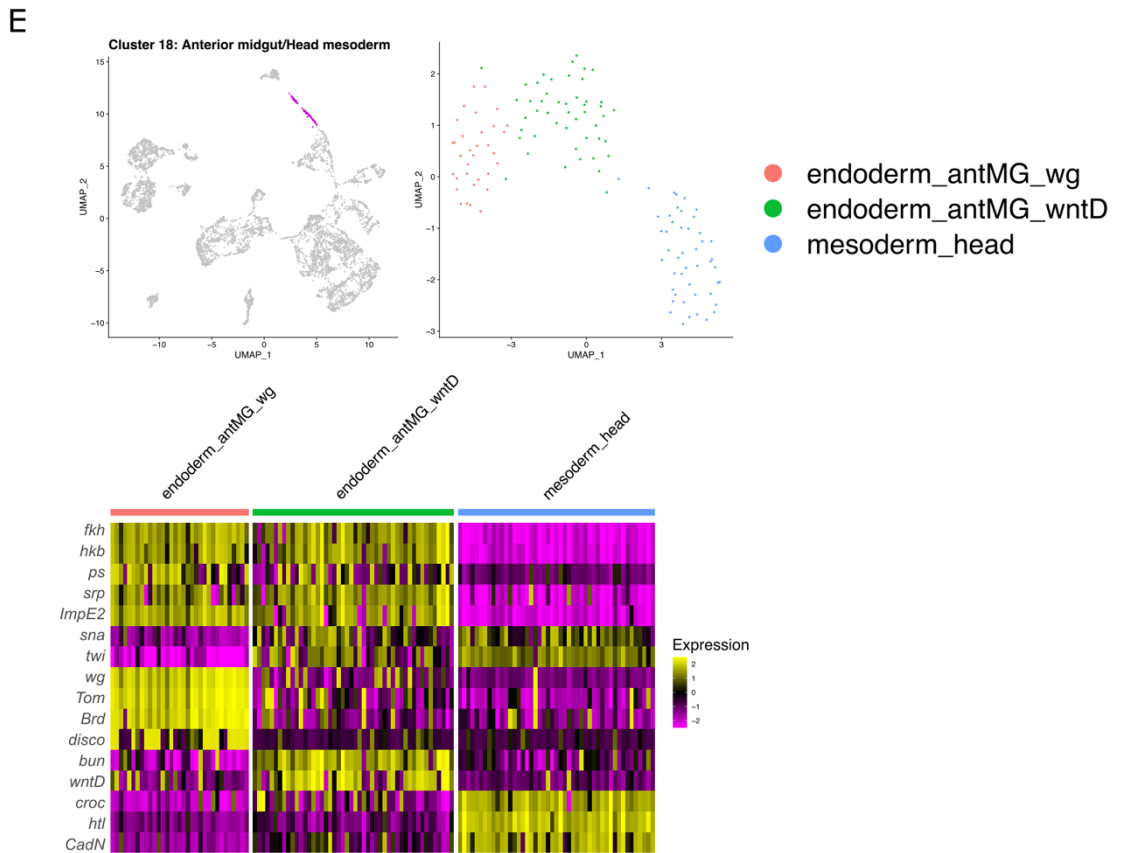


図 17 続き 各クラスターのサブクラスティング結果 (4/5)

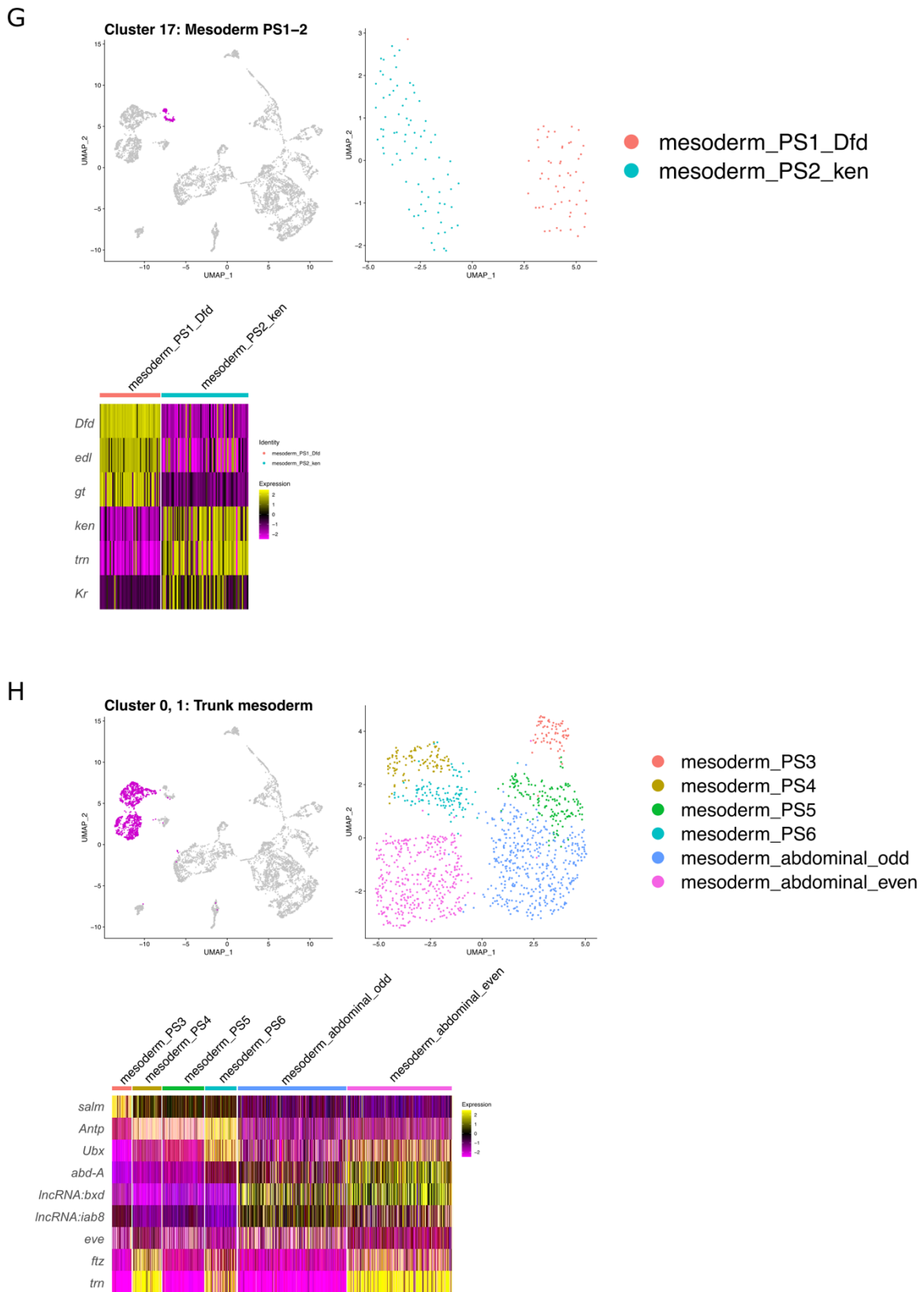


図 17 続き 各クラスタのサブクラスタリング結果 (5/5)

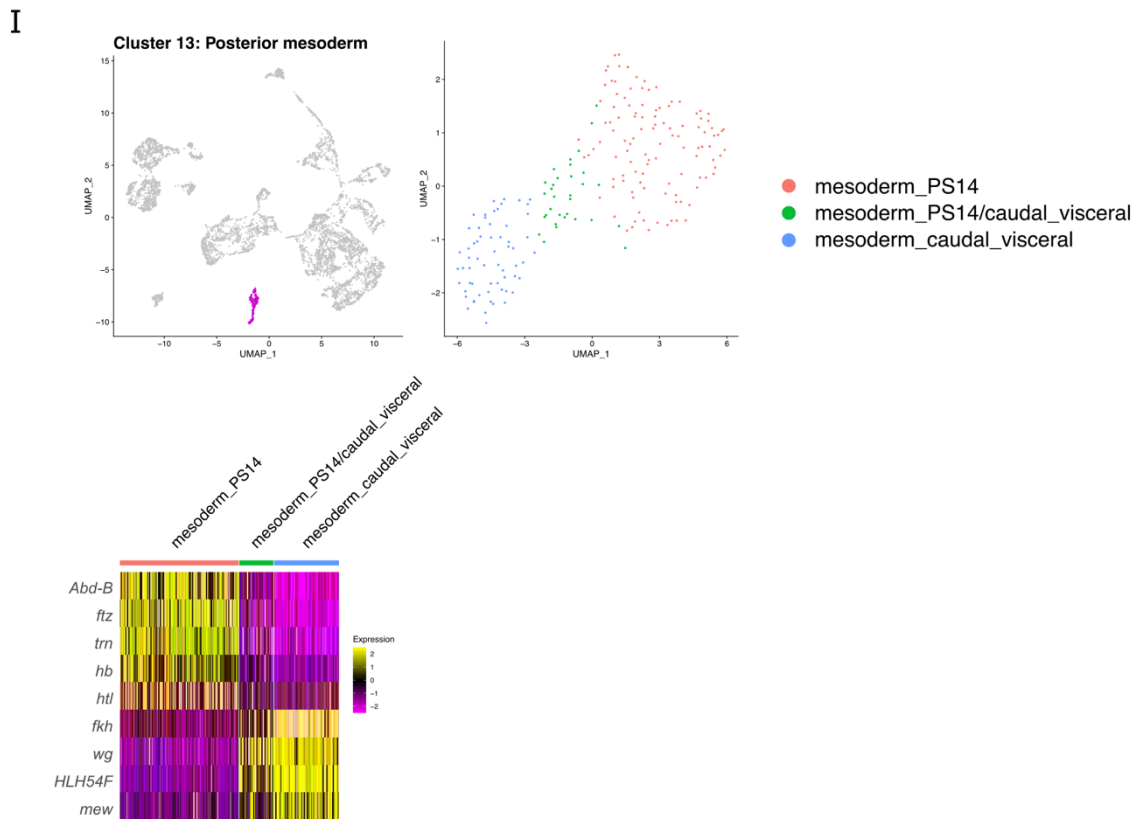
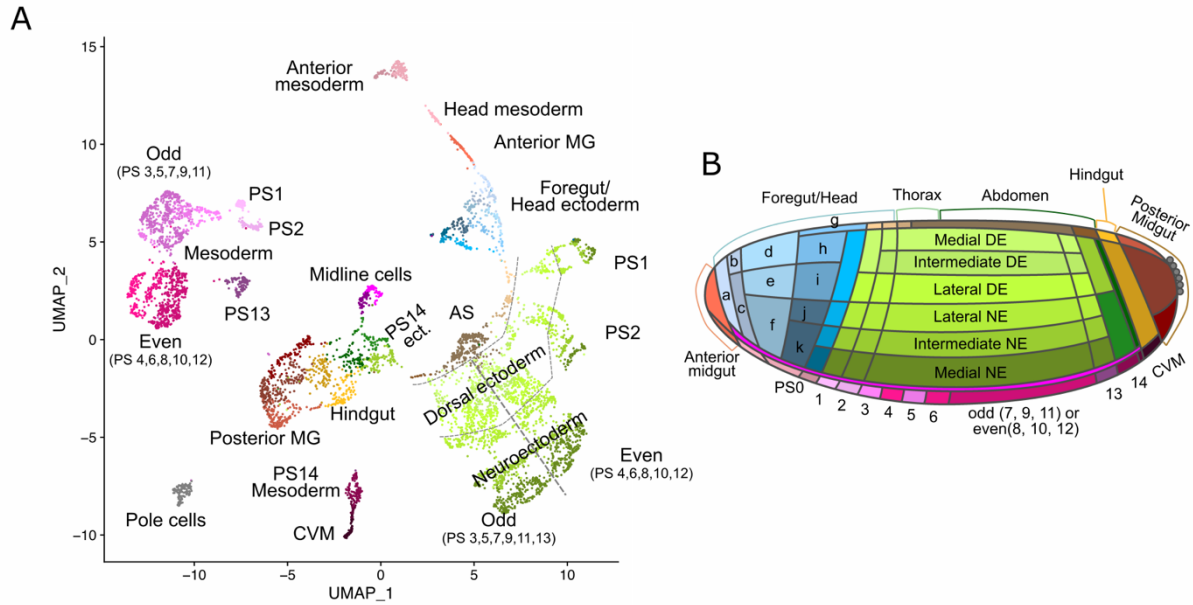


図 17

各パネルの左上はサブクラスタリングの対象となった細胞を UMAP 上にマゼンタで示した。左上のパネルは UMAP 上でサブクラスタ情報を色で表示してある。右下のパネルはサブクラスタの主要なマーカーの発現をヒートマップで示した。

- A. 頭部外胚葉 (Seurat cluster 8 および 14)
- B. パラセグメント 1 の外胚葉 (Seurat cluster 11)
- C. 胚後方外胚葉および内胚葉 (Seurat cluster 5、9 および 12)
- D. 正中線細胞 (Seurat cluster 16)
- E. 胚前方内胚葉および頭部中胚葉 (Seurat cluster 18)
- F. 胚前方中胚葉 (Seurat cluster 15)
- G. パラセグメント 1-2 の中胚葉 (Seurat cluster 17)
- H. 体幹部中胚葉 (Seurat cluster 0 および 1)
- I. 後方中胚葉 (Seurat cluster 13)

図 18 Set 3 でアノテーションされた 77 サブクラスタ



Ectoderm

Head region

- a. ectoderm_head_croc
- b. ectoderm_head_Optix_Six4
- c. ectoderm_head_Optix_sog
- d. ectoderm_head_oc_so
- e. ectoderm_head_oc_Doc2
- f. ectoderm_head_oc_CenG1A
- g. ectoderm_head_oc_Oaz
- h. ectoderm_head_oc_Pvf3_medial
- i. ectoderm_head_oc_Pvf3_lateral
- j. ectoderm_head_kn_lateral
- k. ectoderm_head_kn_medial
- ectoderm_DE_PS0
- ectoderm_lateral_NE_PS0
- ectoderm_intermediate_NE_PS0
- ectoderm_medial_NE_PS0

PS1-2

- ectoderm_medial_DE_PS1
- ectoderm_lateral_DE_PS1
- ectoderm_lateral_NE_PS1
- ectoderm_intermediate_NE_PS1
- ectoderm_medial_NE_PS1
- ectoderm_medial_DE_PS2
- ectoderm_intermediate_DE_PS2
- ectoderm_lateral_DE_PS2
- ectoderm_lateral_NE_PS2
- ectoderm_intermediate_NE_PS2
- ectoderm_medial_NE_PS2

Trunk region

- ectoderm_medial_DE_abdominal_odd
- ectoderm_medial_DE_abdominal_even
- ectoderm_medial_DE_PS13
- ectoderm_intermediate_DE_abdominal_odd
- ectoderm_intermediate_DE_abdominal_even
- ectoderm_intermediate_DE_PS13
- ectoderm_lateral_DE_abdominal_odd
- ectoderm_lateral_DE_abdominal_even
- ectoderm_lateral_DE_PS13
- ectoderm_lateral_NE_abdominal_odd
- ectoderm_lateral_NE_abdominal_even
- ectoderm_lateral_NE_PS13
- ectoderm_intermediate_NE_abdominal_odd
- ectoderm_intermediate_NE_abdominal_even
- ectoderm_intermediate_NE_PS13
- ectoderm_medial_NE_abdominal_odd
- ectoderm_medial_NE_abdominal_even
- ectoderm_medial_NE_PS13

- ectoderm_PS14_dorsal
- ectoderm_PS14_ventral
- ectoderm_PS14/hindgut

Midline cells

- midline_cells_even
- midline_cells_odd

Hindgut

- ectoderm_hindgut_dorsal
- ectoderm_hindgut_ventral

Amnioserosa

- amnioserosa_PS1
- amnioserosa_PS2
- amnioserosa_PS3
- amnioserosa_trunk
- amnioserosa_PS14

Mesoderm

- mesoderm_head
- mesoderm_gcm
- mesoderm_gcm_Dfd
- mesoderm_PS1_Dfd
- mesoderm_PS2_ken
- mesoderm_PS3
- mesoderm_PS4
- mesoderm_PS5
- mesoderm_PS6
- mesoderm_abdominal_odd
- mesoderm_abdominal_even
- mesoderm_PS13
- mesoderm_PS14
- mesoderm_PS14/caudal_visceral
- mesoderm_caudal_visceral

Endoderm

- endoderm_antMG_wg
- endoderm_antMG_wntD
- endoderm_postMG_dorsal
- endoderm_postMG_lateral
- endoderm_postMG_ventral

Germ cell

- pole_cells

図 18

A. Set 3 の 77 サブクラスタの情報を UMAP 上に色で示した。

B. サブクラスタが対応する胚における位置を示した模式図。各領域の色は A の点の色と対応している。

図 19 ダブレット細胞の例

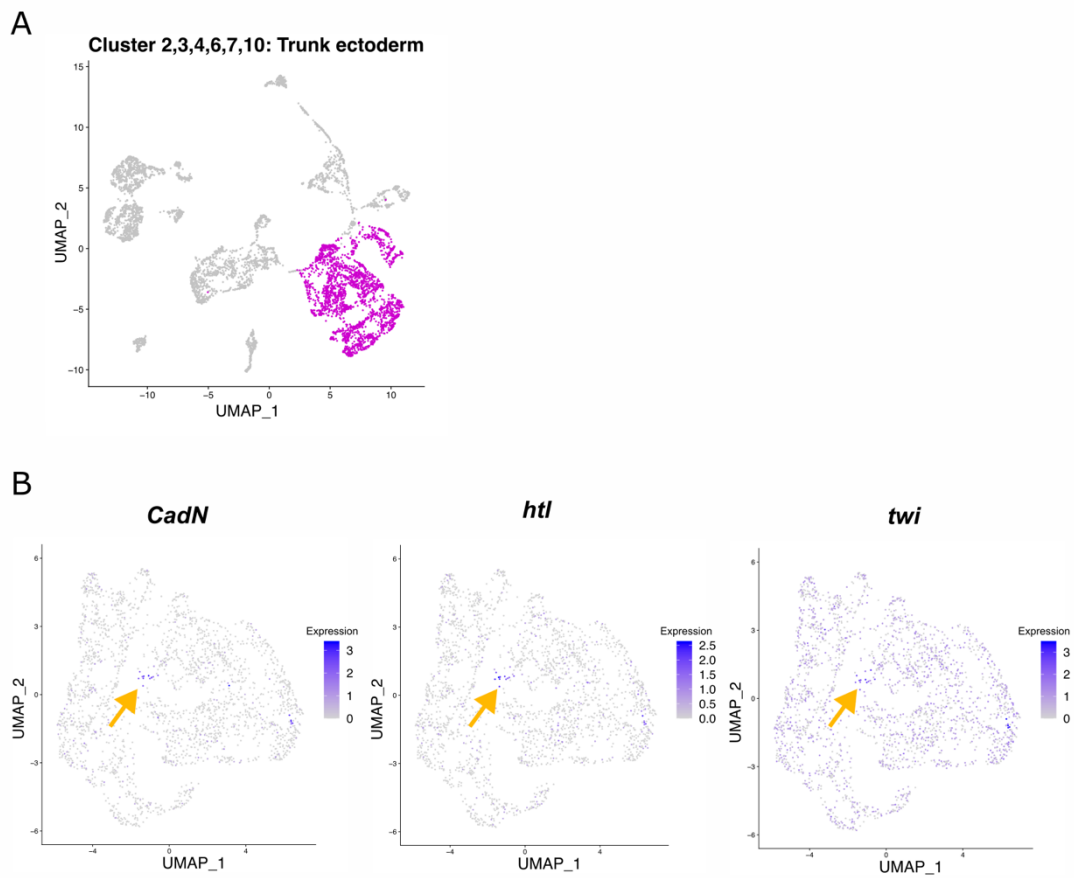


図 19

A. 体幹部の外胚葉を UMAP 上にマゼンタで示した (図 16 の A と同一)

B. A で示した細胞における中胚葉マーカー遺伝子の発現を UMAP 上に示したもの。

Cadherin-N (*CadN*) (左)、*heartless* (*htl*) (中央)、*twi*(右)。黄色の矢印がダブルット細胞を示す。

図 20 Set 2 のクラスタリング結果

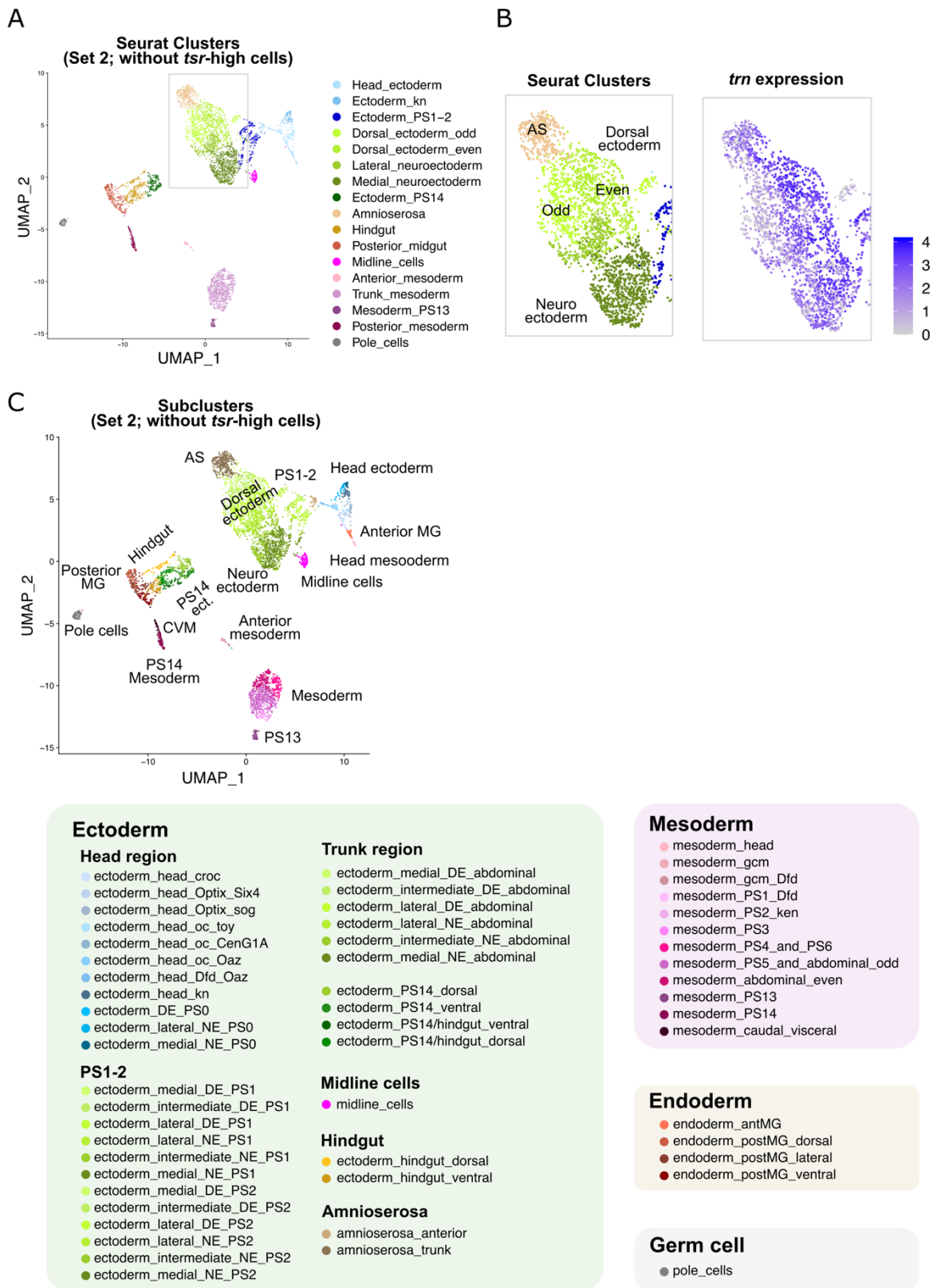


図 20

A. Set 2 における Seurat clusters の情報を UMAP プロット上に色で示した。四角の領域を B に拡大して示している。

B. A において四角で示した領域の拡大図（左）と同じ領域における *trn* 発現量（右）。

C. Set 2 におけるサブクラスタの情報を UMAP プロット上に色で示した。

図 21 Set1 および NK-data のサブクラスタリング結果

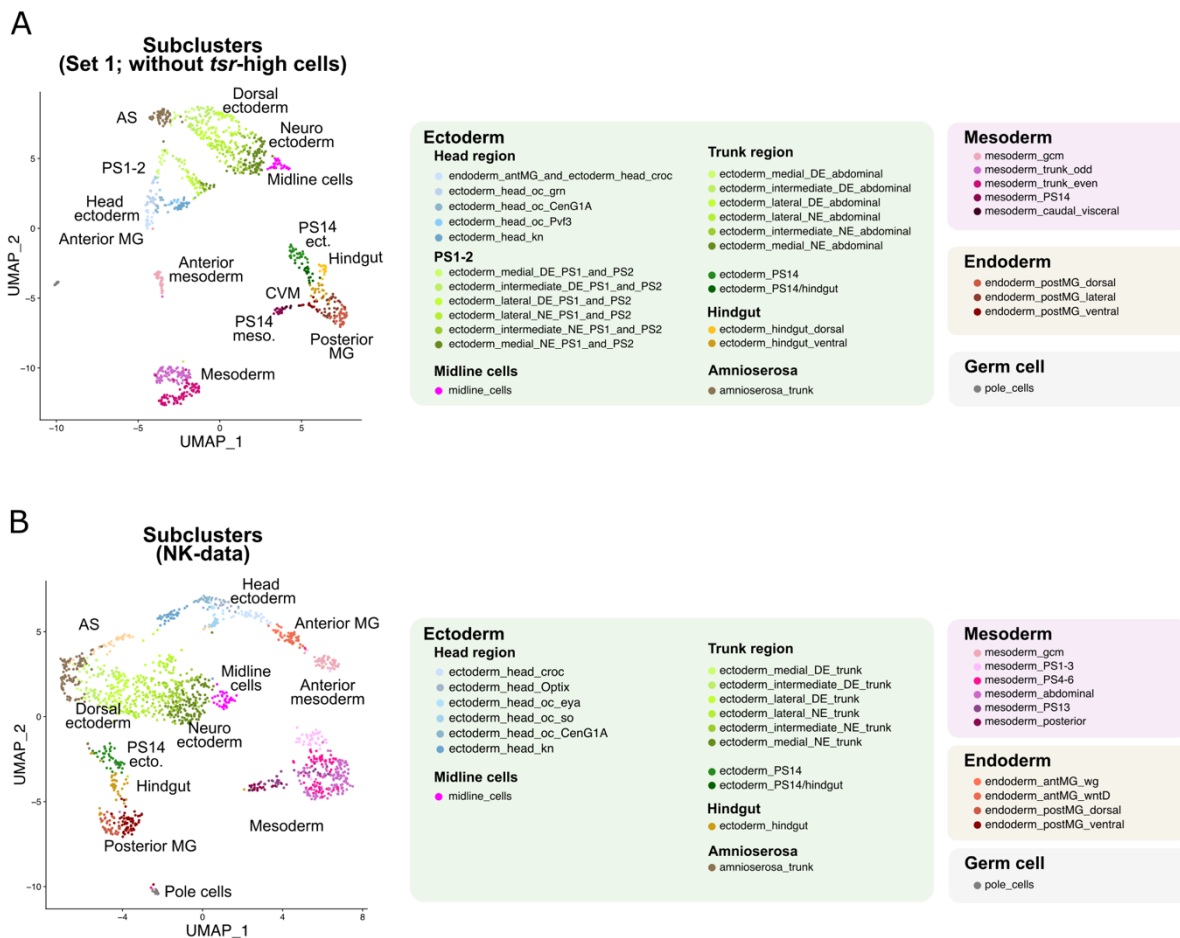


図 21

A. Set 1 におけるサブクラスタの情報を UMAP プロット上に色で示した。

B. NK-data におけるサブクラスタの情報を UMAP プロット上に色で示した。

図 22 データの統合解析

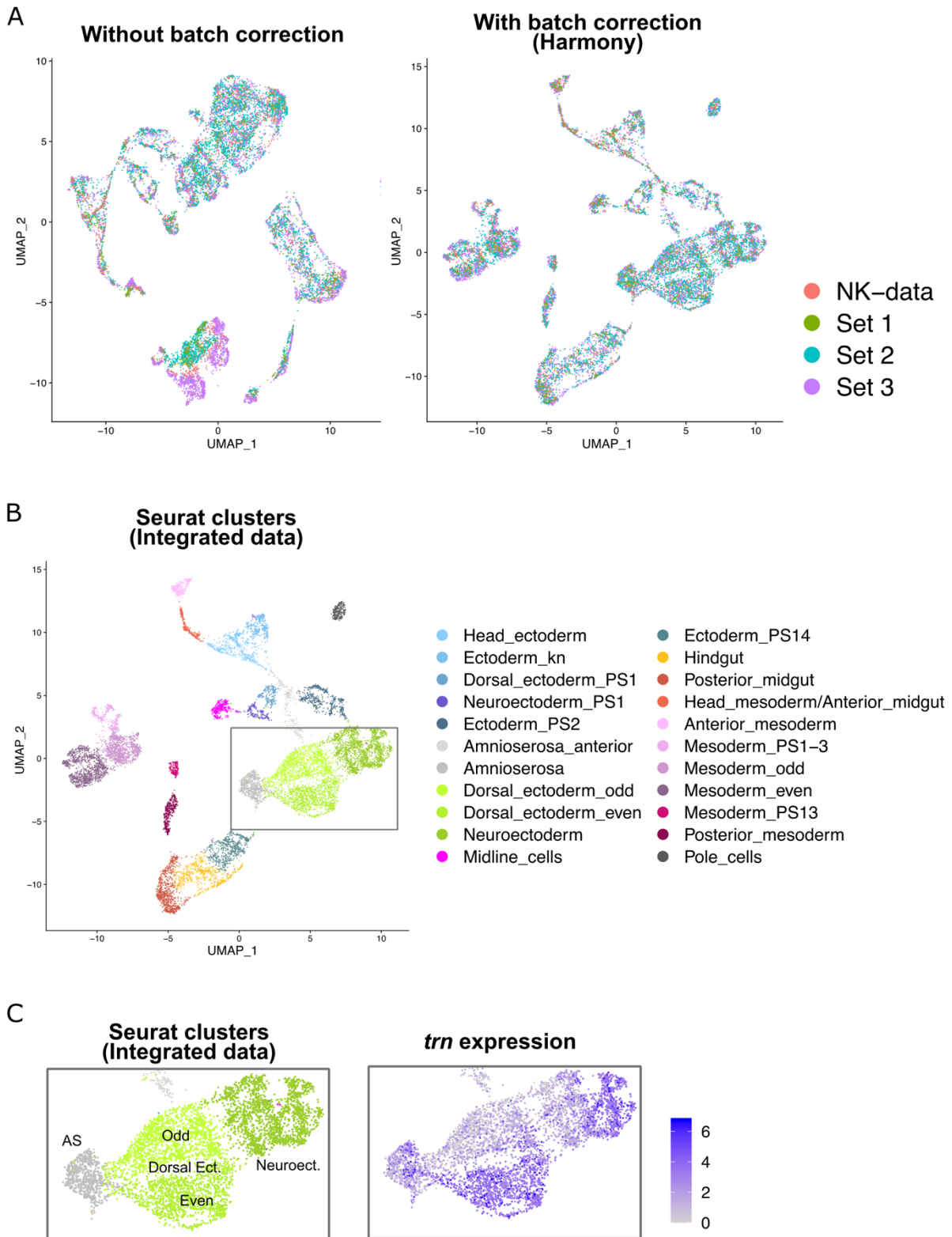
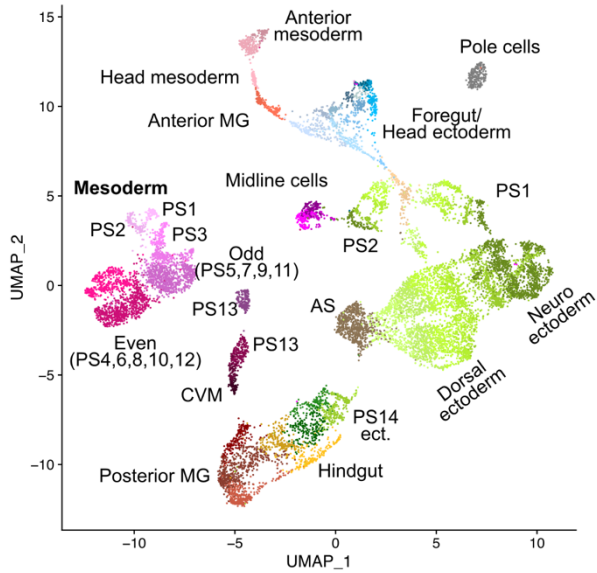


図 22

- A. 4つのデータセットを合わせ、(左) バッチコレクションなし、(右) バッチコレクションありで次元削減した UMAP プロット。点の色でどのデータに由来するかを示している。
- B. 統合データの Seurat クラスタの情報を UMAP 上に色で示した。四角で囲った領域を C で拡大して表示している。
- C. B で示した領域の拡大図 (左) と同じ領域における *trn* の発現パターン。

図 23 統合データのサブクラスティング結果



Ectoderm

Head region

- ectoderm_head_fkh
- ectoderm_head_croc
- ectoderm_head_Optix_Six4_SoxN
- ectoderm_head_Optix_Six4_grn
- ectoderm_head_Optix_toy
- ectoderm_head_Optix_CenG1A
- ectoderm_head_oc_eya
- ectoderm_head_oc_so_toy
- ectoderm_head_oc_so
- ectoderm_head_oc_Doc2
- ectoderm_head_oc_CenG1A
- ectoderm_head_oc_Pvf3
- ectoderm_head_kn_lateral
- ectoderm_head_kn_medial
- ectoderm_head_Dfd_Oaz
- ectoderm_PS0_ems
- ectoderm_DE_PS0
- ectoderm_lateral_NE_PS0
- ectoderm_intermediate_NE_PS0
- ectoderm_medial_NE_PS0

PS1-2

- ectoderm_medial_DE_PS1
- ectoderm_lateral_DE_PS1
- ectoderm_lateral_NE_PS1
- ectoderm_intermediate_NE_PS1
- ectoderm_medial_NE_PS1
- ectoderm_medial_DE_PS2
- ectoderm_intermediate_DE_PS2
- ectoderm_lateral_DE_PS2
- ectoderm_lateral_NE_PS2
- ectoderm_intermediate_NE_PS2
- ectoderm_medial_NE_PS2

Midline cells

- midline_cells_PS1-3
- midline_cells_abdominal_even
- midline_cells_abdominal_odd

Trunk region

- ectoderm_medial_DE_abdominal
- ectoderm_intermediate_DE_abdominal
- ectoderm_lateral_DE_abdominal
- ectoderm_lateral_NE_abdominal
- ectoderm_intermediate_NE_abdominal
- ectoderm_medial_NE_abdominal

- ectoderm_PS14_dorsal
- ectoderm_PS14_ventral
- ectoderm_PS14/hindgut

Amnioserosa

- amnioserosa_PS1
- amnioserosa_PS2
- amnioserosa_trunk

Hindgut

- ectoderm_hindgut_dorsal
- ectoderm_hindgut_ventral

Mesoderm

- mesoderm_head
- mesoderm_gcm
- mesoderm_gcm_Dfd
- mesoderm_PS1_Dfd
- mesoderm_PS2_ken
- mesoderm_PS3
- mesoderm_PS4_and_PS6
- mesoderm_PS5
- mesoderm_abdominal_odd
- mesoderm_abdominal_even
- mesoderm_PS13
- mesoderm_PS14
- mesoderm_PS14/caudal_visceral
- mesoderm_caudal_visceral

Endoderm

- endoderm_antMG_wg
- endoderm_antMG_wntD
- endoderm_postMG_dorsal
- endoderm_postMG_lateral
- endoderm_postMG_ventral

Germ cell

- pole_cells

図 23

統合データにおけるサブクラスタの情報を UMAP プロット上に色で示した。

AS: Amnioserosa, CVM: Caudal visceral mesoderm

図 24 FISH による intermediate 細胞の存在確認

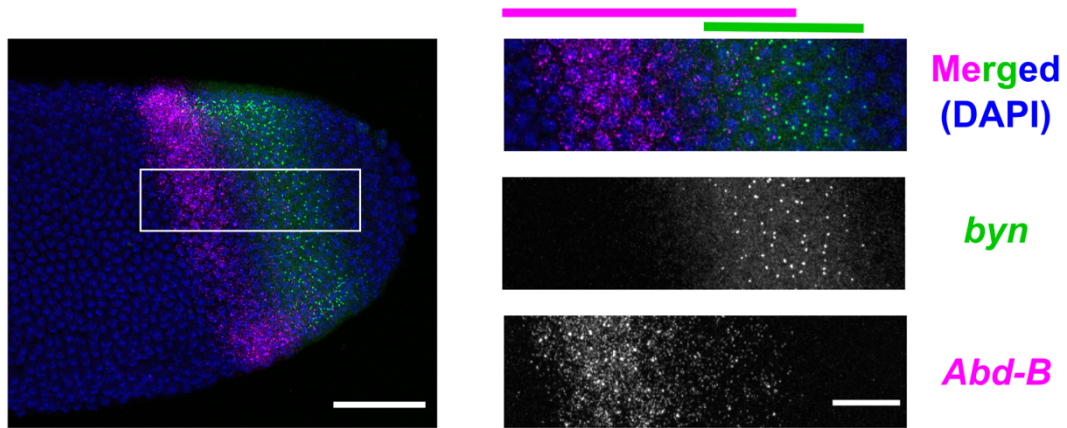


図 24

ショウジョウバエ原腸胚の後方における *Abd-B* (マゼンタ) および *byn* (緑) の mRNA 発現を SABER-FISH によって検出したもの。核は DAPI により染色した (青)。右のパネルは左のパネルの四角で囲った領域の拡大図。スケールバーは 50 μ m(左) および 20 μ m(右)。

図 25 3,000HVGs の Gene Ontology enrichment 解析の結果

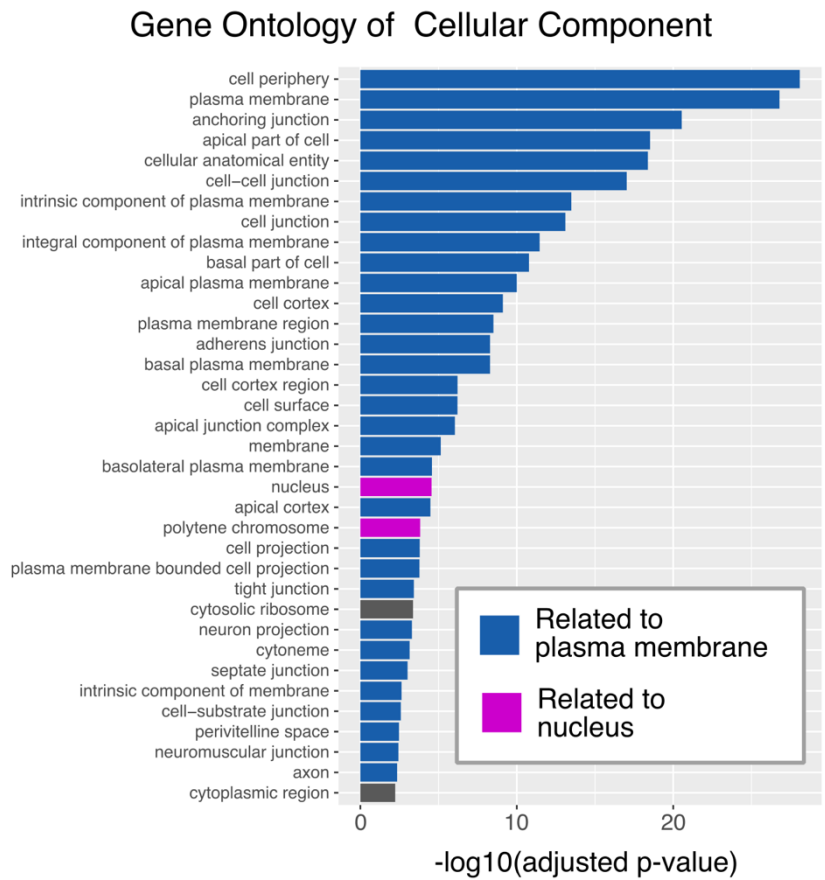


図 25

3,000 HVGs における g:Profiler を用いた Gene Ontology term enrichment analysis の結果。Cellular Components のタームのみを示してある。横軸に補正 p-value の逆数の対数を取って棒グラフとして表示した。棒の色は青が細胞膜に関連するターム、マゼンタが核に関連するターム、黒がいずれでもない（細胞質に関連する）タームであることをそれぞれ示す。

図 26 GLAD カテゴリの enrichment 解析

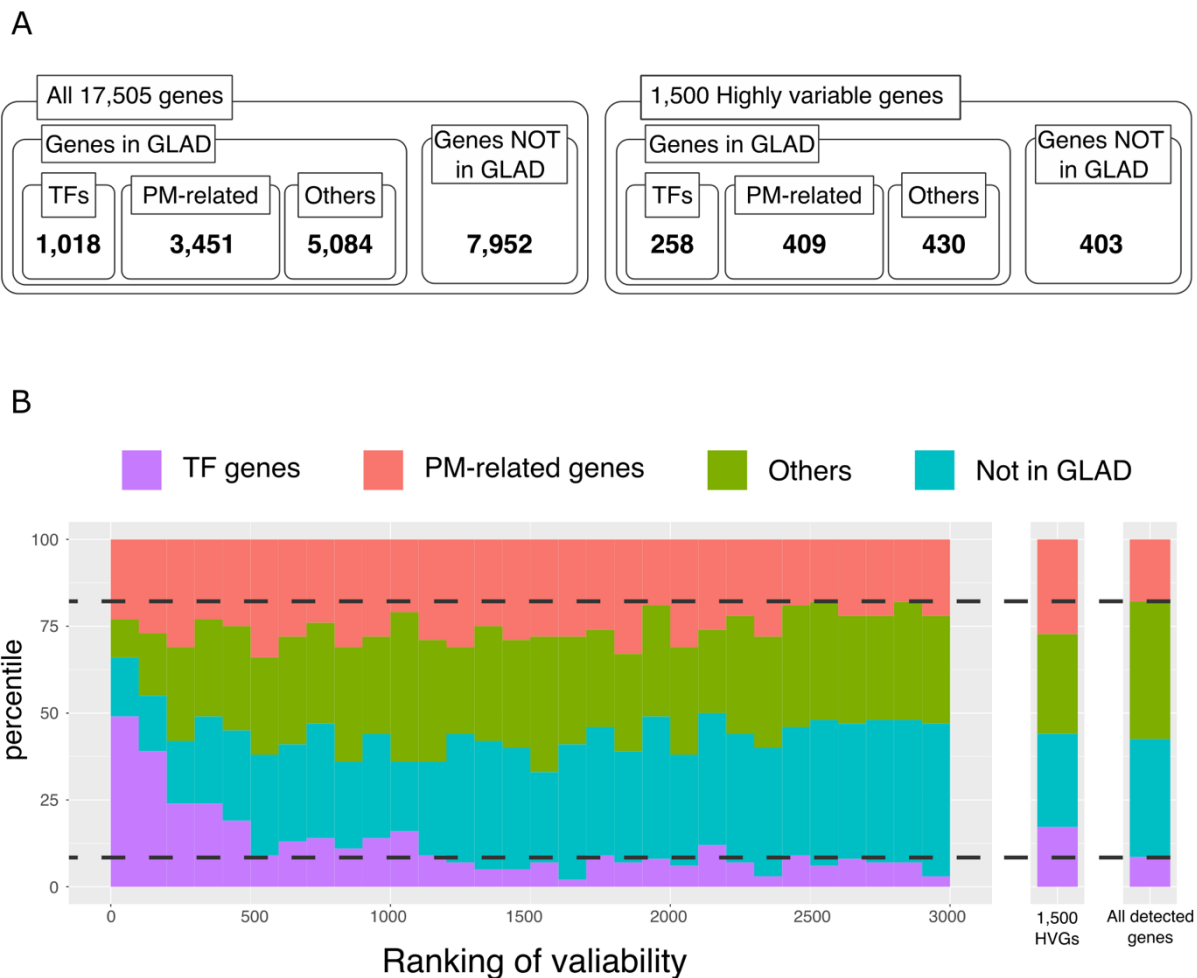


図 26

A. 各 GLAD カテゴリに属する遺伝子数を示した。左が全遺伝子、右が上位 1,500HVGs について示したものである。

B. 3,000 HVGs を細胞間分散の大きさに順位づけして 100 個ずつ 30 のビンに振り分け、各遺伝子がどのカテゴリに属するかの割合をビンごとに示した。黒い点線はそれぞれ全検出遺伝子からランダムに選んだ場合の細胞膜関連遺伝子 (PM-related genes) の割合と転写因子 (TF) の割合を示す。

図 27 1,500 HVGs を用いたサブクラスタの階層的クラスタリング

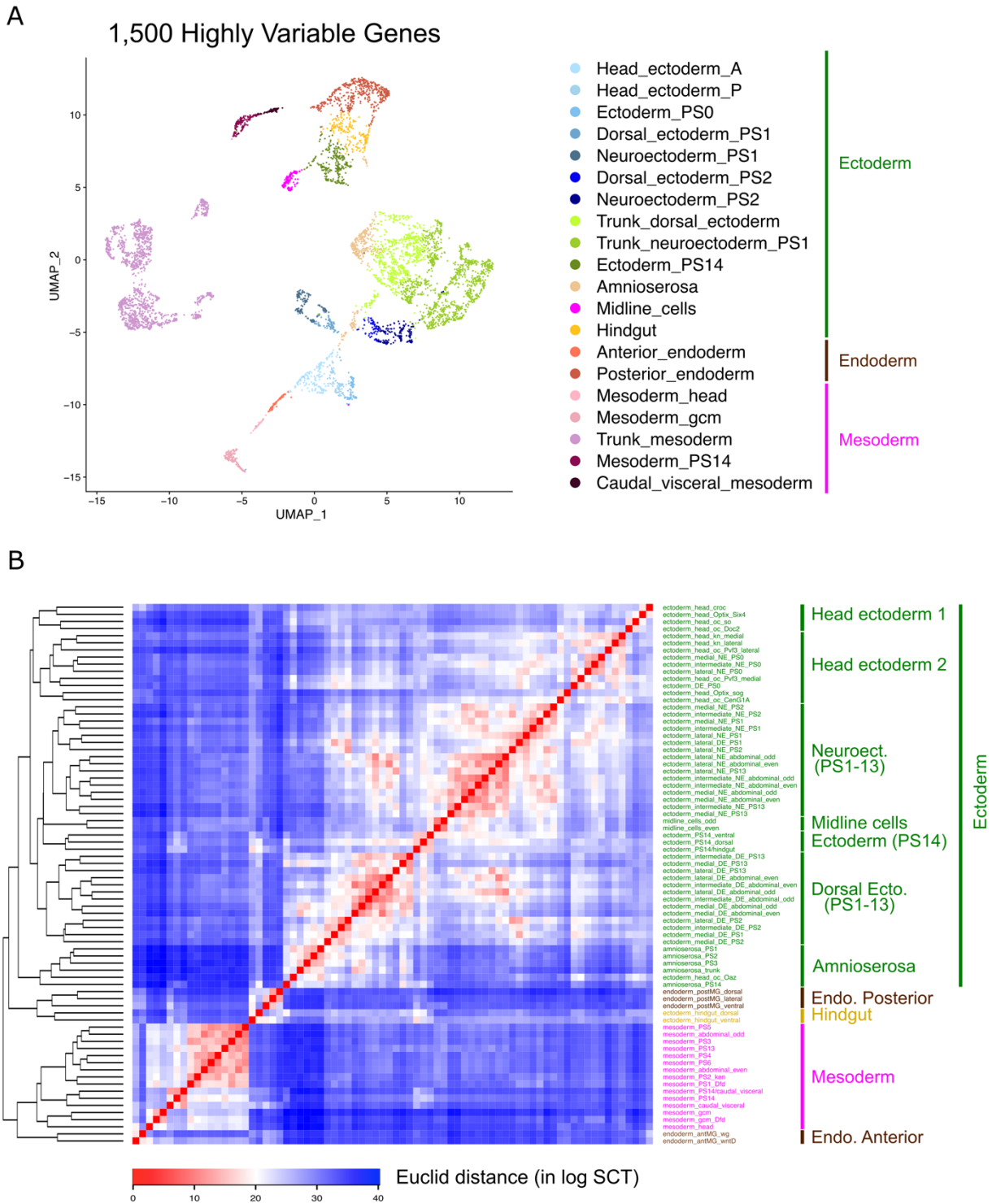


図 27

- A. 上位 1,500 HVGs を用いた UMAP プロット。各点の色はスーパークラスター (表 2) を示す。
- B. 上位 1,500 HVGs を用いた 76 サブクラスターの階層的クラスタリング。ヒートマップの色はサブクラスター間の距離をユークリッド距離で示している。サブクラスター名のラベルは予定胚葉ごとに色分けしてある。

図 28 転写因子を用いたサブクラスターの階層的クラスタリング

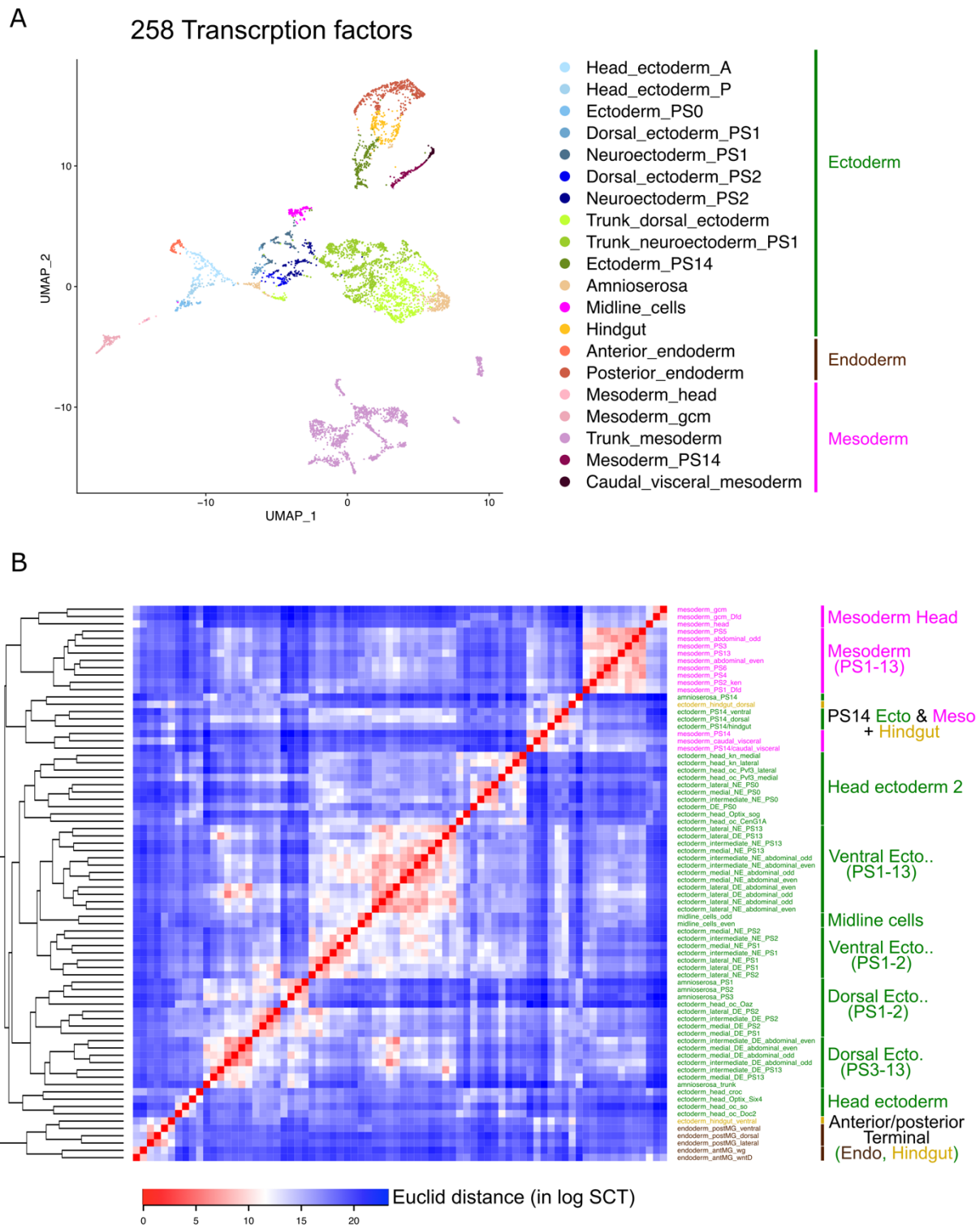


図 28

- A. 258 個の転写因子 (TFs) を用いた UMAP プロット。各点の色はスーパークラスタ (表 2) を示す。
- B. 258 個の転写因子 (TFs) を用いた 76 サブクラスタの階層的クラスタリング。ヒートマップの色はサブクラスタ間の距離をユークリッド距離で示している。サブクラスタ名のラベルは予定胚葉ごとに色分けしてある。

図 29

A. 258 個の細胞膜関連遺伝子 (PM-related genes) を用いた UMAP プロット。各点の色はスーパークラスタ (表 2) を示す。

B. 258 個の細胞膜関連遺伝子 (PM-related genes) を用いた 76 サブクラスタの階層的クラスタリング。ヒートマップの色はサブクラスタ間の距離をユークリッド距離で示している。サブクラスタ名のラベルは予定胚葉ごとに色分けしてある。

図 30 転写因子/細胞膜関連遺伝子以外によるサブクラスタの階層的クラスタリング

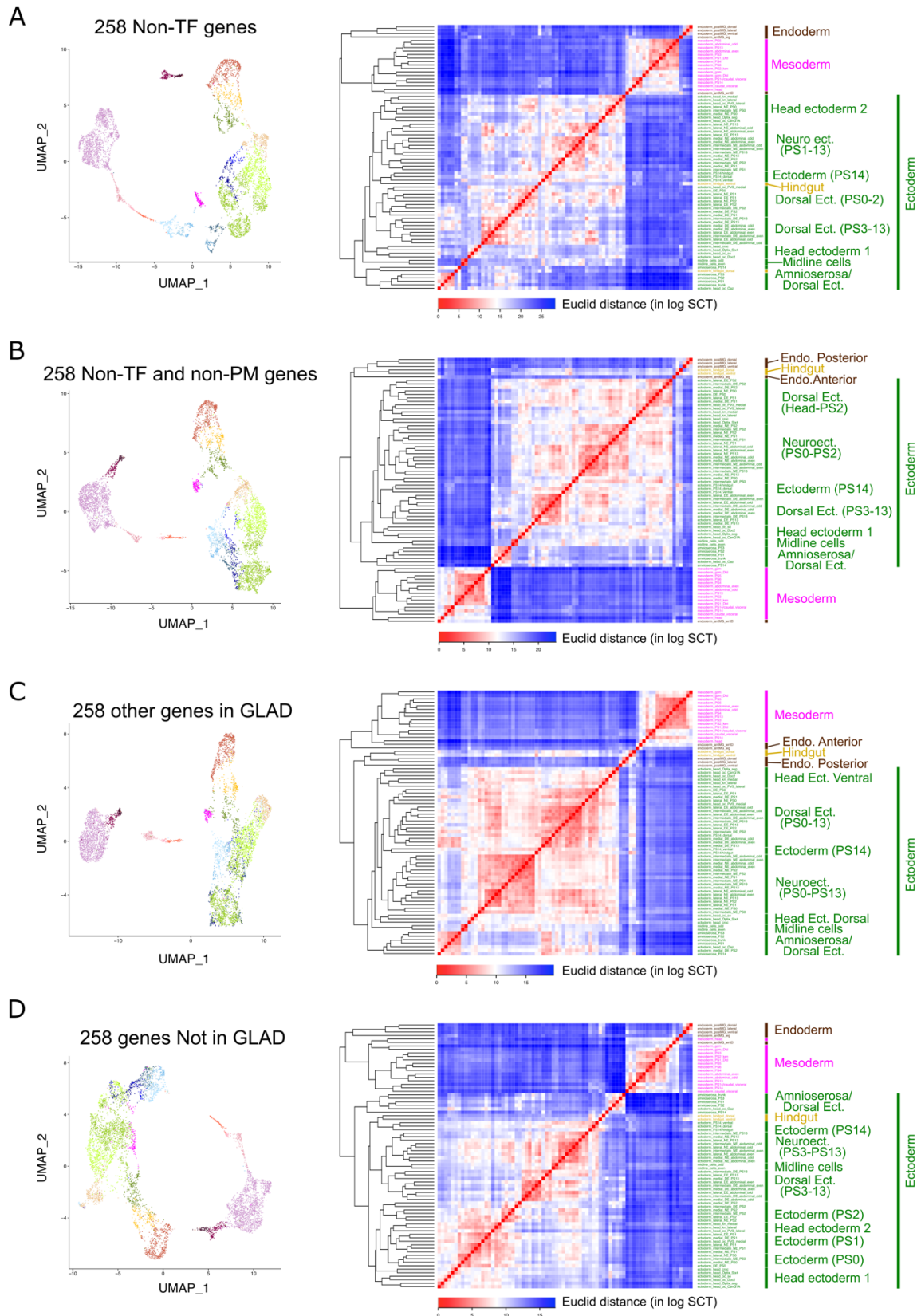


図 30

各パネルの左は UMAP プロットであり、各点の色はスーパークラスタ（表 2）を示す。右 76 サブクラスタの階層的クラスタリング結果であり、ヒートマップの色はサブクラスタ間の距離をユークリッド距離で示している。サブクラスタ名のラベルは予定胚葉ごとに色分けしてある。

- A. 258 個の非転写因子遺伝子を用いた場合。
- B. 258 個の非転写因子かつ非細胞膜関連遺伝子を用いた場合。
- C. GLAD に含まれる遺伝子のうち非転写因子かつ非細胞膜関連遺伝子（Others in GLAD）258 個を用いた場合。
- D. GLAD に含まれない遺伝子（Not in GLAD）258 個を用いた場合。

図 31 ストライプへの割り当てに用いたランドマーク遺伝子

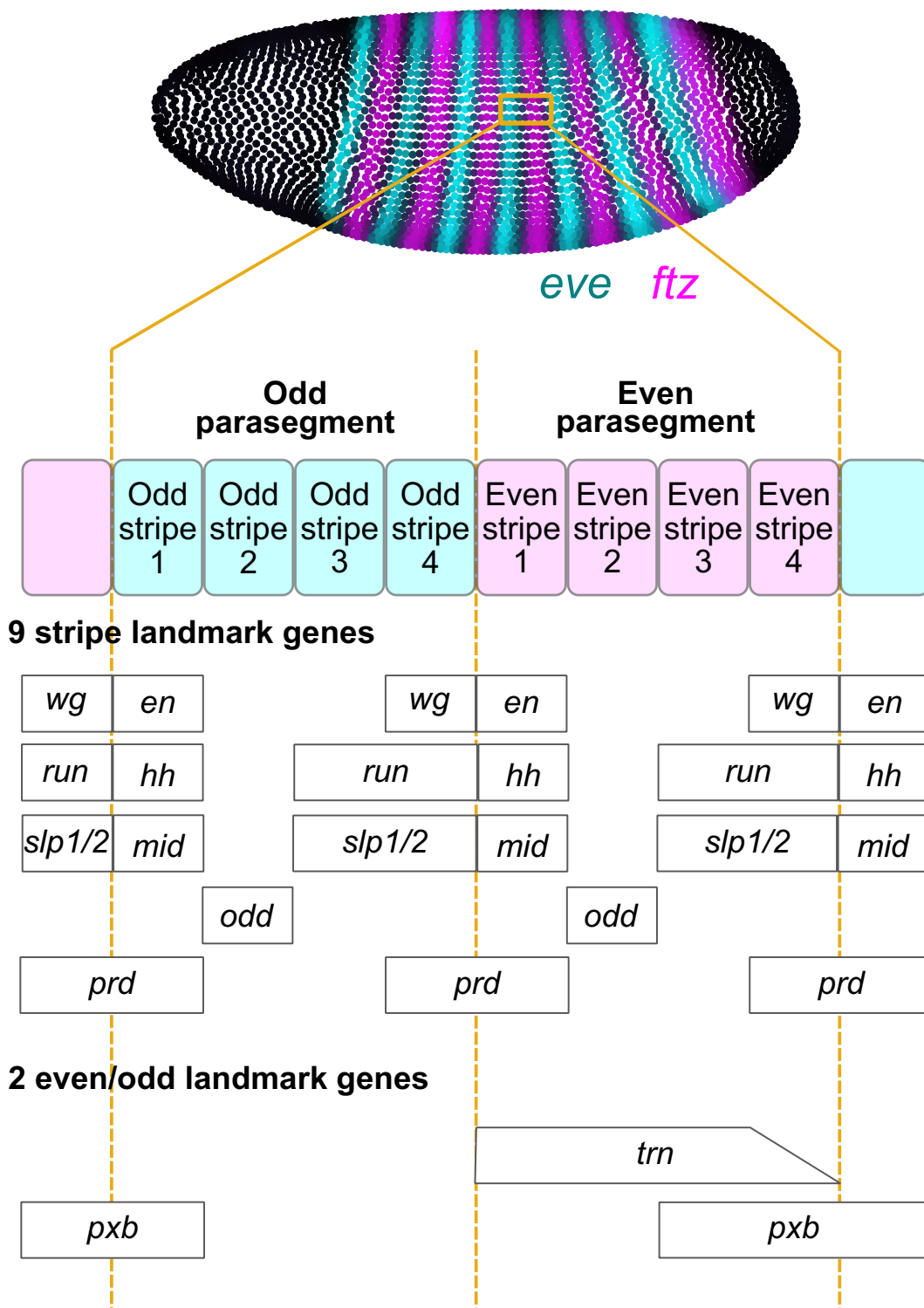


図 31

(上) BDTNP データベースから取得したペアルール遺伝子 (*eve* および *ftz*) のストライプ状発現の例。*eve* は奇数番目(even)のパラセグメント、*ftz* は偶数番目(odd)のパラセグメントで発現する。(図 5 の上と同一) (下) 9つのストライプランドマーク遺伝子および2つの even/odd パラセグメントランドマーク遺伝子の発現パターン

図 32 ストライプへの割り当て手法

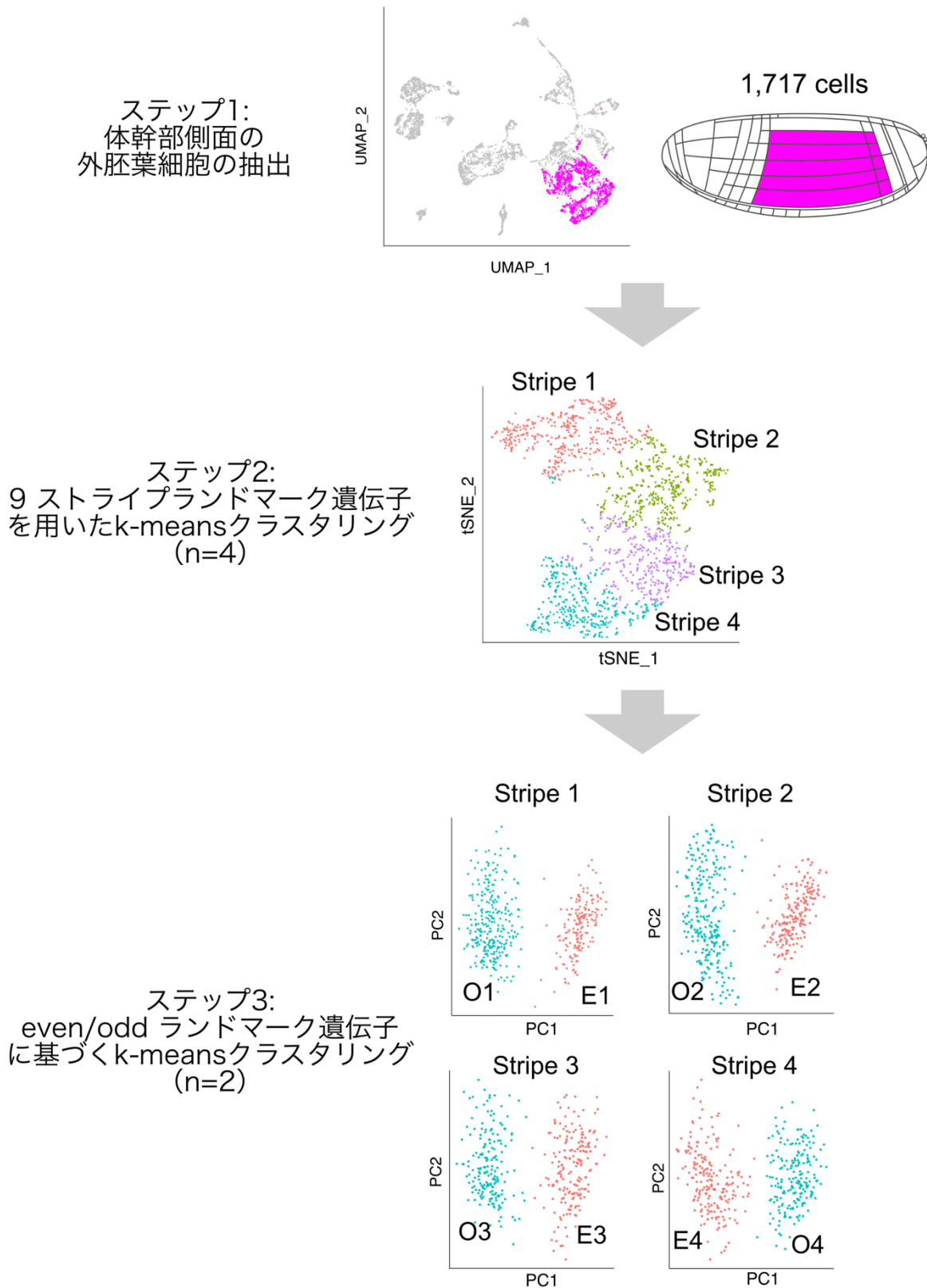


図 32

(ステップ 1) 割り当てに使用する体幹部側方外胚葉の抽出。使用した細胞を UMAP および胚の模式図上にマゼンタで示した。(ステップ 2) 9つのストライプランドマーク遺伝子を用いた t-SNE 上でのクラスタ数 4 の k-means クラスタリング。t-SNE プロット上の各点の色はどのストライプに割り当てられたかを示す。(ステップ 3) 2つの even/odd ランドマーク遺伝子の発現に基づき、ストライプごとに偶数番目、奇数番目のいずれのパラセグメントに由来する細胞かを分類した。結果を PCA プロット上の点の色で示した。

図 33 各ストライプに割り当てられた細胞数

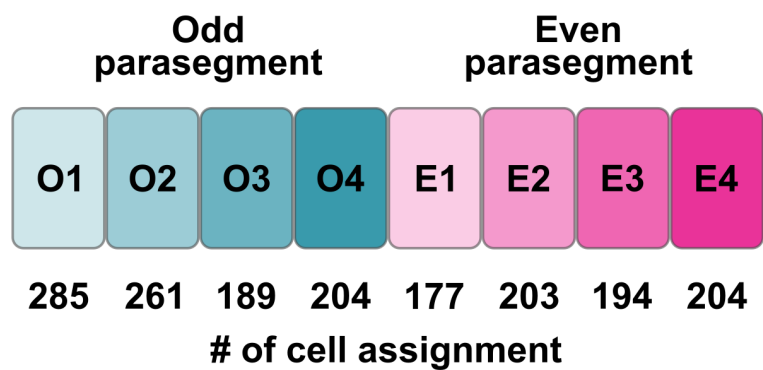


図 33
各ストライプに割り当てられた細胞数。

図 34 ストライプ状の遺伝子発現の再構成

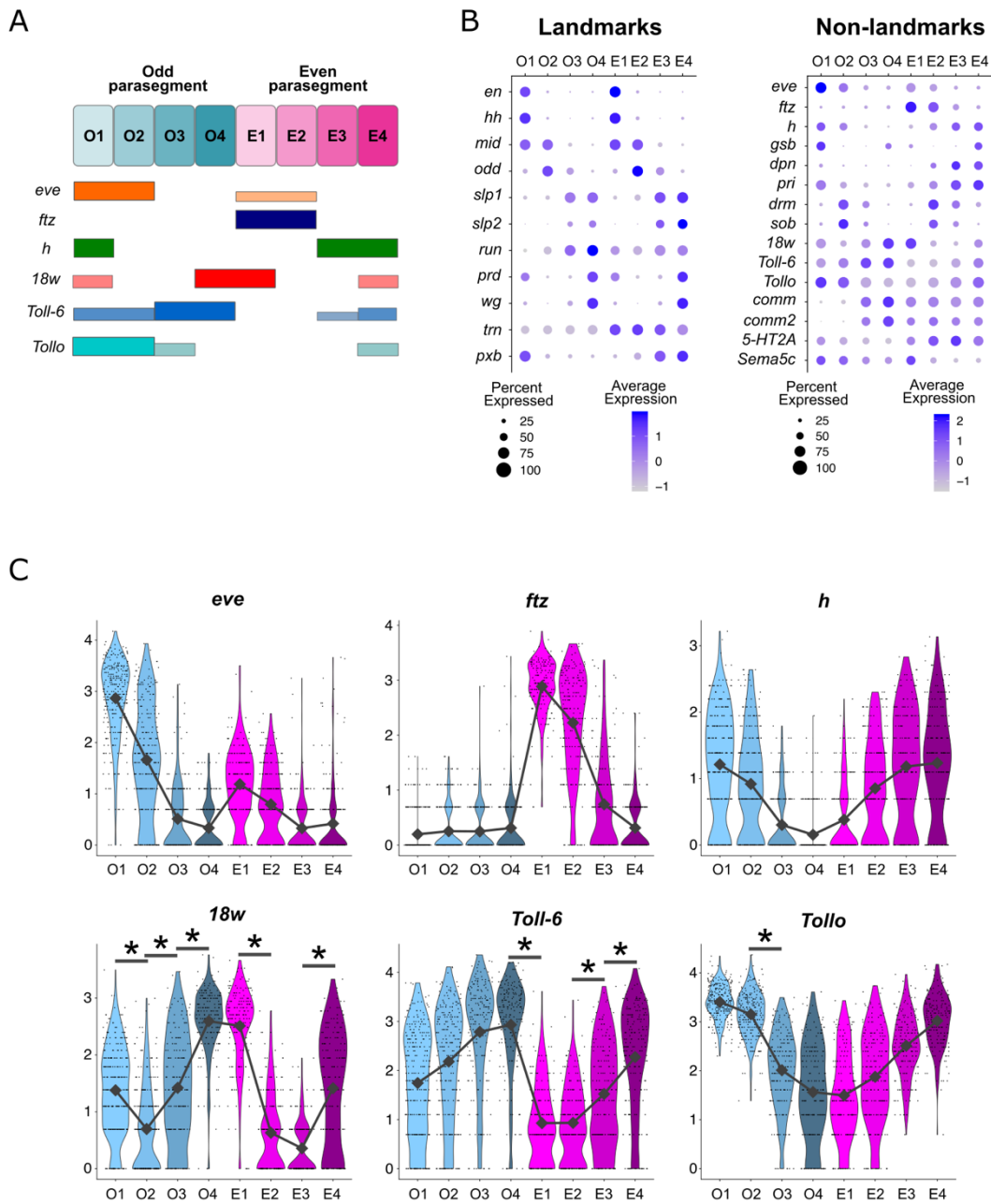


図 34

A. 報告されているストライプ状遺伝子発現 (*eve*、*ftz*、*h*、*18w*、*Toll-6*、*Tollo*) のパターン (Clark and Akam, 2016; Paré et al., 2014)。

B. ストライプへの割り当てに基づく(左)ランドマーク遺伝子の再構成パターン (右) 非ランドマーク遺伝子の再構成パターン。

C. *eve*、*ftz*、*h*、*18w*、*Toll-6*、*Tollo* の遺伝子発現のストライプ再構成結果のバイオリンプロット。縦軸は log スケールであり、灰色のラインは各ストライプにおける発現の中央値を示す。下の三つのパネルのアスタリスクは細胞境界における有意な発現の違い ($|FC| \geq 1.75$ かつ $FWER < 0.01$) を示す。

図 35 ストライプ間の遺伝子発現変動解析

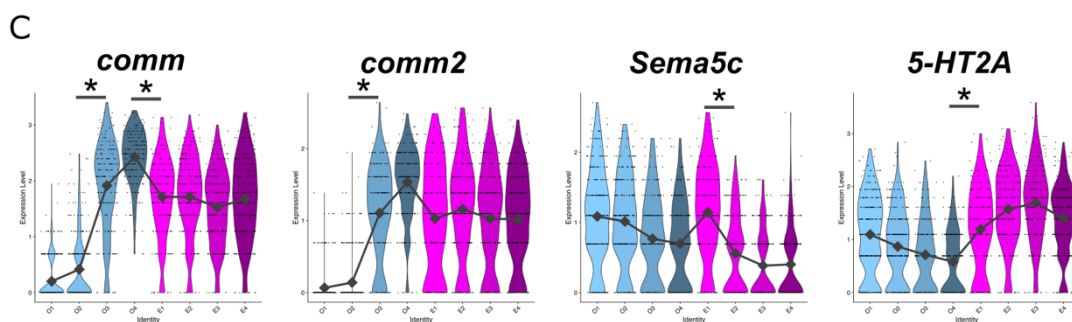
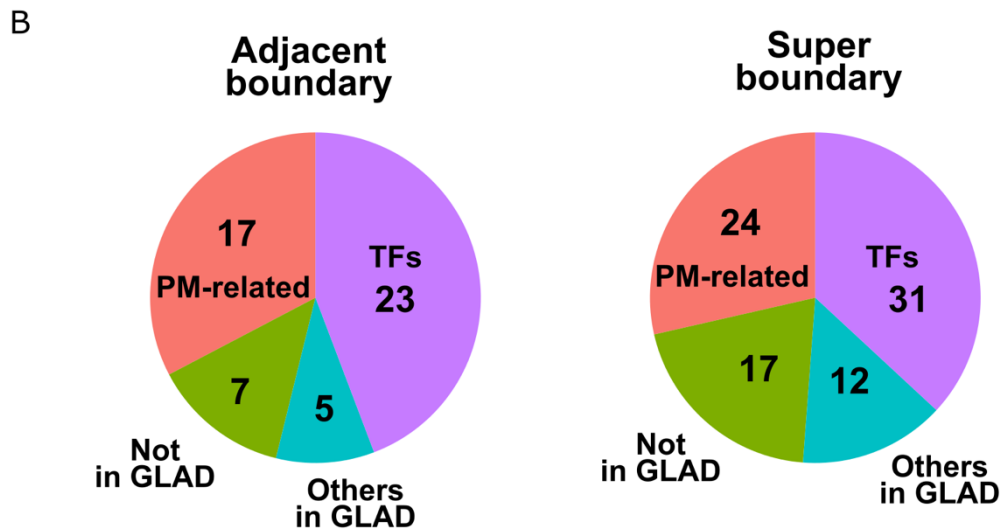
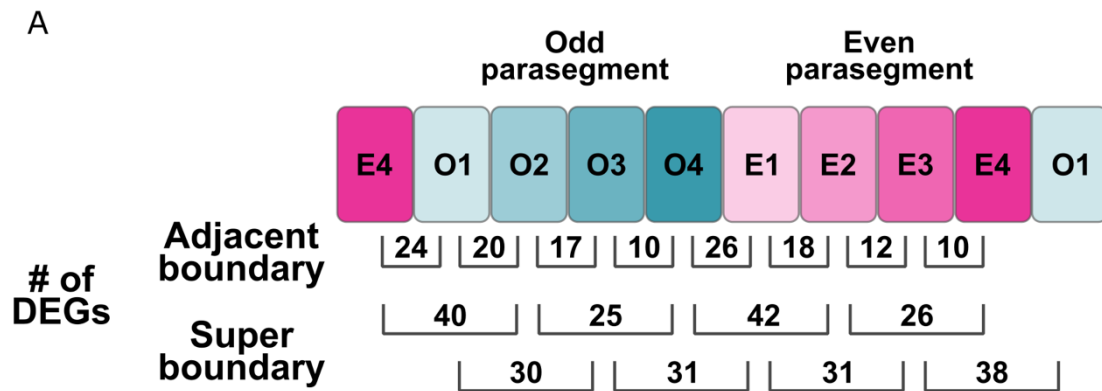


図 35

- A. 各 boundary および super-boundary において検出された DEGs の数。
- B. 各 boundary および super-boundary において検出された DEGs がどの GLAD カテゴリに属しているかの内訳を円グラフで示した。
- C. DEGs として検出された細胞膜関連遺伝子の例。発現量の再構成結果をバイオリンプロットで示した。縦軸は発現量の log スケールであり、灰色のラインは各ストライプにおける発現の中央値を示す。アスタリスクは細胞境界における有意な発現の違い ($|FC| \geq 1.75$ かつ $FWER < 0.01$) を示す。

図 36 胚前方マーカーの *bcd*-RNAi データセットにおける発現

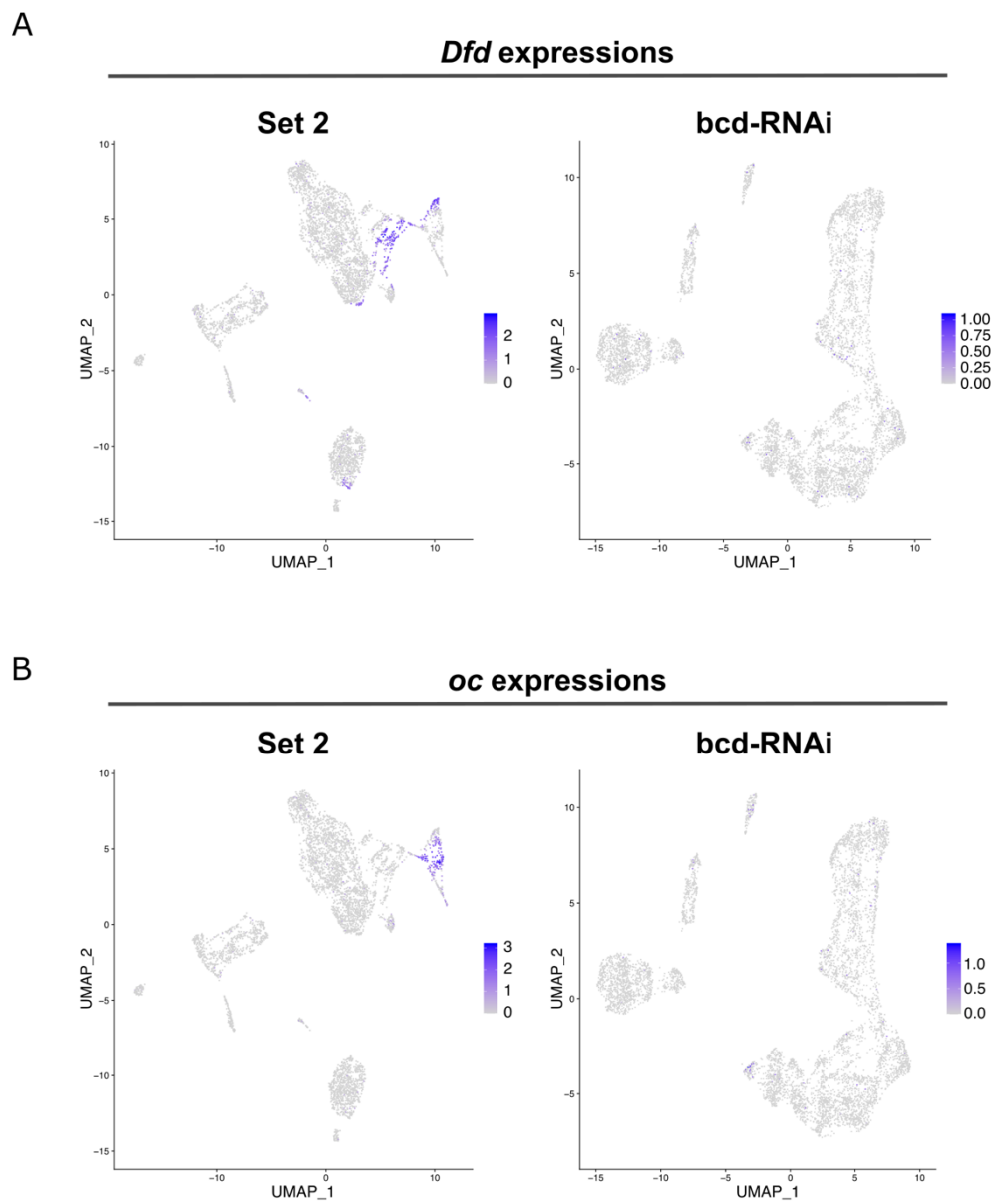


図 36

A. *Dfd* の発現を UMAP 上に示した。(左) Set 2、(右) *bcd*-RNAi

B. *oc* の発現を UMAP 上に示した。(左) Set 2、(右) *bcd*-RNAi

図 37 *bcd*-RNAi データのサブクラスタリング結果

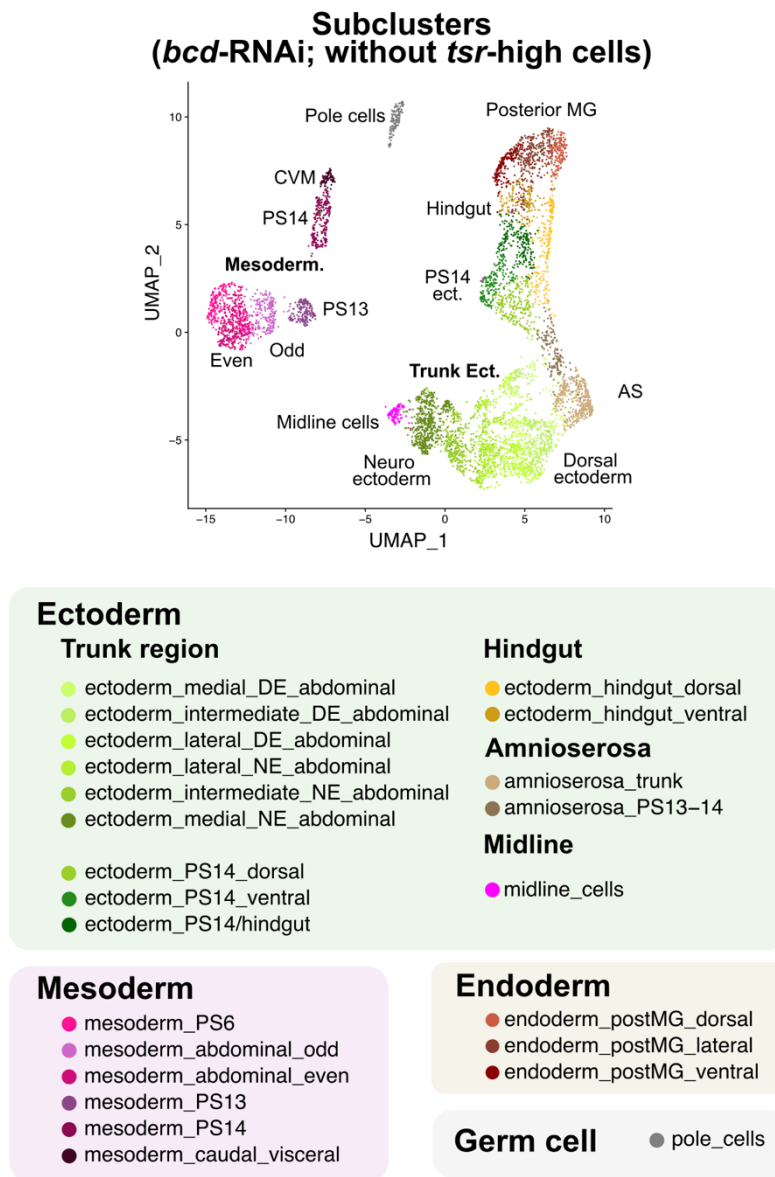


図 37

bcd-RNAi データの 24 サブクラスタの情報を UMAP 上に色で示した。

図 38 Set2 と *bcd*-RNAi データの統合解析

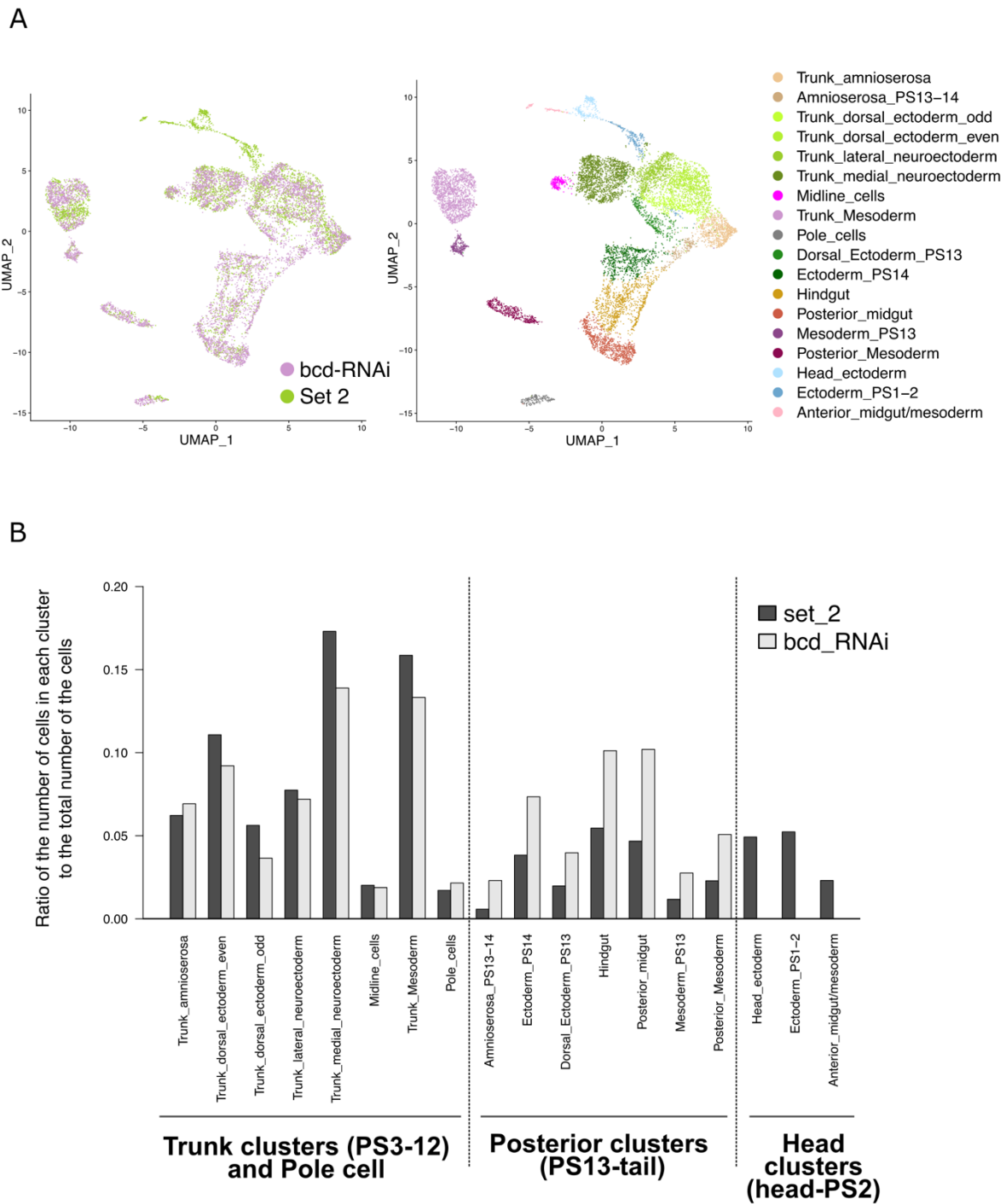


図 38

A. Set 2 と *bcd*-RNAi データを統合したデータの Seurat クラスタを UMAP 上に色で示した。
 B. データセットごとに、各 Seurat クラスタに割り当てられた細胞数の全体の細胞数に対する比率を棒グラフで示した。

図 39 それぞれの DEGs が検出されたサブクラスタ数

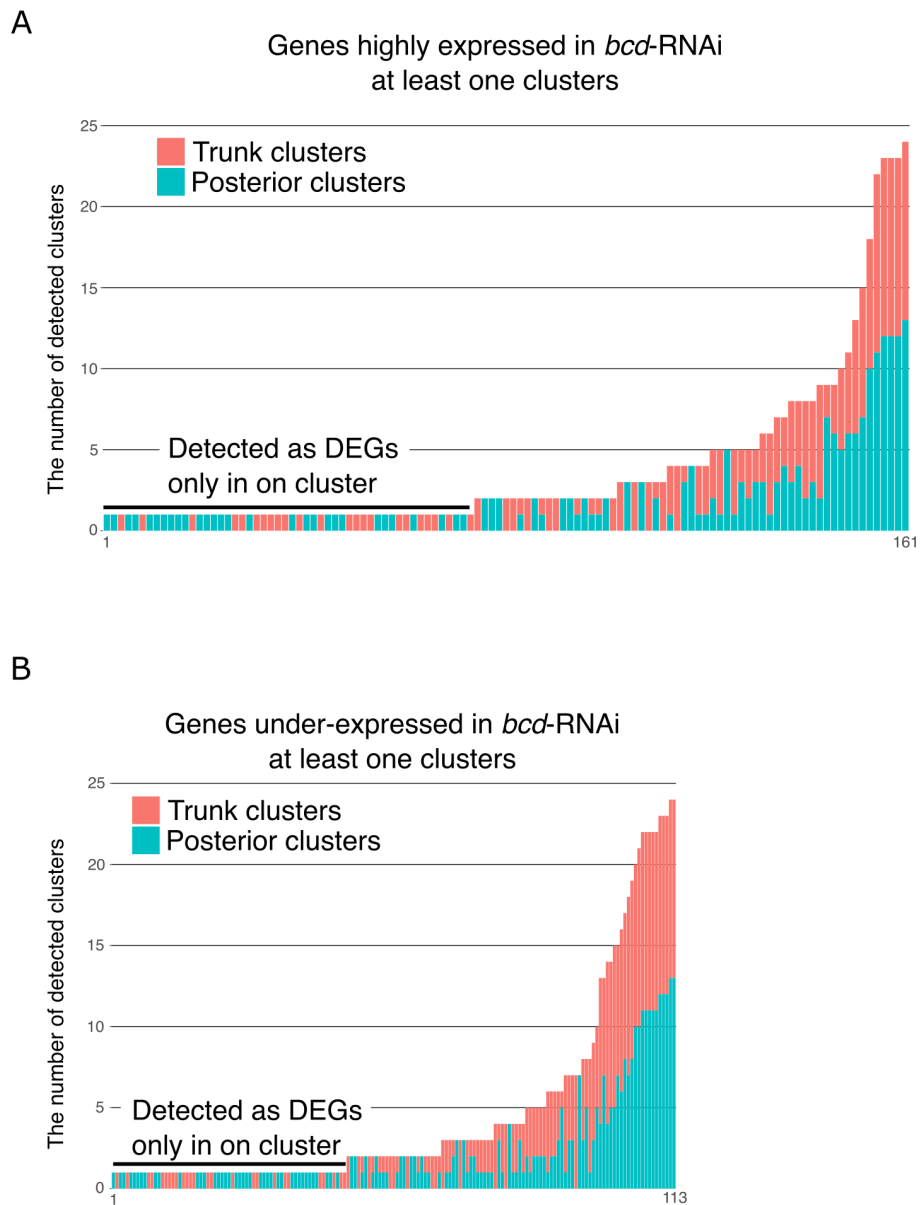


図 39

bcd-RNAi データと control (Set 2)の間で DEGs として検出された遺伝子のそれぞれについて、その遺伝子が DEGs として検出されたサブクラスタの数を棒グラフとして示した。遺伝子は検出サブクラスタ数の少ない順に並べてある。サブクラスタは体幹部（赤）と後部（青）に分類してある。(A)*bcd*-RNAi データで発現が高かった遺伝子 (B)*bcd*-RNAi データで発現の低かった遺伝子。

図 40 *Dll* の発現パターン

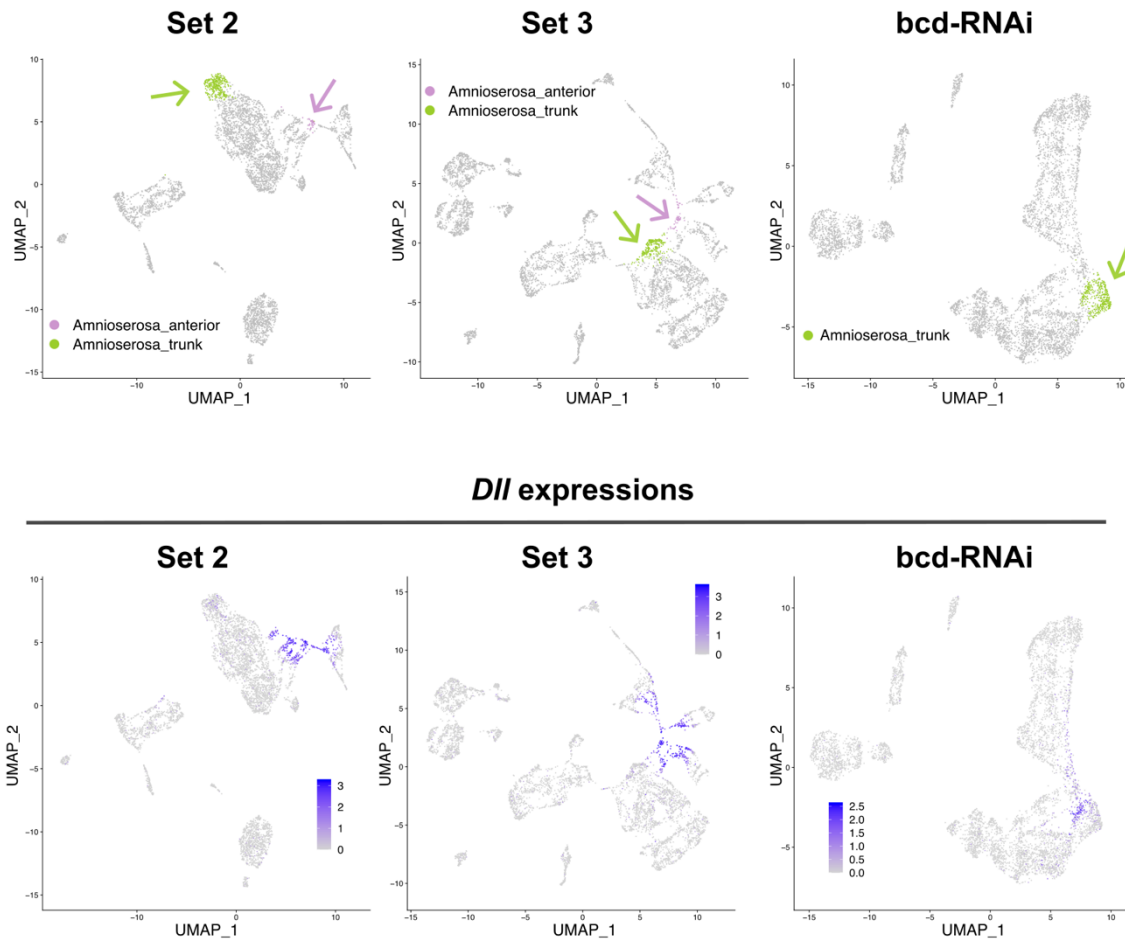


図 40

(上) UMAP プロット上で前方の羊漿膜 (マゼンタ、Amnioserosa_anterior) と体幹部の羊漿膜 (緑、Amnioserosa_trunk) に属する細胞をそれぞれ示した。左から、Set 2、Set 3、bcd-RNAi である。

(下) *Dll* の遺伝子発現を UMAP プロット上に示した。左から、Set 2、Set 3、bcd-RNAi である。

図 41 遺伝子発現の空間再構成手法 Perler

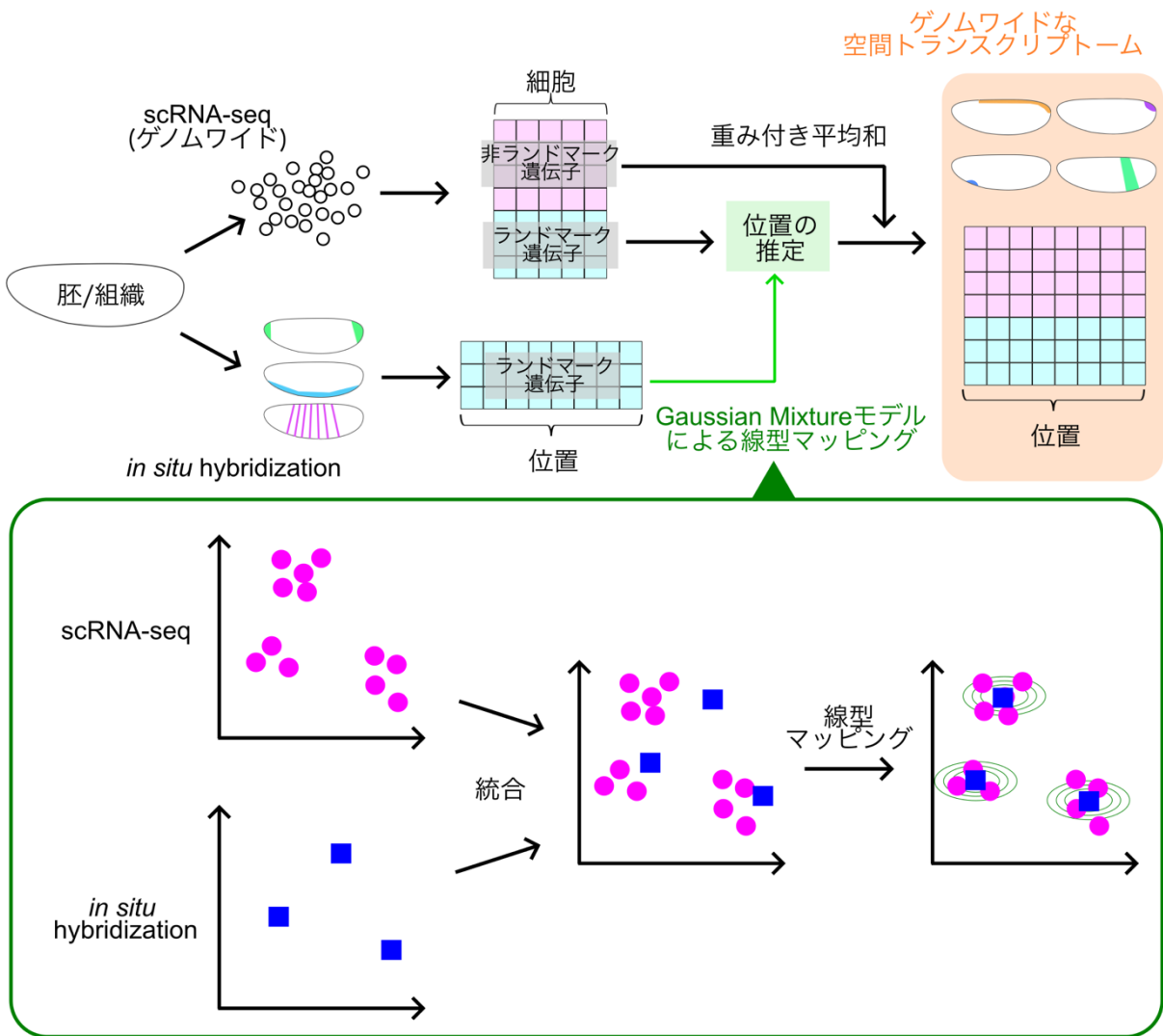


図 41

遺伝子発現の空間再構成手法 Perler の概要図。in situ hybridization によって空間発現パターンが既知となっている遺伝子発現をランドマークとし、scRNA-seq データの各細胞が胚のどの位置に由来するかを推定し、推定した位置に基づいて非ランドマーク遺伝子の scRNA-seq における発現の重み付き平均和を各位置について計算し、ゲノムワイドな空間トランスクリプトームを得る。この過程で、scRNA-seq と in situ hybridization の間の比較が必要になるが、Gaussian Mixture モデルに基づく線型マッピングによって両者のずれを補正する。Okochi et al., 2021 を参考に作成した。

図 42 Set 3 と NK-data それぞれを用いた Perler による空間再構成結果の定量比較

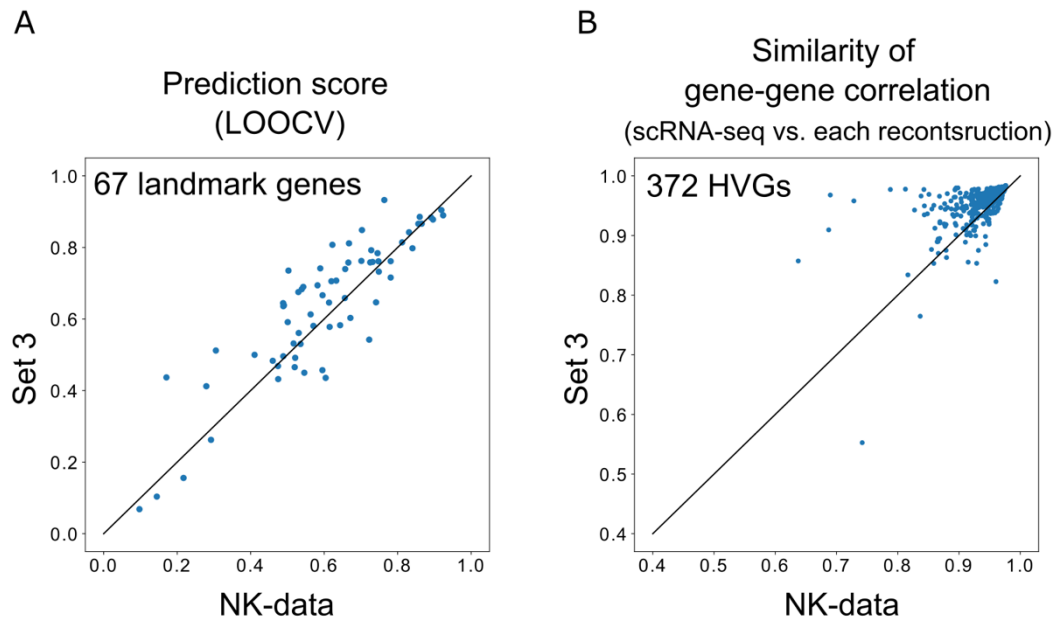


図 42

A. Leave-one-gene-out cross-validation (LOOCV) のスコアを Set 3 と NK-data それぞれを用いた Perler による空間再構成の間で比較した散布図。各点は遺伝子を表し、横軸は NK-data を用いた場合のスコア、縦軸が Set 3 を用いた場合のスコアである。各スコアは対応する遺伝子に関する空間再構成結果とレファレンスの空間パターンのピアソン相関係数である。67 のランドマーク遺伝子についての比較を示してある。

B. 遺伝子発現の相関構造がどの程度維持されているかを Set 3 と NK-data それぞれを用いた Perler による空間再構成の間で比較した散布図。各点は遺伝子を表し、横軸は NK-data を用いた場合のスコア、縦軸が Set 3 を用いた場合のスコアである。スコアの計算方法については「材料と手法」を参照。Set 3 と NK-data 双方に共通する HVGs である 372 遺伝子についての比較を示してある。

図 43 Perler による空間再構成結果の例

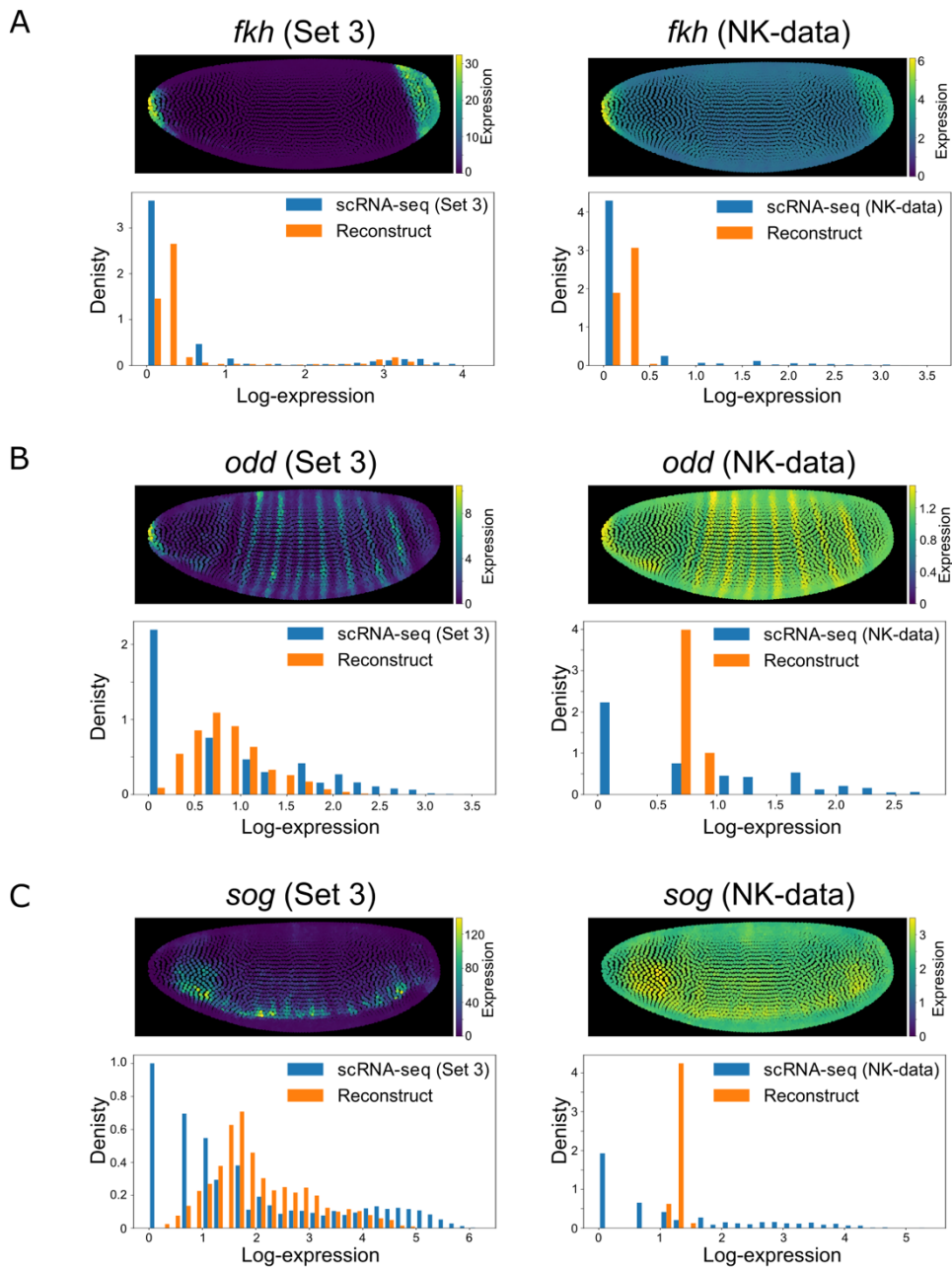


図 43

それぞれのパネルの上のプロットは Perler による空間再構成結果の例 (A: *fkh*、B: *odd*、C: *sog*) を示す。カラーマップは発現量に対して線型で、min-max スケーリングされている。下のプロットは遺伝子発現の分布密度のヒストグラムであり、青の棒が元の scRNA-seq における発現の分布、オレンジの棒が再構成された発現の分布である。各パネルの左が Set 3 を用いた場合、右が NK-data を用いた場合である。

図 44 Perler によるストライプ状遺伝子発現の再構成結果

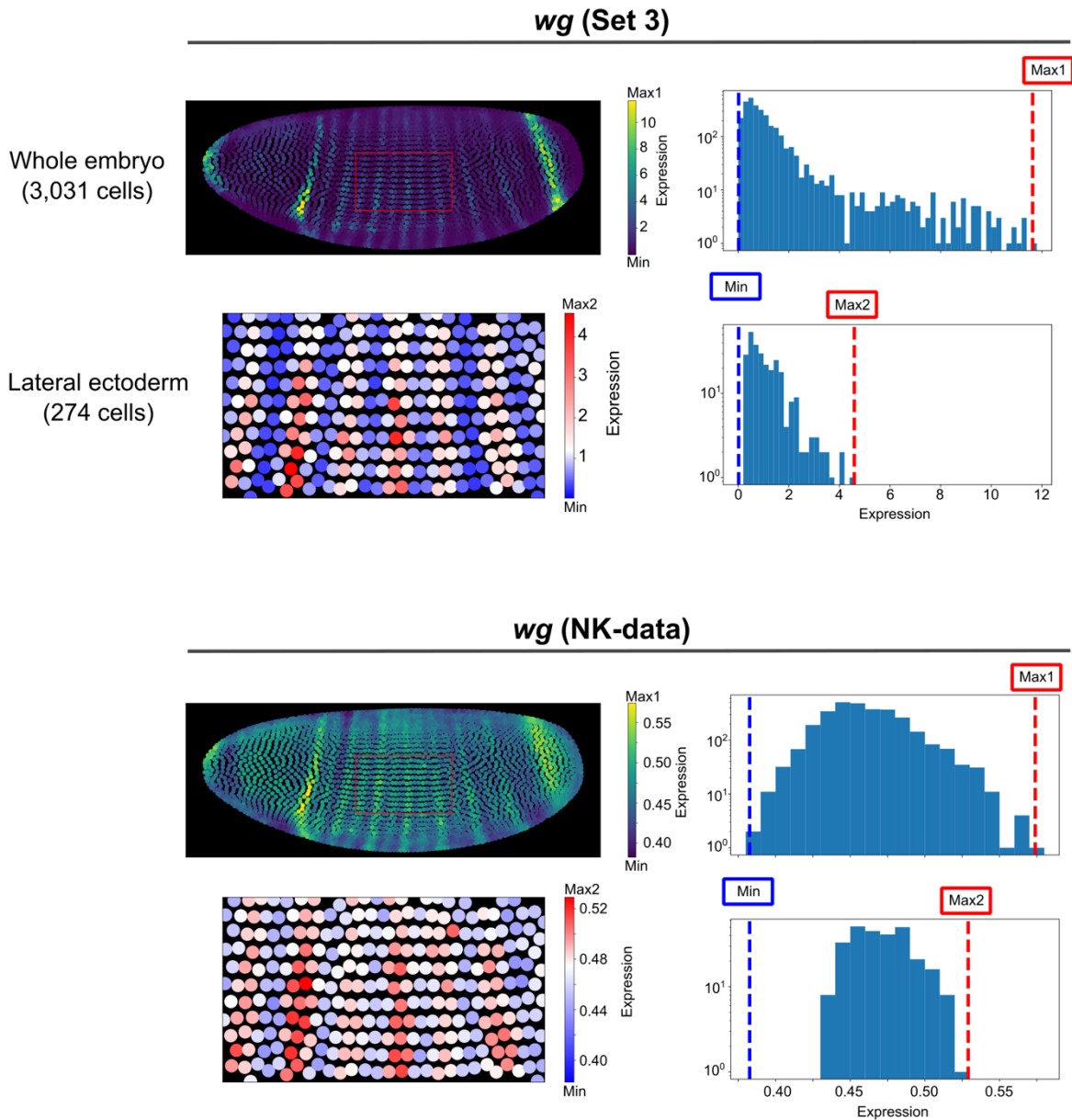


図 44

(上) Set 3 を用いた場合 (下) NK-data を用いた場合の *wg* の遺伝子発現再構成結果。それぞれのパネルについて、左上のプロットは再構成された遺伝子発現の胚全体におけるパターンを示し、赤枠で囲った部分の拡大図が左下のプロットである。右上、左下のプロットはそれぞれ左上、右下のプロットにおける遺伝子発現の分布をヒストグラムで示したものである。遺伝子発現は線型スケールで示されており、胚全体および拡大図における最大値はそれぞれ Max1、Max2 で右のプロットに示してある。胚全体の遺伝子発現の最小値は Min として左のプロットに示してある。

図 45 *C15* および *egr* の遺伝子発現の空間再構成結果

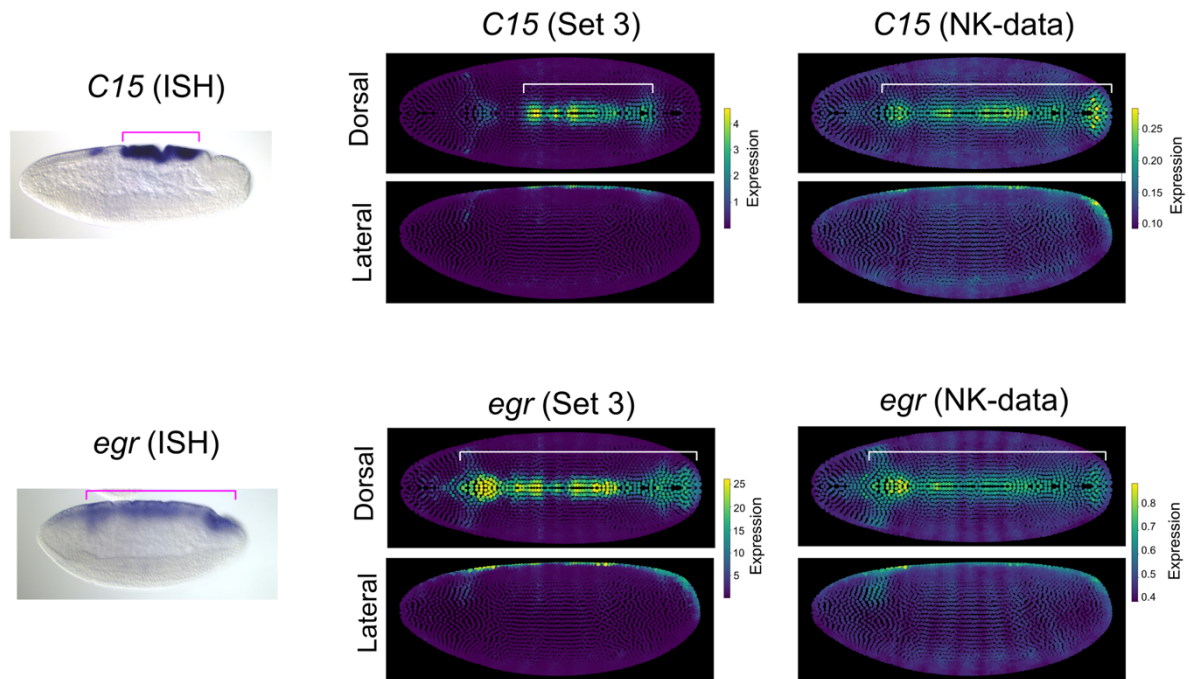


図 45

(左) *C15* および *egr* の *in situ* hybridization を胚の側面から見たもの。画像は Berkeley Drosophila Genome Project *in situ* データベースから引用した(<https://insitu.fruitfly.org/>、Hammonds et al., 2013; Tomancak et al., 2007, 2002)。(中央) Set 3 を用いた Perler による *C15* および *egr* の発現の空間再構成結果の側面から見た図(Lateral)と背側から見た図 (Dorsal)。(右) NK-data を用いた Perler による *C15* および *egr* の発現の空間再構成結果の胚の側面から見た図(Lateral)と背側から見た図 (Dorsal)。赤と白の線は背側の正中線に沿った遺伝子発現の範囲を示す。

図 46 Perler と NovoSpaRc による空間再構成結果の定量比較

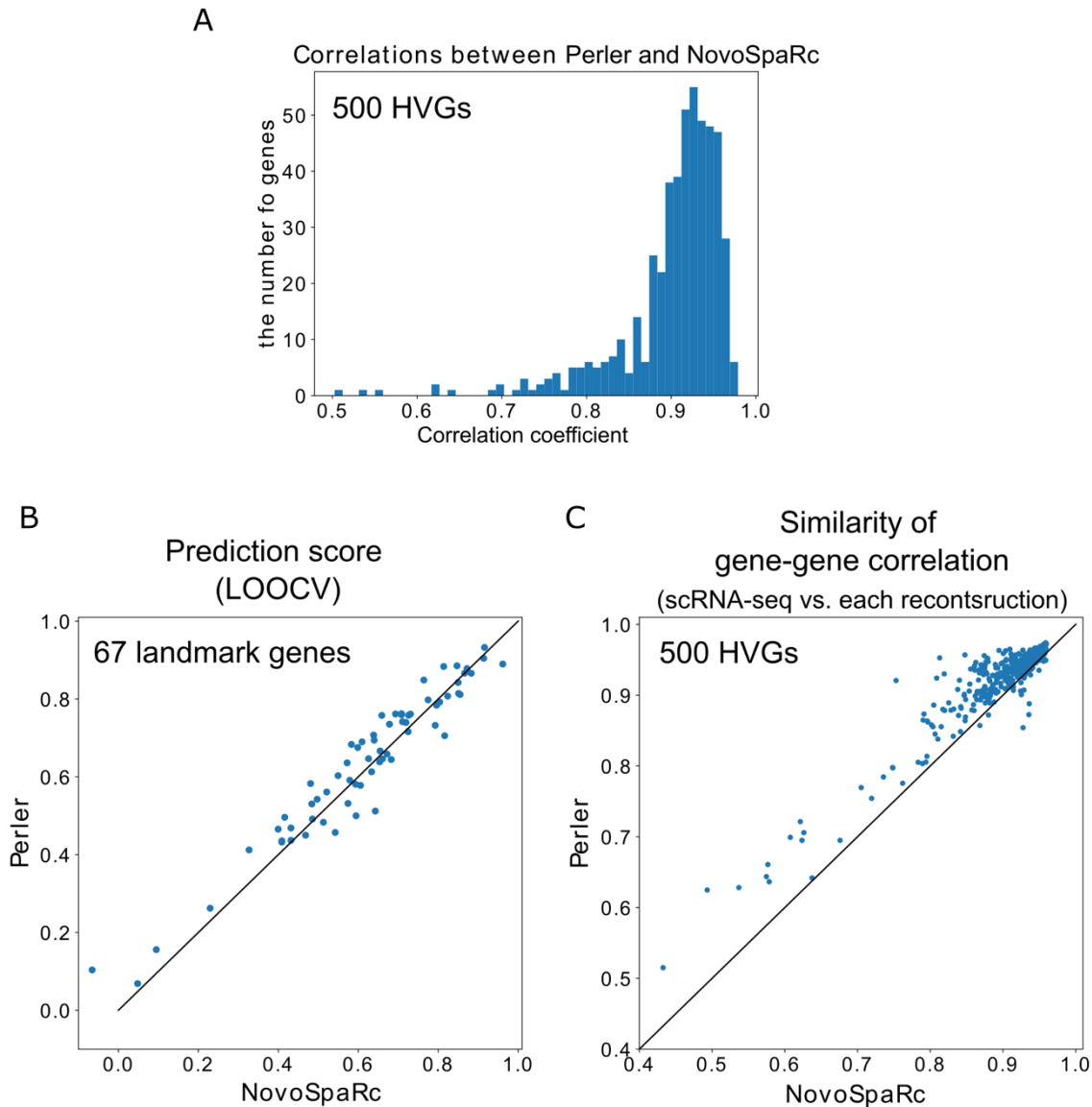


図 46

A. Perler と NovoSpaRc による、Set 3 に基づく空間再構成結果のピアソン相関係数のヒストグラム。上位 500HVGs に関する結果を示した。

B. Perler と NovoSpaRc による、Set 3 に基づく空間再構成の LOOCV スコアの比較。各点が遺伝子を示し、横軸が NovoSpaRc のスコア、縦軸が Perler のスコアを示す。各スコアは対応する遺伝子に関する空間再構成結果とレファレンスの空間パターンのピアソン相関係数である。67 のランドマーク遺伝子についての比較を示してある。

C. 遺伝子発現の相関構造がどの程度維持されているかを Set 3 を用いた Perler による空間再構成と NovoSpaRc による空間再構成の間で比較した散布図。各点は遺伝子を表し、横軸は NovoSpaRc のスコア、縦軸が Perler のスコアである。スコアの計算方法については「材料と手法」を参照。Set 3 と NK-data 双方に共通する HVGs である 372 遺伝子についての比較を示してある。

図 47 Perler と NovoSpaRc の空間再構成結果の定性比較

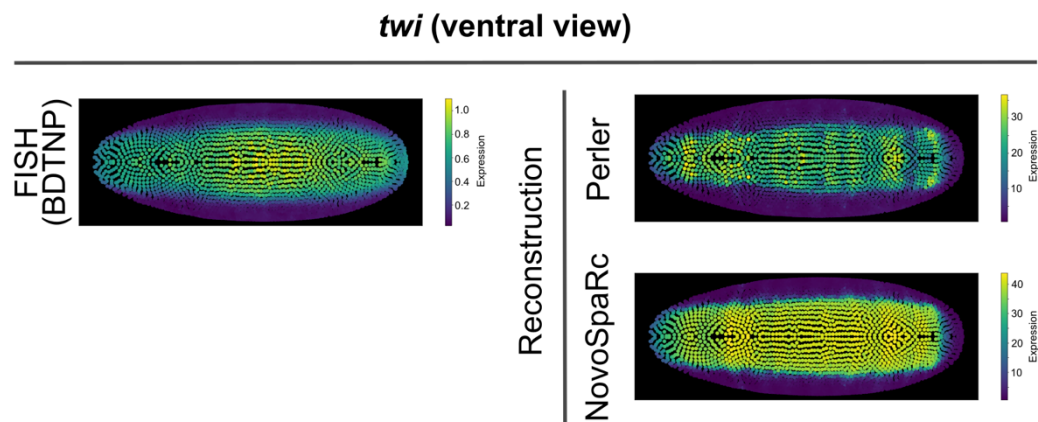


図 47

(左) レファレンス (BDTNP データベース) における *twi* の遺伝子発現。(右) 空間再構成における *twi* の遺伝子発現。上が Perler による再構成、下が NovoSpaRc による再構成である。全てのプロットにおいて、遺伝子発現のスケールは線型で、min-max スケーリングで表示している。また、胚を腹側から見た図である。

図 48 背腹軸の空間再構成の不正確性

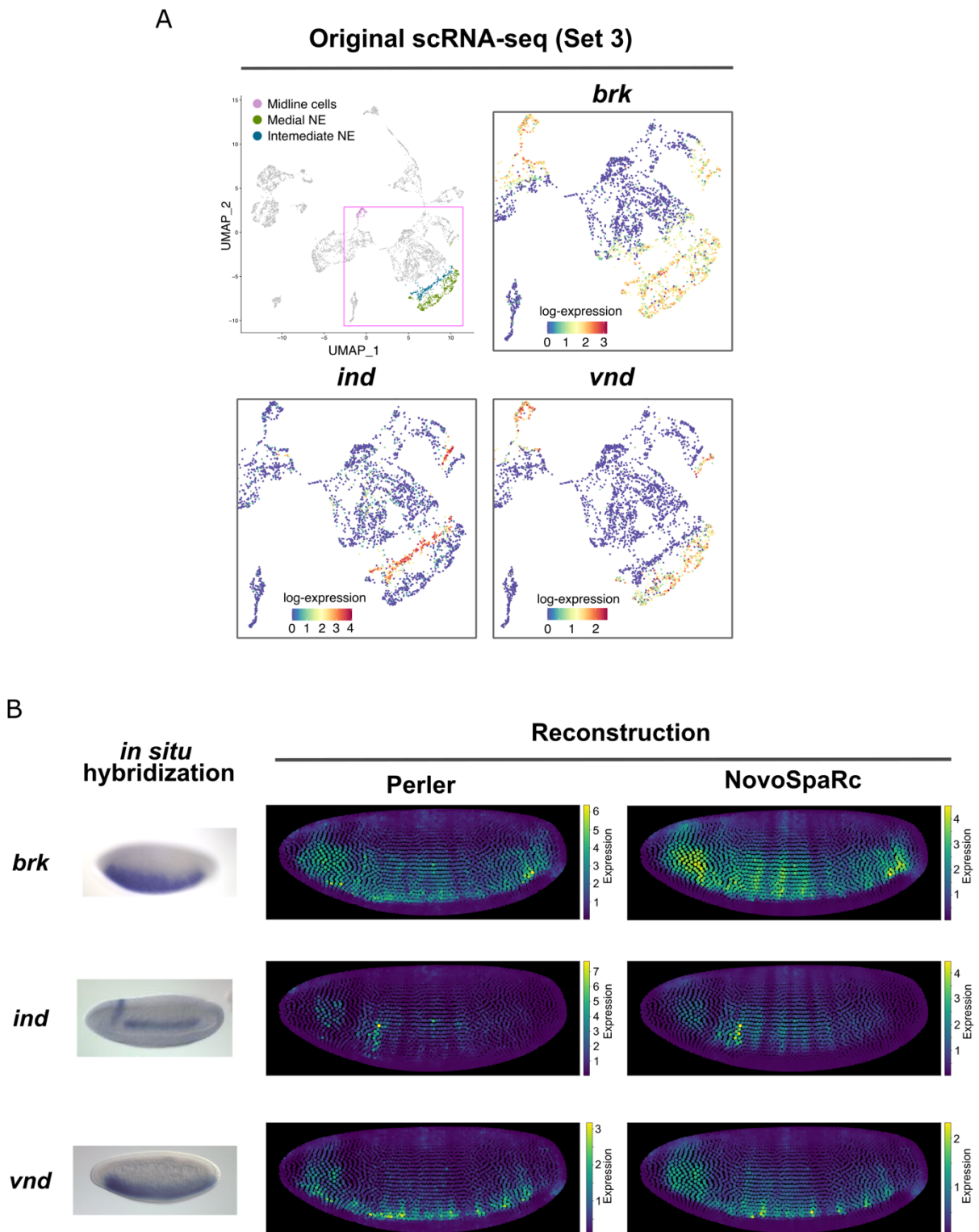


図 48

A. 左上のパネルは図 49 B でプロットに用いた細胞群を Set 3 の UMAP 上に示した。他のパネルは Set 3 における *brk* (右上)、*ind* (左下)、*vnd* (右下) の発現を Set 3 の UMAP 上で示した。

B. *brk* (上)、*ind* (中央)、*vnd* (下) の *in situ* hybridization と空間再構成結果の側方から見た図。左が BDGP(<https://insitu.fruitfly.org/>、Hammonds et al., 2013; Tomancak et al., 2007, 2002) から引用した *in situ* hybridization 画像、中央が Perler による再構成、右が NovoSpaRc による再構成を示す。遺伝子発現のスケールは線型であり、min-max スケーリングされている。

図 49 空間再構成における *vnd* と *ind* の発現の分離不全

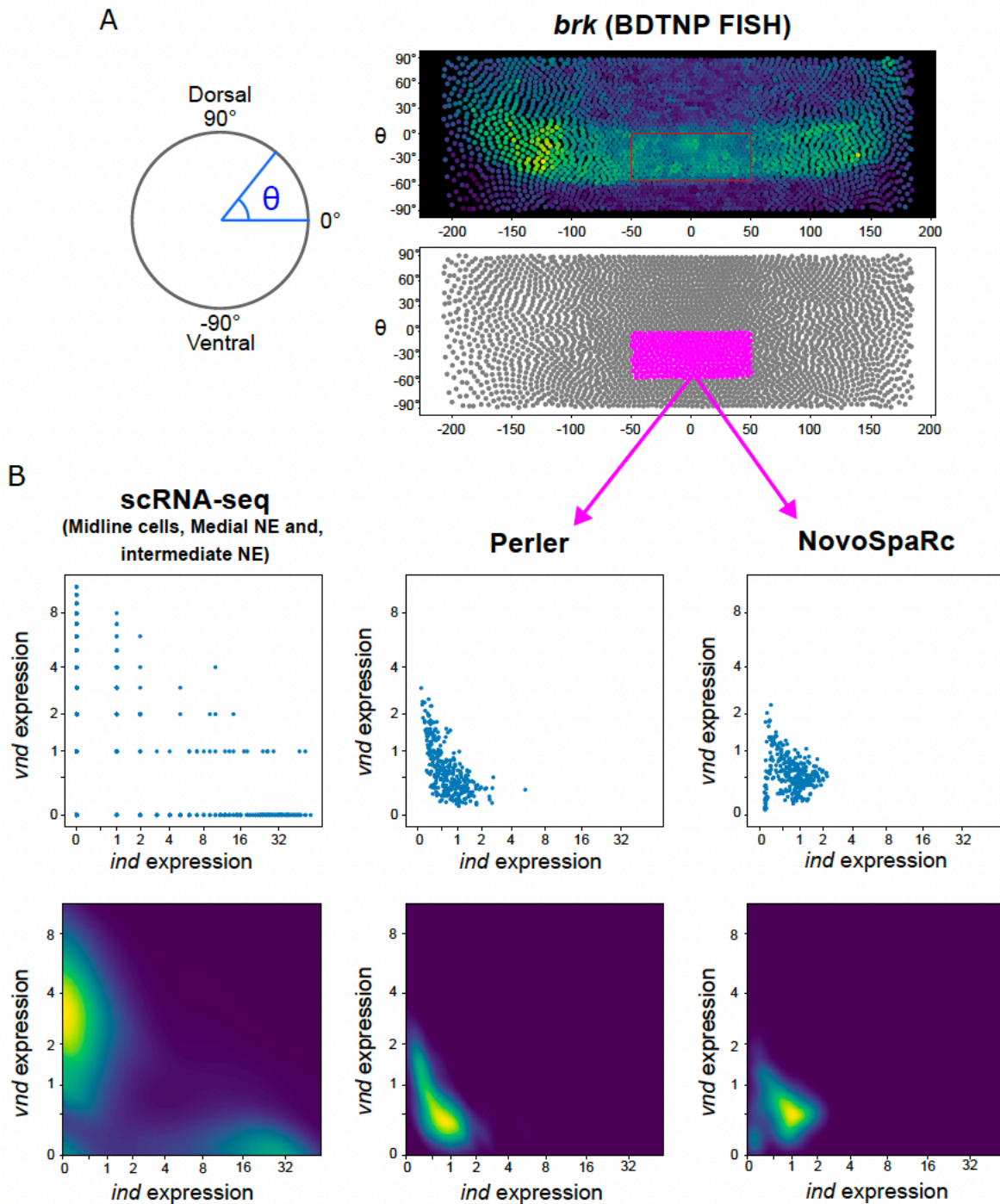


図 49

A. (左) 背腹軸における細胞位置の定義の模式図 (右上) レファレンスにおける *brk* の発現。胚を平面化して示してある。(右下) 右上のパネルで赤い四角で囲んだ範囲の細胞をマ

ゼンタで示した。B の空間再構成についてのプロットではこれらの細胞のみを用いている。B. (上) *ind* (横軸) と *vnd* (縦軸) の発現の散布図。左から scRNA-seq (Set 3)、Perler による空間再構成、NovoSpaRc による空間再構成を示す。(下) *ind* (横軸) と *vnd* (縦軸) の発現の密度分布プロット。左から scRNA-seq (Set 3)、Perler による空間再構成、NovoSpaRc による空間再構成を示す。カラーマップは min-max スケーリングされている。scRNA-seq についてのプロットでは図 48A で着色して示した細胞のみを用いた。

表

表 1 各データセットの要約

	Datasets	protocol	Number of Cells	Median total UMI counts per cells	Total number of detected genes	Median number of detected genes per cells
before removing tsr-high cell removal and subclustering	Set 1 (batch1)	trypsin dissociation and C1HT	229	153,324	11,950	4,632
	Set 1 (batch2)	trypsin dissociation and C1HT	297	145,152	12,578	4,324
	Set 1 (batch3)	trypsin dissociation and C1HT	409	155,883	13,350	4,510
	Set 1 (batch4)	trypsin dissociation and C1HT	308	147,736	12,991	4,442
	Set 1 (merged)	trypsin dissociation and C1HT	1,243	150,429	15,059	4,480
	Set 2	trypsin dissociation and 10x	7,314	22,505.5	13,606	3,222
	Set 3	CAP dissociation and 10x	6,180	37,610	13,349	4,053
	NK-data	Mechanical dissociation and Drop-seq	1,592	22,207.5	12,686	3,033
bcd-RNAi	trypsin dissociation and 10x	9,223	17,589	13,670	2,960	
after removing tsr-high cell	Set 1 (batch1)	trypsin dissociation and C1HT	191	153,849	11,690	4,650
	Set 1 (batch2)	trypsin dissociation and C1HT	262	146,220	12,361	4,339
	Set 1 (batch3)	trypsin dissociation and C1HT	330	156,913	13,057	4,537
	Set 1 (batch4)	trypsin dissociation and C1HT	235	150,528	12,479	4,459
	Set 1 (merged)	trypsin dissociation and C1HT	1,018	151,665	14,785	4,493.5
	Set 2	trypsin dissociation and 10x	4,897	20,346	13,222	3,085
	Set 3	CAP dissociation and 10x	-	-	-	-
	NK-data	Mechanical dissociation and Drop-seq	-	-	-	-
	bcd-RNAi	trypsin dissociation and 10x	6,473	16,111	13,353	2,825
after removing potential doublet cells during subclustering	Set 1 (batch1)	trypsin dissociation and C1HT	no potential doublet cells were found			
	Set 1 (batch2)	trypsin dissociation and C1HT				
	Set 1 (batch3)	trypsin dissociation and C1HT				
	Set 1 (batch4)	trypsin dissociation and C1HT				
	Set 1 (merged)	trypsin dissociation and C1HT				
	Set 2	trypsin dissociation and 10x	4,855	20,336	13,214	3,084
	Set 3	CAP dissociation and 10x	6,118	37,521	13,335	4,048
	NK-data	Mechanical dissociation and Drop-seq	1,476	22,334.5	12,609	3,048.5
	bcd-RNAi	trypsin dissociation and 10x	6,168	15,926.5	13,311	2,808

表 2 Set 3 のサブクラスタリング結果

Seurat Clusters (markers)	Subclusters	Germ layer	Super Cluster	positive markers	negative markers	for stripe reconstruction
8: Head ectoderm (oc) 14: Head ectoderm_kn (kn)	ectoderm_head_croc	Ectoderm	Head_ectoderm_oc	croc, Oplx, fkh	hkb	
	ectoderm_head_Oplx_Six4	Ectoderm	Head_ectoderm_oc	oc, Oplx, Six4		
	ectoderm_head_Oplx_sog	Ectoderm	Head_ectoderm_oc	oc, Oplx, sog, SoxN		
	ectoderm_head_oc_CenG1A	Ectoderm	Head_ectoderm_oc	oc, CenG1A, sog, SoxN	Oplx	
	ectoderm_head_oc_Doc2	Ectoderm	Head_ectoderm_oc	oc, Doc2, grn	so	
	ectoderm_head_oc_so	Ectoderm	Head_ectoderm_oc	oc, so, Doc2, grn		
	ectoderm_head_oc_Pvf3_medial	Ectoderm	Head_ectoderm_oc	oc, Pvf3, grn	so, Oaz	
	ectoderm_head_oc_Pvf3_lateral	Ectoderm	Head_ectoderm_oc	oc, Pvf3, Dr	so, Oaz	
	ectoderm_head_oc_Oaz	Ectoderm	Head_ectoderm_oc	oc, Oaz, zen, grn		
	ectoderm_head_kn_lateral	Ectoderm	Head_ectoderm_kn	kn, Atx-1	Dfd	
	ectoderm_head_kn_medial	Ectoderm	Head_ectoderm_kn	kn, sog	Dfd	
	ectoderm_medial_NE_PS0	Ectoderm	Ectoderm_PS0	kn, Dfd, SoxN, sog, vnd		
	ectoderm_intermediate_NE_PS0	Ectoderm	Ectoderm_PS0	kn, Dfd, SoxN, sog, ind		
	ectoderm_lateral_NE_PS0	Ectoderm	Ectoderm_PS0	kn, Dfd, SoxN, Atx-1	grn	
	ectoderm_DE_PS0	Ectoderm	Ectoderm_PS0	kn, Dfd, grn		
	amnioserosa_PS1	Amnioserosa	Amnioserosa	Dfd, zen, peb, salm, eve		
amnioserosa_PS2	Amnioserosa	Amnioserosa	Scr, zen, peb, salm, ftz			
amnioserosa_PS3	Amnioserosa	Amnioserosa	zen, peb, salm, eve	Dfd, Scr, trn		
11: Ectoderm_PS1 (Dfd)	ectoderm_medial_DE_PS1	Ectoderm	Dorsal_Ectoderm_PS1	Dfd, dpp, tup, Doc2		
	ectoderm_lateral_DE_PS1	Ectoderm	Dorsal_Ectoderm_PS1	Dfd, dpp, Atx-1		
	ectoderm_lateral_NE_PS1	Ectoderm	Neuroectoderm_PS1	Dfd, SoxN, Dr, Atx-1		
	ectoderm_intermediate_NE_PS1	Ectoderm	Neuroectoderm_PS1	Dfd, SoxN, sog, ind		
	ectoderm_medial_NE_PS1	Ectoderm	Neuroectoderm_PS1	Dfd, SoxN, sog, vnd		
2: Drosal Ectoderm odd (Antp, dpp, eve) 3: Neuroectoderm odd (Antp, SoxN, eve) 4: Drosal Ectoderm even (Antp, Abx-1, ftz, trn) 6: Neuroectoderm even (Antp, SoxN, ftz, trn) 7: Amnioserosa (zen, peb, egr) 10: Ectoderm PS2 (Scr, ken, SoxN)	ectoderm_medial_DE_PS2	Ectoderm	Dorsal_Ectoderm_PS2	Scr, trn, dpp, tup, Doc2		
	ectoderm_intermediate_DE_PS2	Ectoderm	Dorsal_Ectoderm_PS2	Scr, trn, dpp, tup		
	ectoderm_lateral_DE_PS2	Ectoderm	Dorsal_Ectoderm_PS2	Scr, trn, dpp, Atx-1		
	ectoderm_lateral_NE_PS2	Ectoderm	Neuroectoderm_PS2	Scr, trn, SoxN, Dr, Atx-1		
	ectoderm_intermediate_NE_PS2	Ectoderm	Neuroectoderm_PS2	Scr, trn, SoxN, sog, ind		
	ectoderm_medial_NE_PS2	Ectoderm	Neuroectoderm_PS2	Scr, trn, SoxN, sog, vnd		
	ectoderm_medial_DE_abdominal_even	Ectoderm	Trunk_Dorsal_Ectoderm	Antp, ftz, trn, dpp, tup, Doc2		
	ectoderm_medial_DE_abdominal_odd	Ectoderm	Trunk_Dorsal_Ectoderm	Antp, eve, dpp, tup, Doc2	trn	
	ectoderm_medial_DE_PS13	Ectoderm	Trunk_Dorsal_Ectoderm	Abd-B, eve, dpp, tup, Doc2	trn	
	ectoderm_intermediate_DE_abdominal_even	Ectoderm	Trunk_Dorsal_Ectoderm	Antp, ftz, trn, dpp, tup		Trunk ectoderm 2
	ectoderm_intermediate_DE_abdominal_odd	Ectoderm	Trunk_Dorsal_Ectoderm	Antp, eve, dpp, tup	trn	Trunk ectoderm 2
	ectoderm_intermediate_DE_PS13	Ectoderm	Trunk_Dorsal_Ectoderm	Abd-B, eve, dpp, tup	trn	Trunk ectoderm 2
	ectoderm_lateral_DE_abdominal_even	Ectoderm	Trunk_Dorsal_Ectoderm	Antp, ftz, trn, dpp, Atx-1		Trunk ectoderm 2
	ectoderm_lateral_DE_abdominal_odd	Ectoderm	Trunk_Dorsal_Ectoderm	Antp, eve, dpp, tup, Atx-1	trn	Trunk ectoderm 2
	ectoderm_lateral_DE_PS13	Ectoderm	Trunk_Dorsal_Ectoderm	Abd-B, eve, dpp, tup	trn	Trunk ectoderm 2
	ectoderm_lateral_NE_abdominal_even	Ectoderm	Trunk_Neuroectoderm	Antp, ftz, trn, SoxN, Dr, Atx-1		Trunk ectoderm 2
	ectoderm_lateral_NE_abdominal_odd	Ectoderm	Trunk_Neuroectoderm	Antp, eve, SoxN, Dr, Atx-1	trn	Trunk ectoderm 2
	ectoderm_lateral_NE_PS13	Ectoderm	Trunk_Neuroectoderm	Abd-B, eve, SoxN, Dr, Atx-1	trn	Trunk ectoderm 2
	ectoderm_intermediate_NE_abdominal_even	Ectoderm	Trunk_Neuroectoderm	Antp, ftz, trn, SoxN, sog, ind		Trunk ectoderm 2
ectoderm_intermediate_NE_abdominal_odd	Ectoderm	Trunk_Neuroectoderm	Antp, eve, SoxN, sog, vnd	trn	Trunk ectoderm 2	
ectoderm_intermediate_NE_PS13	Ectoderm	Trunk_Neuroectoderm	Abd-B, eve, SoxN, sog, ind	trn	Trunk ectoderm 2	
ectoderm_medial_NE_abdominal_even	Ectoderm	Trunk_Neuroectoderm	Antp, ftz, trn, SoxN, sog, vnd		Trunk ectoderm 2	
ectoderm_medial_NE_abdominal_odd	Ectoderm	Trunk_Neuroectoderm	Antp, eve, SoxN, sog, vnd		Trunk ectoderm 2	
ectoderm_medial_NE_PS13	Ectoderm	Trunk_Neuroectoderm	Abd-B, eve, SoxN, sog, vnd		Trunk ectoderm 2	
amnioserosa_trunk	Amnioserosa	Amnioserosa	Antp, zen, peb			
16: Midline cells (sim, E(spl)5-HLH)	midline_cells_even	Ectoderm	Midline_Cell	ftz, trn, sim, E(spl)-C		
	midline_cells_odd	Ectoderm	Midline_Cell	eve, sim, E(spl)-C	trn	
5: Posterior Midgut (fkh, hkb) 9: Ectoderm_PS14 (Abd-B, trn) 12: Hindgut (fkh, byn, ct)	amnioserosa_PS14	Amnioserosa	Amnioserosa	Abd-B, trn, ftz, zen		
	ectoderm_PS14_ventral	Ectoderm	Ectoderm_PS14	Abd-B, ftz, trn, sog	byn	
	ectoderm_PS14_dorsal	Ectoderm	Ectoderm_PS14	Abd-B, ftz, trn, Atx-1	byn	
	ectoderm_PS14/hindgut	Ectoderm	Ectoderm_PS14	Abd-B, trn, wg, byn	ftz	
	ectoderm_hindgut_ventral	Ectoderm	Hindgut	fkh, wg, byn	Abd-B	
	ectoderm_hindgut_dorsal	Ectoderm	Hindgut	fkh, wg, byn, tup	Abd-B	
	endoderm_postMG_ventral	Endoderm	Posterior_endoderm	fkh, hkb, sog, sim	peb	
endoderm_postMG_lateral	Endoderm	Posterior_endoderm	fkh, hkb, sim, peb, exex			
endoderm_postMG_dorsal	Endoderm	Posterior_endoderm	fkh, hkb, exex, peb	sim		
18: Anterior midgut/ Head mesoderm (fkh, kni, twi, sna)	endoderm_antMG_wg	Endoderm	Anterior_endoderm	fkh, srp, wg	CadN, sna, twi	
	endoderm_antMG_wntD	Endoderm	Anterior_endoderm	fkh, twi, sna, srp, wntD	CadN	
	mesoderm_head	Mesoderm	Mesoderm_head	twi, sna, CadN	fkh, srp	
15: Anterior mesoderm (twi, sna, gcm)	mesoderm_gcm	Mesoderm	Mesoderm_gcm	sna, twi, gcm	Dfd	
	mesoderm_gcm_Dfd	Mesoderm	Mesoderm_gcm	sna, twi, gcm, Dfd		
17: Mesoderm PS1-2 (twi, sna, Dfd, ken)	mesoderm_PS1_Dfd	Mesoderm	Trunk_Mesoderm	sna, twi, Dfd, eve, salm		
	mesoderm_PS2_ken	Mesoderm	Trunk_Mesoderm	sna, twi, ken, salm, trn		
0: Trunk mesoderm odd (sna, twi, eve) 1: Trunk mesoderm even (sna, twi, ftz, trn)	mesoderm_PS3	Mesoderm	Trunk_Mesoderm	sna, twi, eve, salm	trn, Antp	
	mesoderm_PS4	Mesoderm	Trunk_Mesoderm	sna, twi, Antp, trn, ftz	Ubx	
	mesoderm_PS5	Mesoderm	Trunk_Mesoderm	sna, twi, Antp, eve	trn	
	mesoderm_PS6	Mesoderm	Trunk_Mesoderm	sna, twi, Antp, Ubx, trn, ftz		
	mesoderm_abdominal_odd	Mesoderm	Trunk_Mesoderm	sna, twi, Ubx, abd-A, eve	trn, Antp	
	mesoderm_abdominal_even	Mesoderm	Trunk_Mesoderm	sna, twi, Ubx, abd-A, trn, ftz	Antp	
20: Mesoderm PS13 (sna, twi, Abd-B, eve)	mesoderm_PS13	Mesoderm	Trunk_Mesoderm	sna, twi, Abd-B	trn, Ubx	
13: Mesoderm PS14/ Caudal visceral mesoderm (sna, twi, Abd-B, trn, fkh,)	mesoderm_PS14	Mesoderm	Mesoderm_PS14	Abd-B, trn, ftz	HLH54F	
	mesoderm_PS14/mesoderm_caudal_viscerai	Mesoderm	Mesoderm_PS14	Abd-B, trn, HLH54F, wg	ftz	
	mesoderm_caudal_viscerai	Mesoderm	Caudal_viscerai_mesoderm	HLH54F, wg	ftz	
19: Pole cell (pgc, stai, nos)	pole_cells	Pole_cell	Pole_cell	pgc		

表 3 Set 2 のサブクラスタリング結果

Seurat Clusters (markers)	Subclusters	Germ layer	positive markers	negative markers
8: Head_ectoderm(oc) 12: Head_ectoderm_kn (kn)	mesoderm_head	Mesoderm	sna, twi, hti, CadN	srp
	endoderm_antMG	Endoderm	fkf, hkb, srp	
	ectoderm_head_croc	Ectoderm	croc, Optix, fkf	srp
	ectoderm_head_Optix_Six4	Ectoderm	oc, Optix, Six4, SoxN	
	ectoderm_head_Optix_sog	Ectoderm	oc, Optix, sog, SoxN	
	ectoderm_head_oc_CenG1A	Ectoderm	oc, CenG1A	Optix
	ectoderm_head_oc_toy	Ectoderm	oc, toy	
	ectoderm_head_oc_Oaz	Ectoderm	oc, Oaz	
	ectoderm_head_Dfd_Oaz	Ectoderm	Dfd, Oaz	
	ectoderm_head_kn	Ectoderm	kn	Dfd
	ectoderm_DE_PS0	Ectoderm	Dfd, kn, sog	
	ectoderm_lateral_NE_PS0	Ectoderm	Dfd, kn, Atx-1	grn
	ectoderm_medial_NE_PS0	Ectoderm	Dfd, kn, grn	
	amnioserosa_anterior	Amnioserosa	salm, zen	
1: Medial_neuroectoderm (sog, SoxN) 5: Lateral_neuroectoderm (SoxN) 2: Dorsal_ectoderm_even (dpp, trn) 3: Dorsal_ectoderm_odd (dpp) 4: Amnioserosa (dap, zen) 7: Ectoderm_PS1-2 (Dfd, Scr)	ectoderm_medial_DE_PS1	Ectoderm	Dfd, dpp, tup, Doc2	
	ectoderm_intermediate_DE_PS1	Ectoderm	Dfd, dpp, tup	
	ectoderm_lateral_DE_PS1	Ectoderm	Dfd, dpp, Atx-1	
	ectoderm_lateral_NE_PS1	Ectoderm	Dfd, SoxN, Dr	
	ectoderm_intermediate_NE_PS1	Ectoderm	Dfd, SoxN, sog, ind	
	ectoderm_medial_NE_PS1	Ectoderm	Dfd, SoxN, sog, vnd	
	ectoderm_medial_DE_PS2	Ectoderm	Scr, dpp, tup, Doc2	
	ectoderm_intermediate_DE_PS2	Ectoderm	Scr, dpp, tup	
	ectoderm_lateral_DE_PS2	Ectoderm	Scr, dpp, Atx-1	
	ectoderm_lateral_NE_PS2	Ectoderm	Scr, SoxN, Dr	
	ectoderm_intermediate_NE_PS2	Ectoderm	Scr, SoxN, sog, ind	
	ectoderm_medial_NE_PS2	Ectoderm	Scr, SoxN, sog, vnd	
	ectoderm_medial_DE_abdominal	Ectoderm	Antp, dpp, tup, Doc2	
	ectoderm_intermediate_DE_abdominal	Ectoderm	Antp, dpp, tup	
	ectoderm_lateral_DE_abdominal	Ectoderm	Antp, dpp, Atx-1	
	ectoderm_lateral_NE_abdominal	Ectoderm	Antp, SoxN, Dr	
	ectoderm_intermediate_NE_abdominal	Ectoderm	Antp, SoxN, sog, ind	
ectoderm_medial_NE_abdominal	Ectoderm	Antp, SoxN, sog, vnd		
amnioserosa_trunk	Amnioserosa	zen, peb		
13: Midline cells (sim)	midline_cells	Ectoderm	sim	
6: Posterior_midgut (fkf, hkb) 9: Hindgut (fkf, byn) 10: Ectoderm_PS14 (Abd-B, trn)	ectoderm_PS14_ventral	Ectoderm	Abd-B, SoxN	byn
	ectoderm_PS14_dorsal	Ectoderm	Abd-B, dpp	byn
	ectoderm_PS14/hindgut_ventral	Ectoderm	Abd-B, sog, wg, byn	ftz
	ectoderm_PS14/hindgut_dorsal	Ectoderm	Abd-B, dpp, wg, byn	ftz, sog
	ectoderm_hindgut_ventral	Ectoderm	fkf, wg, byn	Abd-B
	ectoderm_hindgut_dorsal	Ectoderm	fkf, wg, byn, tup	Abd-B
	endoderm_postMG_ventral	Endoderm	fkf, hkb, sog, sim	peb
	endoderm_postMG_lateral	Endoderm	fkf, hkb, exex, sim, peb	
	endoderm_postMG_dorsal	Endoderm	fkf, hkb, exex, peb	sim
16: Anterior_mesoderm (sna, twi, gcm)	mesoderm_gcm	Mesoderm	sna, twi, gcm	Dfd
	mesoderm_gcm_Dfd	Mesoderm	sna, twi, gcm, Dfd	
0: Trunk_mesoderm (sna, twi, Antp, Ubx)	mesoderm_PS1_Dfd	Mesoderm	sna, twi, Dfd	Antp, Ubx
	mesoderm_PS2_ken	Mesoderm	sna, twi, ken	Antp, Ubx
	mesoderm_PS3	Mesoderm	sna, twi	Antp, Ubx, Dfd, ken
	mesoderm_PS4_and_PS6	Mesoderm	sna, twi, Antp, trn	
	mesoderm_PS5_and_abdominal_odd	Mesoderm	sna, twi, Antp, Ubx, abd-A	trn
mesoderm_abdominal_even	Mesoderm	sna, twi, Ubx, abd-A, trn	Antp	
15: Mesoderm_PS13 (sna, twi, Abd-B)	mesoderm_PS13	Mesoderm	sna, twi, Abd-B	
11: Posterior_mesoderm (sna, twi, Abd-B, fkf)	mesoderm_PS14	Mesoderm	sna, twi, Abd-B, ftz	HLH54F
	mesoderm_caudal_visceral	Mesoderm	sna, twi, fkf, wg, HLH54F	Abd-B
14: Pole_cells	pole_cells	Pole_cell	pgc	

表 4 Set 1 のサブクラスタリング結果

Seurat Clusters (markers)	Subclusters	Germ layer	positive markers	negative markers
6: Head ectoderm (oc) 10: Head ectoderm_kn (kn)	endoderm_antlMG_and_ectoderm_head_croc	Endoderm	fkf, croc, Optix, hkb	
	ectoderm_head_oc_grn	Ectoderm	oc, grn	
	ectoderm_head_oc_CenG1A	Ectoderm	oc, CenG1A	
	ectoderm_head_oc_Pvf3	Ectoderm	oc, Pvf3	
	ectoderm_head_kn	Ectoderm	kn	
0: Trunk_medial_neuroectoderm (sog, SoxN) 5: Trunk_lateral_neuroectoderm (SoxN) 1: Trunk_dorsal_ectoderm (dpp) 8: Amnioserosa (dap) 14: Dorsal_ectoderm_PS2 (Scr, dpp) 9: Ectoderm_PS1 (Dfd)	ectoderm_medial_DE_PS1_and_PS2	Ectoderm	Dfd or Scr, dpp, tup, Doc2,	
	ectoderm_intermediate_DE_PS1_and_PS2	Ectoderm	Dfd or Scr, dpp, tup	
	ectoderm_lateral_DE_PS1_and_PS2	Ectoderm	Dfd or Scr, dpp, Atx-1	
	ectoderm_lateral_NE_PS1_and_PS2	Ectoderm	Dfd or Scr, SoxN, Dr	
	ectoderm_intermediate_NE_PS1_and_PS2	Ectoderm	Dfd or Scr, SoxN, sog, ind	
	ectoderm_medial_NE_PS1_and_PS2	Ectoderm	Dfd or Scr, SoxN, sog, vnd	
	ectoderm_medial_DE_abdominal	Ectoderm	Antp, dpp, tup, Doc2	
	ectoderm_intermediate_DE_abdominal	Ectoderm	Antp, dpp, tup	
	ectoderm_lateral_DE_abdominal	Ectoderm	Antp, dpp, Atx-1	
	ectoderm_lateral_NE_abdominal	Ectoderm	Antp, SoxN, Dr	
	ectoderm_intermediate_NE_abdominal	Ectoderm	Antp, SoxN, sog, ind	
	ectoderm_medial_NE_abdominal	Ectoderm	Antp, SoxN, sog, vnd	
	amnioserosa_trunk	Amnioserosa	zen, peb	
13: Midline_cells (sim)	midline_cells	Ectoderm	sim	
12: Ectoderm_PS14 (Abd-B, trn) 4: Hindgut (fkf, byn, ct) 7: Posterior_midgut (fkf, hkb)	ectoderm_PS14	Ectoderm	Abd-B, SoxN	byn
	ectoderm_PS14/hindgut	Ectoderm	Abd-B, SoxN, wg, byn	
	ectoderm_hindgut_ventral	Ectoderm	fkf, wg, byn	Abd-B
	ectoderm_hindgut_dorsal	Ectoderm	fkf, wg, byn, tup	Abd-B
	endoderm_postlMG_ventral	Endoderm	fkf, hkb, sog, sim	peb
	endoderm_postlMG_lateral	Endoderm	fkf, hkb, peb, exex, sim	
endoderm_postlMG_dorsal	Endoderm	fkf, hkb, exex, peb	sim	
11: Anterior_mesoderm (gcm)	mesoderm_gcm	Mesoderm	sna, twi, gcm	
2: Trunk_mesoderm_odd (sna, twi, eve) 3: Trunk_mesoderm_even (sna, twi, ftz, trn)	mesoderm_trunk_odd	Mesoderm	sna, twi, eve	trn, ftz
	mesoderm_trunk_even	Mesoderm	sna, twi, ftz, trn	eve
15: Posterior_mesoderm (sna, twi, Abd-B, trn, fkf)	mesoderm_PS14	Mesoderm	sna, twi, Abd-B, ftz	HLH54F
	mesoderm_caudal_visceral	Mesoderm	sna, twi, fkf, wg, HLH54F	ftz, Abd-B
16: Pole_cells (pgc)	pole_cells	Pole_cell	pgc	

表 5 NK-data のサブクラスタリング結果

Seurat Clusters (markers)	Subclusters	Germ layer	positive markers	negative markers
6: Head_ectoderm (oc) 16: Ectoderm_kn (kn) 9: Anterior_amnioserosa (zen, Dll)	ectoderm_head_croc	Ectoderm	croc, fkh	
	ectoderm_head_Optix	Ectoderm	oc, Optix	
	ectoderm_head_oc_CenG1A	Ectoderm	oc, CenG1A	
	ectoderm_head_oc_eya	Ectoderm	oc, so, eya	
	ectoderm_head_oc_so	Ectoderm	oc, so	
	ectoderm_head_kn	Ectoderm	Dfd, kn	
	amnioserosa_anterior	Amnioserosa	zen, mirr, peb, salm	
1: Medial_neuroectoderm (sog, SoxN) 4: Lateral_neuroectoderm (SoxN) 2: Dorsal_ectoderm (dpp) 7: Amnioserosa (zen) 15: Dorsal_ectoderm_anterior	amnioserosa_trunk	Amnioserosa	zen, mirr, peb	
	ectoderm_medial_DE_trunk	Ectoderm	dpp, tup, Doc2	
	ectoderm_intermediate_DE_trunk	Ectoderm	dpp, tup	
	ectoderm_lateral_DE_trunk	Ectoderm	dpp, Atx-1	
	ectoderm_lateral_NE_trunk	Ectoderm	SoxN, Dr	
	ectoderm_intermediate_NE_trunk	Ectoderm	SoxN, sog, ind	
	ectoderm_medial_NE_trunk	Ectoderm	SoxN, sog, vnd	
13: Midline cells (sim, E(spl)-C)	midline_cells	Ectoderm	sim, E(spl)-C	
5: Posterior_midgut (fkh, hkb) 8: Ectoderm_PS14/Hindgut (Abd-B, byn)	ectoderm_PS14	Ectoderm	Abd-B	byn
	ectoderm_hindgut	Ectoderm	fkh, wg, byn	Abd-B
	endoderm_postMG_ventral	Endoderm	fkh, hkb, sog, sim	peb
	endoderm_postMG_dorsal	Endoderm	fkh, hkb, exex, peb	sim
11: Anterior_midgut (fkh, hkb, kni)	endoderm_antMG_wg	Endoderm	fkh, wg	sna, twi
	endoderm_antMG_wntD	Endoderm	fkh, sna, twi, wntD	CadN
10: Anterior_mesoderm (sna, twi, gcm)	mesoderm_gcm	Mesoderm	sna, twi, gcm	
12: Mesoderm_PS1-3 (sna, twi, salm)	mesoderm_PS1-3	Mesoderm	sna, twi, salm	
0, 3: Trunk_mesoderm (sna, twi)	mesoderm_PS4-6	Mesoderm	sna, twi, Antp	
	mesoderm_abdominal	Mesoderm	sna, twi	Antp, Abd-B
	mesoderm_PS13	Mesoderm	sna, twi, Abd-B	trn
16: Posterior_mesoderm (sna, twi, Abd-B)	mesoderm_posterior	Mesoderm	sna, twi, Abd-B, trn	
17: Pole_cells (pgc)	pole_cells	Pole_cell	pgc	

表 6 統合データのサブクラスタリング結果

Seurat Clusters	Subclusters	Germ layer	positive markers	negative markers
9: Head_ectoderm (oc) 13: Ectoderm_kn (kn)	ectoderm_head_fkh	Ectoderm	Optix, croc, fkh	
	ectoderm_head_croc	Ectoderm	Optix, croc	fkh
	ectoderm_head_Optix_Six4_SoxN	Ectoderm	oc, Optix, Six4, SoxN	grn
	ectoderm_head_Optix_Six4_grn	Ectoderm	oc, Optix, Six4, grn	SoxN
	ectoderm_head_Optix_toy	Ectoderm	oc, Optix, toy, SoxN	
	ectoderm_head_Optix_CenG1A	Ectoderm	oc, Optix, CenG1A, SoxN	
	ectoderm_head_oc_CenG1A	Ectoderm	oc, CenG1A, SoxN	Optix
	ectoderm_head_oc_eya	Ectoderm	oc, eya, SoxN	
	ectoderm_head_oc_so_toy	Ectoderm	oc, so, toy, grn	SoxN
	ectoderm_head_oc_so	Ectoderm	oc, so, grn	toy
	ectoderm_head_oc_Doc2	Ectoderm	oc, so, Doc2, SoxN	
	ectoderm_head_oc_Pvf3	Ectoderm	oc, Pvf3	
	ectoderm_head_kn_lateral	Ectoderm	kn	SoxN
	ectoderm_head_kn_medial	Ectoderm	kn, SoxN	
	ectoderm_PS0_ems	Ectoderm	kn, Dfd, ems	
	ectoderm_medial_NE_PS0	Ectoderm	kn, Dfd, grn	ems, SoxN
	ectoderm_intermediate_NE_PS0	Ectoderm	kn, Dfd, SoxN	ems
	ectoderm_lateral_NE_PS0	Ectoderm	kn, Dfd, SoxN, ind	ems
ectoderm_DE_PS0	Ectoderm	kn, Dfd, SoxN, vnd	ems	
0: Neuroectoderm (SoxN, sog) 1: Dorsal_ectoderm_odd (dpp) 2: Dorsal_ectoderm_even (dpp, trn) 6: Amnioserosa (dap, zen) 10: Ectoderm_PS2 (Scr) 19: Dorsal_ectoderm_PS1 (Dfd, dpp) 21: Neuroectoderm_PS1 (Dfd, SoxN, sog) 15: Amnioserosa_anterior (dap, zen, salm)	ectoderm_head_Dfd_Oaz	Ectoderm	Dfd, Oaz, oc	
	amnioserosa_PS1	Amnioserosa	zen, Dfd, salm	oc
	amnioserosa_PS2	Amnioserosa	zen, Scr, salm	oc
	ectoderm_medial_DE_PS1	Ectoderm	Dfd, dpp, tup, Doc2	oc
	ectoderm_lateral_DE_PS1	Ectoderm	Dfd, dpp, Atx-1	oc
	ectoderm_lateral_NE_PS1	Ectoderm	Dfd, SoxN, Dr	oc
	ectoderm_intermediate_NE_PS1	Ectoderm	Dfd, SoxN, sog, ind	oc
	ectoderm_medial_NE_PS1	Ectoderm	Dfd, SoxN, sog, vnd	oc
	ectoderm_medial_DE_PS2	Ectoderm	Scr, dpp, tup, Doc2	oc
	ectoderm_intermediate_DE_PS2	Ectoderm	Scr, dpp, Doc2	oc
	ectoderm_lateral_DE_PS2	Ectoderm	Scr, dpp, Atx-1	oc
	ectoderm_lateral_NE_PS2	Ectoderm	Scr, SoxN, Dr	oc
	ectoderm_intermediate_NE_PS2	Ectoderm	Scr, SoxN, sog, ind	oc
	ectoderm_medial_NE_PS2	Ectoderm	Scr, SoxN, sog, vnd	oc
	ectoderm_medial_DE_abdominal	Ectoderm	Antp, dpp, tup, Doc2	oc
	ectoderm_intermediate_DE_abdominal	Ectoderm	Antp, dpp, Doc2	oc
	ectoderm_lateral_DE_abdominal	Ectoderm	Antp, dpp, Atx-1	oc
	ectoderm_lateral_NE_abdominal	Ectoderm	Antp, SoxN, Dr	oc
	ectoderm_intermediate_NE_abdominal	Ectoderm	Antp, SoxN, sog, ind	oc
	ectoderm_medial_NE_abdominal	Ectoderm	Antp, SoxN, sog, vnd	oc
amnioserosa_trunk	Amnioserosa	zen, mirr, peb	oc	
14: Midline cells (sim, E(spl)-C)	midline_cells_PS1-3	Ectoderm	sim, salm	
	midline_cells_abdominal_even	Ectoderm	sim, trn	
	midline_cells_abdominal_odd	Ectoderm	sim	trn
5: Posterior_midgut(fkh, hkb) 7: Hindgut (fkh, byn) 8: Ectoderm_PS14 (Abd-B)	ectoderm_PS14_ventral	Ectoderm	Abd-B, SoxN, sog	
	ectoderm_PS14_dorsal	Ectoderm	Abd-B, dpp	sog
	ectoderm_PS14/hindgut	Ectoderm	Abd-B, SoxN, wg, byn	
	ectoderm_hindgut_ventral	Ectoderm	fkh, wg, byn	Abd-B
	ectoderm_hindgut_dorsal	Ectoderm	fkh, wg, byn, tup	Abd-B
	endoderm_postMG_ventral	Endoderm	fkh, hkb, sog, sim	peb
endoderm_postMG_lateral	Endoderm	fkh, hkb, exex, sim, peb		
endoderm_postMG_dorsal	Endoderm	fkh, hkb, exex, peb	sim	
17: Head_mesoderm/Anterior_midgut (sna, twi, fkh, kni)	endoderm_antMG_wg	Endoderm	fkh, wg	sna, twi
	endoderm_antMG_wntD	Endoderm	fkh, sna, twi, wntD	CadN
	mesoderm_head	Mesoderm	sna, twi, htl	fkh
16: Anterior_mesoderm (sna, twi, gcm)	mesoderm_gcm	Mesoderm	sna, twi, gcm,	Dfd
	mesoderm_gcm_Dfd	Mesoderm	sna, twi, gcm, Dfd	
12: Mesoderm_PS1-3 (sna, twi, salm)	mesoderm_PS1_Dfd	Mesoderm	sna, twi, salm, Dfd	
	mesoderm_PS2_ken	Mesoderm	sna, twi, salm, ken	
	mesoderm_PS3	Mesoderm	sna, twi, salm	Dfd, ken
3: Mesoderm_even (sna, twi, trn) 4: Mesoderm_odd (sna, twi)	mesoderm_PS4_and_PS6	Mesoderm	sna, twi, Antp, trn	
	mesoderm_PS5	Mesoderm	sna, twi, Antp, Ubx	trn
	mesoderm_abdominal_odd	Mesoderm	sna, twi, Ubx, abd-A	Antp, trn
	mesoderm_abdominal_even	Mesoderm	sna, twi, Ubx, abd-A, trn	Antp
20: Mesoderm_PS13 (sna, twi, Abd-B)	mesoderm_PS13	Mesoderm	sna, twi, Abd-B	trn
11: Posterior_mesoderm (sna, twi, Abd-B, trn, fkh)	mesoderm_PS14	Mesoderm	sna, twi, Abd-B, ftz	fkh, HLH54F
	mesoderm_PS14/mesoderm_caudal_visceral	Mesoderm	Abd-B, trn, HLH54F, wg	
	mesoderm_caudal_visceral	Mesoderm	sna, twi, fkh, wg, HLH54F	Abd-B
18: Pole_cells (pgc)	pole_cells	Pole_cell	pgc	

表 7 *bcd*-RNAi データのサブクラスタリング結果

Seurat Clusters (markers)	Subclusters	Germ layer	positive markers	negative markers
1: Trunk_dorsal_ectoderm (dpp) 2: Trunk_lateral_neuroectoderm (SoxN) 4: Trunk_medial_neuroectoderm (SoxN, sog) 6: Amnioserosa (zen) 9: Dorsal_ectoderm_PS13 (Abd-B, dpp)	ectoderm_medial_DE_abdominal	Ectoderm	Antp, dpp, tup, Doc2	
	ectoderm_intermediate_DE_abdominal	Ectoderm	Antp, dpp, tup	
	ectoderm_lateral_DE_abdominal	Ectoderm	Antp, dpp, Atx-1	
	ectoderm_lateral_NE_abdominal	Ectoderm	Antp, SoxN, Dr	
	ectoderm_intermediate_NE_abdominal	Ectoderm	Antp, SoxN, sog, ind	
	ectoderm_medial_NE_abdominal	Ectoderm	Antp, SoxN, sog, vnd	
	amnioserosa_trunk	Amnioserosa	zen, mirr, peb	
	amnioserosa_PS13-14	Amnioserosa	zen, mirr, peb, Abd-B	
	midline_cells	Ectoderm	sim	
15: Midline_cells	ectoderm_PS14_ventral	Ectoderm	Abd-B, SoxN	byn
	ectoderm_PS14_dorsal	Ectoderm	Abd-B, dpp	byn, sog
	ectoderm_PS14/ectoderm_hindgut	Ectoderm	Abd-B, wg, byn	
	ectoderm_hindgut_ventral	Ectoderm	fkh, wg, byn	Abd-B
	ectoderm_hindgut_dorsal	Ectoderm	fkh, wg, byn, tup	Abd-B
	endoderm_postMG_ventral	Endoderm	fkh, hkb, sog, sim	peb
	endoderm_postMG_lateral	Endoderm	fkh, hkb, exex, sim, peb	
	endoderm_postMG_dorsal	Endoderm	fkh, hkb, exex, peb	sim
3: Posterior_midgut (fkh, hkb) 7: Hindgut (fkh, byn) 5: Ectoderm_PS14 (Abd-B)	mesoderm_PS6	Mesoderm	sna, twi, Antp, Ubx	
	mesoderm_abdomial_odd	Mesoderm	sna, twi, Ubx	Antp, trn
	mesoderm_abdominal_even	Mesoderm	sna, twi, Ubx, trn	Antp
0: Trunk_mesoderm (sna, twi) 10: Trunk_mesoderm_eve (sna, twi, eve)	mesoderm_PS13	Mesoderm	sna, twi, Abd-B	trn
12: Mesoderm_PS13 (sna, twi, Abd-B)	mesoderm_PS14	Mesoderm	sna, twi, Abd-B, ftz	HLH54F
8: Posterior_mesoderm (sna, twi, Abd-B, trn)	mesoderm_caual visceral	Mesoderm	sna, twi, fkh, wg, HLH54F	Abd-B
14: Pole_cells	pole_cells	Pole_cell	pgc	

表 8 bcd-RNAi / コントロール細胞比の仮説検定の p-value の一覧

	cluster name	p-values
Trunk & Pole cells	Trunk_amnioserosa	0.14259457
	Trunk_dorsal_ectoderm_even	0.001253923
	Trunk_dorsal_ectoderm_odd	8.99E-07
	Trunk_lateral_neuroectoderm	0.289076661
	Trunk_medial_neuroectoderm	9.85E-07
	Midline_cells	0.626706448
	Trunk_Mesoderm	0.000184026
	Pole_cells	0.096909871
Posterior	Amnioserosa_PS13.14	2.67E-14
	Ectoderm_PS14	1.51E-15
	Dorsal_Ectoderm_PS13	1.13E-09
	Hindgut	1.53E-19
	Posterior_midgut	3.67E-28
	Mesoderm_PS13	2.81E-09
	Posterior_Mesoderm	1.21E-14
Anterior (no bcd-RNAi cells)	Head_ectoderm	2.70E-87
	Ectoderm_PS1.2	7.91E-93
	Anterior_midgut.mesoderm	6.32E-41

表 9 C1HT の実行に用いたカスタムオリゴ配列

primer name	sequence
oligodTprimer1	ACACTCTTTCCCTACACGACGCTCTCCGATCTCACGTANNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer2	ACACTCTTTCCCTACACGACGCTCTCCGATCTCTCAGANNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer3	ACACTCTTTCCCTACACGACGCTCTCCGATCTTGATCANNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer4	ACACTCTTTCCCTACACGACGCTCTCCGATCTCAGACNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer5	ACACTCTTTCCCTACACGACGCTCTCCGATCTCGATGTNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer6	ACACTCTTTCCCTACACGACGCTCTCCGATCTTACTGCNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer7	ACACTCTTTCCCTACACGACGCTCTCCGATCTATGCTCNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer8	ACACTCTTTCCCTACACGACGCTCTCCGATCTCATCTGNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer9	ACACTCTTTCCCTACACGACGCTCTCCGATCTGACTCANNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer10	ACACTCTTTCCCTACACGACGCTCTCCGATCTAGATCGNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer11	ACACTCTTTCCCTACACGACGCTCTCCGATCTATCAGCANNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer12	ACACTCTTTCCCTACACGACGCTCTCCGATCTGCTACANNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer13	ACACTCTTTCCCTACACGACGCTCTCCGATCTCAGATCANNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer14	ACACTCTTTCCCTACACGACGCTCTCCGATCTCACAGTNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer15	ACACTCTTTCCCTACACGACGCTCTCCGATCTTACGAGNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer16	ACACTCTTTCCCTACACGACGCTCTCCGATCTCGACTANNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer17	ACACTCTTTCCCTACACGACGCTCTCCGATCTGCATCTNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer18	ACACTCTTTCCCTACACGACGCTCTCCGATCTAGACTNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer19	ACACTCTTTCCCTACACGACGCTCTCCGATCTACTGNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer20	ACACTCTTTCCCTACACGACGCTCTCCGATCTCGTAGANNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer21	ACACTCTTTCCCTACACGACGCTCTCCGATCTACTCGANNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer22	ACACTCTTTCCCTACACGACGCTCTCCGATCTACATGCNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer23	ACACTCTTTCCCTACACGACGCTCTCCGATCTCGTCATNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer24	ACACTCTTTCCCTACACGACGCTCTCCGATCTGTAGCANNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer25	ACACTCTTTCCCTACACGACGCTCTCCGATCTGCAGTANNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer26	ACACTCTTTCCCTACACGACGCTCTCCGATCTCACGTNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer27	ACACTCTTTCCCTACACGACGCTCTCCGATCTACGTCANNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer28	ACACTCTTTCCCTACACGACGCTCTCCGATCTCTCGATNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer29	ACACTCTTTCCCTACACGACGCTCTCCGATCTATCGTGNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer30	ACACTCTTTCCCTACACGACGCTCTCCGATCTGCTGATNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer31	ACACTCTTTCCCTACACGACGCTCTCCGATCTGTCTACNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer32	ACACTCTTTCCCTACACGACGCTCTCCGATCTCATGCTNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer33	ACACTCTTTCCCTACACGACGCTCTCCGATCTTAGCACNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer34	ACACTCTTTCCCTACACGACGCTCTCCGATCTGTGCATNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer35	ACACTCTTTCCCTACACGACGCTCTCCGATCTTAGTCGNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer36	ACACTCTTTCCCTACACGACGCTCTCCGATCTTCTCAGNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer37	ACACTCTTTCCCTACACGACGCTCTCCGATCTCTAGTCNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer38	ACACTCTTTCCCTACACGACGCTCTCCGATCTTCTAGCANNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer39	ACACTCTTTCCCTACACGACGCTCTCCGATCTATGACGNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
oligodTprimer40	ACACTCTTTCCCTACACGACGCTCTCCGATCTGAGCTANNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN

表 10 修正した gtf において統合された遺伝子の一覧 (1/3)

original gene symbol			modified gene symbol
CG34178	CG34177		CG34178_CG34177
SP555	CG14042		SP555_CG14042
CG44000	CG33995	CG44001	Edited00002
CG45075	Kr-h1		CG45075_Kr-h1
CG34310	cup		CG34310_cup
CG43355	sala		CG43355_sala
CG14933	CG42486		CG14933_CG42486
RpL7-like	CG34164		RpL7-like_CG34164
Adh	Adhr		Adh_Adhr
CG43816	CG13245		CG43816_CG13245
BicC	CG45084		BicC_CG45084
ScfB	ScfA		ScfB_ScfA
CG34313	CG31743		CG34313_CG31743
beat-IIIa	CG34106		beat-IIIa_CG34106
CG42635	CG42634		CG42635_CG42634
CG17325	CG42305		CG17325_CG42305
CG17344	CG43731		CG17344_CG43731
CG31976	CG31975		CG31976_CG31975
CG15386	CG42371		CG15386_CG42371
CG34314	Rad1		CG34314_Rad1
Snapin	CG9960		Snapin_CG9960
CG31917	Tfb5		CG31917_Tfb5
slmo	CG34179		slmo_CG34179
CG42730	tectonic		CG42730_tectonic
ImgB	ImgA		ImgB_ImgA
lncRNA:CR44872	CG5850		CR44872_CG5850
LSm-4	CG17768		LSm-4_CG17768
CG17139	CG17140		CG17139_CG17140
DNApol-gamma35	CG33649		DNApol-gamma35_CG33649
chif	CG42231		chif_CG42231
CG10570	CG42502		CG10570_CG42502
CG43739	sky		CG43739_sky
CG43345	CG43346		CG43345_CG43346
CG44286	CG8788		CG44286_CG8788
CG33964	CG13175		CG33964_CG13175
Galphaq	CG45086		Galphaq_CG45086
CG33155	mRpl53		CG33155_mRpl53
Obp50c	Obp50b		Obp50c_Obp50b
CG44243	CG42837		CG44243_CG42837
CG30100	CG42372		CG30100_CG42372
CG4984	CG4975		CG4984_CG4975
CG42518	MED9		CG42518_MED9
CG30120	CG5189		CG30120_CG5189
Pepck	CG45087		Pepck_CG45087
CG42363	CG42362		CG42363_CG42362
CG42496	Ppcdc		CG42496_Ppcdc
CG13511	CG42565	CG13510	Edited00046
CG30273	CG30269		CG30273_CG30269
CG42284	CG30274		CG42284_CG30274
CG45068	CG45069		CG45068_CG45069
CG3409	CG45092		CG3409_CG45092
CG30503	kappaB-Ras		CG30503_kappaB-Ras
CG42516	Pbp49		CG42516_Pbp49
spaw	hubl		spaw_hubl
swif	cola		swif_cola
CG33199	CG8229		CG33199_CG8229
CG45085	CG8026		CG45085_CG8026
CG13739	CG13954		CG13739_CG13954
Hr46	CG12912		Hr46_CG12912
CG43172	CG43200		CG43172_CG43200
Edited00061			Edited00061
CG43397	CG12902		CG43397_CG12902
pre-lola-G	lola		pre-lola-G_lola
CG33672	CG33671		CG33672_CG33671
CG17724	Kdm4B	seq	CG17724_Kdm4B_seq

表 10 修正した gtf において統合された遺伝子の一覧 (2/3)

original gene symbol				modified gene symbol
lat	CG34315			lat_CG34315
CG43178	CG43201	CG43171		Edited00067
CG45088	CG6191			CG45088_CG6191
CG30479	CG30480			CG30479_CG30480
CG42662	CG30080			CG42662_CG30080
Menl-1	Menl-2			Menl-1_Menl-2
CG30461	ste24c			CG30461_ste24c
CG43328	CG43327			CG43328_CG43327
fab1	CG33981			fab1_CG33981
Gint3	CG42306			Gint3_CG42306
CG44434	CG44435			CG44434_CG44435
Cpr57A	Obp57d			Cpr57A_Obp57d
CG18065	CG13430			CG18065_CG13430
CG33786	CG33785			CG33786_CG33785
CG9865	CG42379	CG42381	CG42380	Edited00080
CG42497	Tim10			CG42497_Tim10
CG43325	CG43326			CG43325_CG43326
CG42560	CG42559			CG42560_CG42559
CG10332	IM18			CG10332_IM18
CG43776	CG43777	CG43775		Edited00085
CG42568	DnaJ-60			CG42568_DnaJ-60
CG9133	CG9130			CG9133_CG9130
CG33791	CG18170			CG33791_CG18170
MEP-1	CG42245			MEP-1_CG42245
CG42304	Snup			CG42304_Snup
CG12016	CG42456			CG12016_CG42456
Gr64c	Gr64b			Gr64c_Gr64b
Gr64d	Gr64e			Gr64d_Gr64e
PGRP-LD	Pmi			PGRP-LD_Pmi
mus312	CG42307			mus312_CG42307
CG42661	CG42660	Tsp66A		CG42661_CG42660_Tsp66A
Hsp22	Hsp67Bb			Hsp22_Hsp67Bb
vnc	CG42455			vnc_CG42455
SuUR	CG45101			SuUR_CG45101
Ccdc56	mtTFB1			Ccdc56_mtTFB1
Nxf3	ssp2			Nxf3_ssp2
gdl-ORF39	gdl			gdl-ORF39_gdl
asRNA:CR45893	CG4998			CR45893_CG4998
CG45081	CG42374	CG9666		Edited00104
Mkp	CG34140			Mkp_CG34140
lncRNA:CR45184	CG13900			CR45184_CG13900
CG32298	CG32299			CG32298_CG32299
CG32270	CG32269			CG32270_CG32269
CG32388	CG43439			CG32388_CG43439
CG43781	CG43780			CG43781_CG43780
CG33057	mkg-p			CG33057_mkg-p
CG6709	CG14164			CG6709_CG14164
lncRNA:CR45443	lncRNA:CR45671			CR45443_CR45671
Trl	CG42507			Trl_CG42507
CG5151	CG32152			CG5151_CG32152
CG43373	CG42514			CG43373_CG42514
CG43954	Lasp			CG43954_Lasp
frc	CG32174			frc_CG32174
CG4174	CG13380			CG4174_CG13380
CG14077	lr75d			CG14077_lr75d
CG34277	CG31542			CG34277_CG31542
CG33722	CG18749			CG33722_CG18749
CG11760	CG8116			CG11760_CG8116
CG45050	CG43675			CG45050_CG43675
CG42511	CG8135			CG42511_CG8135
CG9288	CG42375			CG9288_CG42375
tal-2A	tal-1A	tal-AA	tal-3A	pri
CG42727	CG42726			CG42727_CG42726
Tgs1	moi			Tgs1_moi
CG42613	CG43732			CG42613_CG43732
Indy-2	CG33934			Indy-2_CG33934
CG7071	Muted			CG7071_Muted

表 10 修正した gtf において統合された遺伝子の一覧 (3/3)

original gene symbol				modified gene symbol
CG6726	CG17110			CG6726_CG17110
CG6000	lncRNA:CR44097			CG6000_CR44097
CG33096	CG33095			CG33096_CG33095
CG11889	CG11891			CG11889_CG11891
CG42488	CG5116			CG42488_CG5116
CG17770	Acam			CG17770_Acam
lncRNA:CR45040	Kaz-m1			CR45040_Kaz-m1
CG42557	CG42558			CG42557_CG42558
Sry-alpha	Sry-beta			Sry-alpha_Sry-beta
CG34317	CG7950			CG34317_CG7950
Hdac3	CG45100			Hdac3_CG45100
CG11768	CG34135			CG11768_CG34135
Pif1A	Pif1B			Pif1A_Pif1B
p24-2	Unc-115b			p24-2_Unc-115b
KP78a	KP78b			KP78a_KP78b
fabp	sea			fabp_sea
CG42504	CG42505	CG31358		Edited00149
CG10096	CG10097			CG10096_CG10097
CG44194	CG17327			CG44194_CG17327
primo-1	primo-2			primo-1_primo-2
CG7215	Prx5			CG7215_Prx5
Ada2a	Rpb4			Ada2a_Rpb4
CG43210	CG43194			CG43210_CG43194
Xport	CG42508			Xport_CG42508
att-ORFB	att-ORFA			att-ORFB_att-ORFA
CG45099	CG6015			CG45099_CG6015
CG43999	CG43998			CG43999_CG43998
CG42503	Mocs2			CG42503_Mocs2
CG17197	CG17198			CG17197_CG17198
CG14550	CG42498			CG14550_CG42498
CG42487	CG4884			CG42487_CG4884
CG2217	CG42740			CG2217_CG42740
sgg	lncRNA:CR45197			sgg_CR45197
CG15465	rg			CG15465_rg
CG42308	CG32736			CG42308_CG32736
CG15717	MFS10			CG15717_MFS10
CG6294	CG6299			CG6294_CG6299
CG32573	CG42512			CG32573_CG42512
Tyler	Shawn			Tyler_Shawn
Rpp20	CG33932			Rpp20_CG33932
CG11566	stg1			CG11566_stg1
waw	bbx			waw_bbx
CG18624	CG45089			CG18624_CG45089
CG10962	lncRNA:CR43836			CG10962_CR43836
lmp	lncRNA:CR45204			lmp_CR45204
Gr10a	Or10a			Gr10a_Or10a
inaF-D	inaF-B	inaF-C	inaF-A	inaF
Sclp	lncRNA:CR32660			Sclp_CR32660
CG45065	CG45064			CG45065_CG45064
CG42353	CG42354			CG42353_CG42354
mei-218	mei-217			mei-218_mei-217
CG12788	Tim9b			CG12788_Tim9b
CG33713	CG33714			CG33713_CG33714
stnB	stnA			stnB_stnA
CG17162	CG17159			CG17162_CG17159
xmas-1	xmas-2			xmas
pre-mod(mdg4)-Z	pre-mod(mdg4)-Y	pre-mod(mdg4)-X	pre-mod(mdg4)-AD	mod(mdg4)s
pre-mod(mdg4)-AE	pre-mod(mdg4)-W	pre-mod(mdg4)-V	pre-mod(mdg4)-U	
pre-mod(mdg4)-AA	pre-mod(mdg4)-AB	pre-mod(mdg4)-B	pre-mod(mdg4)-C	mod(mdg4)as
pre-mod(mdg4)-E	pre-mod(mdg4)-G	pre-mod(mdg4)-H	pre-mod(mdg4)-I	
pre-mod(mdg4)-J	pre-mod(mdg4)-K	pre-mod(mdg4)-L	pre-mod(mdg4)-N	
pre-mod(mdg4)-O	pre-mod(mdg4)-P	pre-mod(mdg4)-T	mod(mdg4)	

表 11 空間再構成に用いたランドマーク遺伝子の一覧

<i>aay</i>	<i>Ama</i>	<i>Ance</i>	<i>apt</i>	<i>Blimp-1</i>	<i>bowl</i>	<i>brk</i>	<i>Btk29A</i>
<i>bun</i>	<i>cad</i>	<i>CenG1A</i>	<i>CG10479</i>	<i>CG11208</i>	<i>CG43394</i>	<i>cnc</i>	<i>croc</i>
<i>Cyp310a1</i>	<i>D</i>	<i>dan</i>	<i>danr</i>	<i>Dfd</i>	<i>disco</i>	<i>Doc2</i>	<i>Doc3</i>
<i>dpn</i>	<i>E(spl)m5-HLH</i>	<i>edl</i>	<i>ems</i>	<i>eve</i>	<i>exex</i>	<i>fkf</i>	<i>ftz</i>
<i>h</i>	<i>hb</i>	<i>hkb</i>	<i>htl</i>	<i>llp4</i>	<i>ImpE2</i>	<i>ken</i>	<i>knrl</i>
<i>Kr</i>	<i>lok</i>	<i>Mdr49</i>	<i>Mes2</i>	<i>MESR3</i>	<i>mfas</i>	<i>NetA</i>	<i>noc</i>
<i>nub</i>	<i>oc</i>	<i>odd</i>	<i>peb</i>	<i>phu</i>	<i>prd</i>	<i>pxb</i>	<i>rau</i>
<i>sna</i>	<i>srp</i>	<i>tkv</i>	<i>tll</i>	<i>toc</i>	<i>Traf4</i>	<i>trn</i>	<i>tsh</i>
<i>twi</i>	<i>zen</i>	<i>CG17724_Kdm4B_seq</i>					

表 12 SABER-FISH のプローブ配列 (1/3)

primer name	sequence
Hairpin (h.27.27)	ACATCATCATGGGCCTTTTGGCCCATGATGATGTATGATGATG/3InvdT/ TGC GTTCTGGGTAAGGATAGGCCAACATGTGCTTttCATCATCAT AAACATAACCGACTTTAGACTACAAC TTATCGAGGGCTCGAttCATCATCAT ATTTTCGTACAGTTGCGAATTTCAAGTGCATCTGCTACTGttCATCATCAT TGATTCGGATCGGACATGGGGTTGGATTTCGAGCTttCATCATCAT ATCTCCCCTCCATCGTTTTACACTCGCTTACTTttCATCATCAT GCGGCATGGTGACCCATGTTCAAGGTAAGGTTCAAGttCATCATCAT ATAGTTCGAAAATTCATAGTGCTCGCCTAGGCTAATGAGAttCATCATCAT GCGGTCGTAGTACCCGGAGCCGTAATGCTGATTGAttCATCATCAT CTCCTCATTGTGGTCACATATTCCTATCCCGCATGCttCATCATCAT ACGACAAAATTTATCCTAGACTGCGGAGTGTGTCTTCTTGttCATCATCAT GCGCGTGGTTGTGGTTGCTGCTCGAATTGTTGTTGttCATCATCAT CTTGTTCTTCATGCGCCGATTCTGGAACCATATCTTGttCATCATCAT TATACAGGAAC TTAGAAGGTATCAAAGGACACGACACGACGttCATCATCAT TTTGCGGAGCTGTGAGGGGTCCGAAAATTGTTAAAttCATCATCAT CTAACTCAGCCGCC TTGTGGCCGCTCTTAATTGGAAttCATCATCAT GGGGTTACCCGAACTCGGATTGGCTAACTGCTAGTttCATCATCAT ATATCAACTACCGAACTAAGCTGCATTATCGTGTTGGGGttCATCATCAT TTGTTCTGCTGATTGGCCTGGCGCTGTGAGTCTTttCATCATCAT GCGCACTTCTGGGCAGTTGCTCATCCACATCCTTGttCATCATCAT CGCTGCTGTTGTCCAAGGGTCACTGGTGCATCTTGttCATCATCAT TTAATTCGAAAGCGACTTCAACTATGGGGTAAACATGCACAAttCATCATCAT CAAGGACCCACTGCTCCCACGGATAATCCACCAGAttCATCATCAT GAAACATACGCATTGAAAAGAACTCCTTCTCCAGCTCCttCATCATCAT AACTTAGACTTAATATCAGGATCAAGCGGCGTCGATACACAAttCATCATCAT AATCTACTACTCGAACGCGTGATTTCATAATTTATGCGGGGttCATCATCAT TAATAATAAACTGCATTTGCATCAACGTCGGTTGGTCACACttCATCATCAT CCCGGATGGCGGGTATATGCAGTGGCGCTAATTGTAAttCATCATCAT AGAACATTTATCTGGGTCTAAAATGTGTTCTGCTGCTCAGCttCATCATCAT GGAATCGCCCTTCCAATCGCTGGACTAGCATGGAAttCATCATCAT AGTCCATGGCCAGCTCGTTCCTTGATTAGATGATGGttCATCATCAT
Abd-B Probe oligo pool	

表 12 続き SABER-FISH のプローブ配列 (2/3)

primer name	sequence
Hairpin (h.30.30)	AAATACTCTCGGGCCTTTTGGCCCGAGAGTATTTGAGAGTATT/3lnvdT/
byn Probe oligo pool	ATTCACCGTTGACGTACTTCCAGCGATGTGAGTCGttAATACTCTC
	AGACTCGGGGTGTACGTAGATGGGATTTGACGGGGttAATACTCTC
	AACCGCATTTCGTGTACAGCTGGTGAGTGGGTGAGGttAATACTCTC
	TATCCGTGGGATGATCCGATCACTGCAGCGTGCTCttAATACTCTC
	ATGTTGTA CTGATGACTGTGTGGCGTCAGGTGATGttAATACTCTC
	TGGTCTTGTGGTCAGCTTTACTTTGGCAAACGAGttAATACTCTC
	TATCCGTGCTCCTTTGGCATCCAGAAAGGCCTTGttAATACTCTC
	ACGCCACTCCGGTTCATAGATATCCGTTTGGGCttAATACTCTC
	GCAAATGGATTATACTTGATCTTCAGCGATGTCACCTCCTCttAATACTCTC
	ATTTGATCCGGATCCGGAACGACCCTGGGTGACttAATACTCTC
	ATTGGAGTTTAGCGCTGCCGCCGATCGTGTGTTGttAATACTCTC
	GACCGGAGGCACTGATCTTCACGACCGGAACATGttAATACTCTC
	AGTTGAATACGCAACACTGCCAAATCACGCACTTTTATttAATACTCTC
	AGATTCTGGAACGCAGCCAGAGCTCCCGATCGTCttAATACTCTC
	TCCAGGGGAATGTATTCCGGCTGGTAATACGCGGCttAATACTCTC
	CAGATTGCGGTCCAGTTCATTGGGACTGCCTGCTCttAATACTCTC
	TTATGACCGAGCCCTGCTCATCGCTCGGCTTGTA CttAATACTCTC
	ATTGCTCTGCCAGCTGCTGGGATAGGAGAACACGGttAATACTCTC
	ATGGGCTCCTTCATCCAGTGGGCTCCAAAATTGGGttAATACTCTC
	GGCGCTCTCGATATAATCCAGTGGTAAACTCCTTATGAAGAttAATACTCTC
	GAGCCAGCCGTAGTGGGTATCGTGTGGATACAGTGttAATACTCTC
	AACATTGCCACTGAGTCCGGTTGTTGGATCCACGGttAATACTCTC
	AATGGATAGGTGACCACGTGACGCTGCTCGGAACCttAATACTCTC
	ATCTGGACGAACTCCAGGAGGACGGTGTACATGGCttAATACTCTC
	AGCACTAATCCCTAGAGGGTTCACTTTATCAAGGATCACTttAATACTCTC
	CGATTGCGATTGGGTGTGCTTTACTAATAGCGTTACTAACAttAATACTCTC
	GATGTCGTGAGAGCTTTCAGGGAGGAACTGGCGCttAATACTCTC
	AAGATAGGGATCTGCCGTAGCGATCGCAACTGCCGttAATACTCTC
	CTGATACGCCGTACCCGCGATAAACTGCGTTTCCGttAATACTCTC
	CATTGTGCCAACCCTGCGGAATAGCTTAGCGCCttAATACTCTC

表 12 続き SABER-FISH のプローブ配列 (3/3)

primer name	Sequence	5' Modification
27.Alexa565	TTATGATGATGTATGATGATGT	Alexa Fluor 555
30.Alexa647	TTGAGAGTATTTGAGAGTATTT	Alexa Fluor 647

引用文献

Adam, M., Potter, A.S., Potter, S.S., 2017. Psychrophilic proteases dramatically reduce single-cell RNA-seq artifacts: A molecular atlas of kidney development. *Dev. Camb.* 144, 3625–3632. <https://doi.org/10.1242/dev.151142>

Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.C., Geurts, P., Aerts, J., Van Den Oord, J., Atak, Z.K., Wouters, J., Aerts, S., 2017. SCENIC: Single-cell regulatory network inference and clustering. *Nat. Methods* 14, 1083–1086. <https://doi.org/10.1038/nmeth.4463>

Alberga, A., Boulay, J.L., Kempe, E., Dennefeld, C., Haenlin, M., 1991. The snail gene required for mesoderm formation in *Drosophila* is expressed dynamically in derivatives of all three germ layers. *Development* 111, 983–992. <https://doi.org/10.1242/dev.111.4.983>

Albright, A.R., Stadler, M.R., Eisen, M.B., 2022. Single-nucleus RNA-sequencing in pre-cellularization *Drosophila melanogaster* embryos. *PLoS ONE* 17, 1–13. <https://doi.org/10.1371/journal.pone.0270471>

Ashton, J.M., Rehrauer, H., Myers, Jason, Myers, Jacqueline, Zanche, M., Balys, M., Foox, J., Mason, C.E., Steen, R., Kuentzel, M., Aquino, C., Garcia-Reyero, N., Chittur, S.V., 2021. Comparative Analysis of Single-Cell RNA Sequencing Platforms and Methods. *J. Biomol. Tech.* JBT 32, 3fc1f5fe.3eccea01. <https://doi.org/10.7171/3fc1f5fe.3eccea01>

Bailles, A., Collinet, C., Philippe, J.-M., Lenne, P.-F., Munro, E., Lecuit, T., 2019. Genetic induction and mechanochemical propagation of a morphogenetic wave. *Nature* 572, 467–473. <https://doi.org/10.1038/s41586-019-1492-9>

Bakken, T.E., Hodge, R.D., Miller, J.A., Yao, Z., Nguyen, T.N., Aevermann, B., Barkan, E., Bertagnolli, D., Casper, T., Dee, N., Garren, E., Goldy, J., Graybuck, L.T., Kroll, M., Lasken, R.S., Lathia, K., Parry, S., Rimorin, C., Scheuermann, R.H., Schork, N.J., Shehata, S.I., Tieu, M., Phillips, J.W., Bernard, A., Smith, K.A., Zeng, H., Lein, E.S., Tasic, B., 2018. Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. *PLOS ONE* 13, e0209648. <https://doi.org/10.1371/journal.pone.0209648>

Basile, G., Kahraman, S., Dirice, E., Pan, H., Dreyfuss, J.M., Kulkarni, R.N., 2021. Using single-nucleus RNA-sequencing to interrogate transcriptomic profiles of archived human pancreatic islets. *Genome Med.* 13, 128. <https://doi.org/10.1186/s13073-021-00941-8>

Beliveau, B.J., Kishi, J.Y., Nir, G., Sasaki, H.M., Saka, S.K., Nguyen, S.C., Wu, C.-T., Yin, P., 2018. OligoMiner provides a rapid, flexible environment for the design of genome-scale

oligonucleotide in situ hybridization probes. *Proc. Natl. Acad. Sci. U. S. A.* 115, E2183–E2192. <https://doi.org/10.1073/pnas.1714530115>

Berleth, T., Burri, M., Thoma, G., Bopp, D., Riehlstein, S., Frigerio, G., Noll, M., Nüsslein-Volhard, C., 1988. The role of localization of bicoid RNA in organizing the anterior pattern of the *Drosophila* embryo. *EMBO J.* 7, 1749–1756. <https://doi.org/10.1002/j.1460-2075.1988.tb03004.x>

Bertet, C., Sulak, L., Lecuit, T., 2004. Myosin-dependent junction remodelling controls planar cell intercalation and axis elongation. *Nature*. <https://doi.org/10.1038/nature02590>

Briggs, J.A., Weinreb, C., Wagner, D.E., Megason, S., Peshkin, L., Kirschner, M.W., Klein, A.M., 2018. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* 360. <https://doi.org/10.1126/science.aar5780>

Briscoe, J., Small, S., 2015. Morphogen rules: Design principles of gradient-mediated embryo patterning. *Dev. Camb.* 142, 3996–4009. <https://doi.org/10.1242/dev.129452>

Calderon, D., Blecher-Gonen, R., Huang, X., Secchia, S., Kentro, J., Daza, R.M., Martin, B., Dulja, A., Schaub, C., Trapnell, C., Larschan, E., O'Connor-Giles, K.M., Furlong, E.E.M., Shendure, J., 2022. The continuum of *Drosophila* embryonic development at single-cell resolution. *Science* 377, 1–40. <https://doi.org/10.1126/science.abn5800>

Cammarota, C., Finegan, T.M., Wilson, T.J., Yang, S., Bergstrahl, D.T., 2020. An Axon-Pathfinding Mechanism Preserves Epithelial Tissue Integrity. *Curr. Biol. CB* 30, 5049-5057.e3. <https://doi.org/10.1016/j.cub.2020.09.061>

Casanova, J., Struhl, G., 1989. Localized surface activity of torso, a receptor tyrosine kinase, specifies terminal body pattern in *Drosophila*. *Genes Dev.* 3, 2025–2038. <https://doi.org/10.1101/gad.3.12b.2025>

Chen, S., Mar, J.C., 2018. Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data. *BMC Bioinformatics* 19, 232. <https://doi.org/10.1186/s12859-018-2217-z>

Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinforma. Oxf. Engl.* 34, i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>

Clark, E., Akam, M., 2016. Odd-paired controls frequency doubling in *Drosophila* segmentation by altering the pair-rule gene regulatory network. *eLife* 5, 1–42. <https://doi.org/10.7554/eLife.18215>

Cowden, J., Levine, M., 2002. The Snail repressor positions Notch signaling in the *Drosophila* embryo. *Development* 129, 1785–1793. <https://doi.org/10.1242/dev.129.7.1785>

Del Álamo, D., Rouault, H., Schweisguth, F., 2011. Mechanism and significance of cis-inhibition in notch signalling. *Curr. Biol.* 21, 40–47. <https://doi.org/10.1016/j.cub.2010.10.034>

Ding, X.B., Jin, J., Tao, Y.T., Guo, W.P., Ruan, L., Yang, Q.L., Chen, P.C., Yao, H., Zhang, H.B., Chen, X., 2020. Predicted *Drosophila* Interactome Resource and web tool for functional interpretation of differentially expressed genes. *Database J. Biol. Databases Curation* 2020, 1–11. <https://doi.org/10.1093/database/baaa005>

Driever, W., Nüsslein-Volhard, C., 1988. The bicoid protein determines position in the *Drosophila* embryo in a concentration-dependent manner. *Cell* 54, 95–104.

Farrell, J.A., Wang, Y., Riesenfeld, S.J., Shekhar, K., Regev, A., Schier, A.F., 2018. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* 360. <https://doi.org/10.1126/science.aar3131>

Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., Slichter, C.K., Miller, H.W., McElrath, M.J., Prlic, M., Linsley, P.S., Gottardo, R., 2015. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* 16, 278. <https://doi.org/10.1186/s13059-015-0844-5>

Finklstein, R., Perrimon, N., 1990. The orthodenticle gene is regulated by bicoid and torso and specifies *Drosophila* head development. *Nature* 346, 485–488. <https://doi.org/10.1038/346485a0>

Fowlkes, C.C., Hendriks, C.L.L., Keränen, S.V.E., Weber, G.H., Rübél, O., Huang, M.-Y., Chatoor, S., DePace, A.H., Simirenko, L., Henriquez, C., Beaton, A., Weizmann, R., Celniker, S., Hamann, B., Knowles, D.W., Biggin, M.D., Eisen, M.B., Malik, J., 2008. A Quantitative Spatiotemporal Atlas of Gene Expression in the *Drosophila* Blastoderm. *Cell* 133, 364–374. <https://doi.org/10.1016/j.cell.2008.01.053>

Frohnhofer, H.G., Nüsslein-Volhard, C., 1986. Organization of anterior pattern in the *Drosophila* embryo by the maternal gene bicoid. *Nature* 324, 120–125. <https://doi.org/10.1038/324120a0>

Ganguly, A., Jiang, J., Ip, Y.T., 2005. *Drosophila* WntD is a target and an inhibitor of the Dorsal / Twist / Snail network in the gastrulating embryo 3419–3429. <https://doi.org/10.1242/dev.01903>

Gavis, E.R., Lehmann, R., 1992. Localization of nanos RNA controls embryonic polarity. *Cell* 71, 301–313. [https://doi.org/10.1016/0092-8674\(92\)90358-J](https://doi.org/10.1016/0092-8674(92)90358-J)

Gilmour, D., Rembold, M., Leptin, M., 2017. From morphogen to morphogenesis and back. *Nature* 541, 311–320. <https://doi.org/10.1038/nature21348>

Graham, P.L., Anderson, W.R., Brandt, E.A., Xiang, J., Pick, L., 2019. Dynamic expression of *Drosophila* segmental cell surface-encoding genes and their pair-rule regulators. *Dev. Biol.* 447, 147–156. <https://doi.org/10.1016/j.ydbio.2019.01.015>

Hammonds, A.S., Bristow, C.A., Fisher, W.W., Weizmann, R., Wu, S., Hartenstein, V., Kellis, M., Yu, B., Frise, E., Celniker, S.E., 2013. Spatial expression of transcription factors in *Drosophila*

embryonic organ development. *Genome Biol.* 14, R140. <https://doi.org/10.1186/gb-2013-14-12-r140>

Heemskerk, J., DiNardo, S., 1994. *Drosophila* hedgehog acts as a morphogen in cellular patterning. *Cell* 76, 449–460. [https://doi.org/10.1016/0092-8674\(94\)90110-4](https://doi.org/10.1016/0092-8674(94)90110-4)

Heimberg, G., Bhatnagar, R., El-Samad, H., Thomson, M., 2016. Low Dimensionality in Gene Expression Data Enables the Accurate Extraction of Transcriptional Programs from Shallow Sequencing. *Cell Syst.* 2, 239–250. <https://doi.org/10.1016/j.cels.2016.04.001>

Hinck, L., 2004. The versatile roles of “axon guidance” cues in tissue morphogenesis. *Dev. Cell* 7, 783–93. <https://doi.org/10.1016/j.devcel.2004.11.002>

Hu, Y., Comjean, A., Perkins, L.A., Perrimon, N., Mohr, S.E., 2015. GLAD: an Online Database of Gene List Annotation for *Drosophila*. *J. Genomics* 3, 75–81. <https://doi.org/10.7150/jgen.12863>

Hunter, J.D., 2007. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* 9, 90–95. <https://doi.org/10.1109/MCSE.2007.55>

Ingham, P.W., Arias, A.M., 1992. Boundaries and fields in early embryos. *Cell* 68, 221–235. [https://doi.org/10.1016/0092-8674\(92\)90467-Q](https://doi.org/10.1016/0092-8674(92)90467-Q)

Irvine, K.D., Wieschaus, E., 1994. Cell intercalation during *Drosophila* germband extension and its regulation by pair-rule segmentation genes. *Development* 120, 827–841.

Islam, S., Zeisel, A., Joost, S., Manno, G.L., Zajac, P., Kasper, M., Lönnerberg, P., Linnarsson, S., 2014. Quantitative single-cell RNA-seq with unique molecular identifiers 11. <https://doi.org/10.1038/nmeth.2772>

Jack, T., McGinnis, W., 1990. Establishment of the Deformed expression stripe requires the combinatorial action of coordinate, gap and pair-rule proteins. *EMBO J.* 9, 1187–1198. <https://doi.org/10.1002/j.1460-2075.1990.tb08226.x>

Janssens, J., Aibar, S., Taskiran, I.I., Ismail, J.N., Gomez, A.E., Aughey, G., Spanier, K.I., De Rop, F.V., González-Blas, C.B., Dionne, M., Grimes, K., Quan, X.J., Pappasokrati, D., Hulselmans, G., Makhzami, S., De Waegeneer, M., Christiaens, V., Southall, T., Aerts, S., 2022. Decoding gene regulation in the fly brain. *Nature* 601, 630–636. <https://doi.org/10.1038/s41586-021-04262-z>

Kaminow, B., Yunusov, D., Dobin, A., Spring, C., 2021. STARsolo : accurate , fast and versatile mapping / quantification of single-cell and single-nucleus RNA-seq data 1–35.

Karaiskos, N., Wahle, P., Alles, J., Boltengagen, A., Ayoub, S., Kipar, C., Kocks, C., Rajewsky, N., Zinzen, R.P., 2017. The *Drosophila* embryo at single-cell transcriptome resolution. *Science* 358, 194–199. <https://doi.org/10.1126/science.aan3235>

Keleman, K., Rajagopalan, S., Cleppien, D., Teis, D., Paiha, K., Huber, L.A., Technau, G.M., Dickson, B.J., 2002. Comm sorts robo to control axon guidance at the *Drosophila* midline. *Cell* 110,

415–27. [https://doi.org/10.1016/s0092-8674\(02\)00901-7](https://doi.org/10.1016/s0092-8674(02)00901-7)

Keleman, K., Ribeiro, C., Dickson, B.J., 2005. Comm function in commissural axon guidance: cell-autonomous sorting of Robo in vivo. *Nat. Neurosci.* 8, 156–63. <https://doi.org/10.1038/nn1388>

Kishi, J.Y., Lapan, S.W., Beliveau, B.J., West, E.R., Zhu, A., Sasaki, H.M., Saka, S.K., Wang, Y., Cepko, C.L., Yin, P., 2019. SABER amplifies FISH: enhanced multiplexed imaging of RNA and DNA in cells and tissues. *Nat. Methods* 16, 533–544. <https://doi.org/10.1038/s41592-019-0404-0>

Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C., Teichmann, S.A., 2015. The Technology and Biology of Single-Cell RNA Sequencing. *Mol. Cell* 58, 610–620. <https://doi.org/10.1016/j.molcel.2015.04.005>

Kondo, T., Hayashi, S., 2019. Two-step regulation of trachealess ensures tight coupling of cell fate with morphogenesis in the drosophila trachea. *eLife* 8, 1–23. <https://doi.org/10.7554/eLife.45145>

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P. ru, Raychaudhuri, S., 2019. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* 16, 1289–1296. <https://doi.org/10.1038/s41592-019-0619-0>

Krueger, D., Tardivo, P., Nguyen, C., De Renzis, S., 2018. Downregulation of basal myosin-II is required for cell shape changes and tissue invagination. *EMBO J.* 37, 1–16. <https://doi.org/10.15252/emj.2018100170>

Leptin, M., Grunewald, B., 1990. Cell shape changes during gastrulation in *Drosophila*. *Development* 110, 73–84.

Letsou, W., Cai, L., 2016. Noncommutative Biology: Sequential Regulation of Complex Networks. *PLOS Comput. Biol.* 12, e1005089. <https://doi.org/10.1371/journal.pcbi.1005089>

Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323. <https://doi.org/10.1186/1471-2105-12-323>

Liu, W., Morgan, K.M., Pine, S.R., 2014. Activation of the Notch1 Stem Cell Signaling Pathway during Routine Cell Line Subculture. *Front. Oncol.* 4, 1–4. <https://doi.org/10.3389/fonc.2014.00211>

Long, H.K., Prescott, S.L., Wysocka, J., 2016. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* 167, 1170–1187. <https://doi.org/10.1016/j.cell.2016.09.018>

Luengo Hendriks, C.L., Keränen, S.V.E., Fowlkes, C.C., Simirenko, L., Weber, G.H., DePace, A.H., Henriquez, C., Kaszuba, D.W., Hamann, B., Eisen, M.B., Malik, J., Sudar, D., Biggin, M.D., Knowles, D.W., 2006. Three-dimensional morphology and gene expression in the *Drosophila* blastoderm at cellular resolution I: Data acquisition pipeline. *Genome Biol.* 7.

<https://doi.org/10.1186/gb-2006-7-12-r123>

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., Trombetta, J.J., Weitz, D.A., Sanes, J.R., Shalek, A.K., Regev, A., McCarroll, S.A., 2015. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* 161, 1202–1214. <https://doi.org/10.1016/j.cell.2015.05.002>

Manning, A.J., Peters, K.A., Peifer, M., Rogers, S.L., 2013. Regulation of Epithelial Morphogenesis by the G Protein–Coupled Receptor *Mist* and Its Ligand *Fog*. *Sci. Signal.* 6, 1–11. <https://doi.org/10.1126/scisignal.2004427>

Martin, A.C., Kaschube, M., Wieschaus, E.F., 2009. Pulsed contractions of an actin–myosin network drive apical constriction. *Nature* 457, 495–499. <https://doi.org/10.1038/nature07522>

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10. <https://doi.org/10.14806/ej.17.1.200>

Martinez-Arias, A., Lawrence, P.A., 1985. Parasegments and compartments in the *Drosophila* embryo. *Nature* 313, 639–642. <https://doi.org/10.1038/313639a0>

Morel, V., Schweisguth, F., 2000. Repression by suppressor of hairless and activation by Notch are required to define a single row of single-minded expressing cells in the *Drosophila* embryo. *Genes Dev.* 14, 377–88.

Moriel, N., Senel, E., Friedman, N., Rajewsky, N., Karaikos, N., Nitzan, M., 2021. NovoSpaRc: flexible spatial reconstruction of single-cell gene expression with optimal transport. *Nat. Protoc.* 16, 4177–4200. <https://doi.org/10.1038/s41596-021-00573-7>

Nüsslein-Volhard, C., Wieschaus, E., 1980. Mutations affecting segment number and polarity in *Drosophila*. *Nature* 287, 795–801. <https://doi.org/10.1038/287795a0>

O’Flanagan, C.H., Campbell, K.R., Zhang, A.W., Kabeer, F., Lim, J.L.P., Biele, J., Eirew, P., Lai, D., McPherson, A., Kong, E., Bates, C., Borkowski, K., Wiens, M., Hewitson, B., Hopkins, J., Pham, J., Ceglia, N., Moore, R., Mungall, A.J., McAlpine, J.N., Shah, S.P., Aparicio, S., 2019. Dissociation of solid tumor tissues with cold active protease for single-cell RNA-seq minimizes conserved collagenase-associated stress responses. *Genome Biol.* 20, 210. <https://doi.org/10.1186/s13059-019-1830-0>

Okochi, Y., Sakaguchi, S., Nakae, K., Kondo, T., Naoki, H., 2021. Model-based prediction of spatial gene expression via generative linear mapping. *Nat. Commun.* 12, 1–13. <https://doi.org/10.1038/s41467-021-24014-x>

Packer, J.S., Zhu, Q., Huynh, C., Sivaramkrishnan, P., Preston, E., Dueck, H., Stefanik, D., Tan, K., Trapnell, C., Kim, J., Waterston, R.H., Murray, J.I., 2019. A lineage-resolved molecular atlas of

C. elegans embryogenesis at single-cell resolution. *Science* 365.

<https://doi.org/10.1126/science.aax1971>

Paré, A.C., Naik, P., Shi, J., Mirman, Z., Palmquist, K.H., Zallen, J.A., 2019. An LRR Receptor-Teneurin System Directs Planar Polarity at Compartment Boundaries. *Dev. Cell* 51, 208-221.e6.

<https://doi.org/10.1016/j.devcel.2019.08.003>

Paré, A.C., Vichas, A., Fincher, C.T., Mirman, Z., Farrell, D.L., Mainieri, A., Zallen, J.A., 2014. A positional Toll receptor code directs convergent extension in *Drosophila*. *Nature* 515, 523–527.

<https://doi.org/10.1038/nature13953>

Paroush, Z., Mark Wainwright, S., Ish-Horowicz, D., 1997. Torso signalling regulates terminal patterning in *Drosophila* by antagonising Groucho-mediated repression. *Development* 124, 3827–3834.

Petkova, M.D., Tkačik, G., Bialek, W., Wieschaus, E.F., Gregor, T., 2019. Optimal Decoding of Cellular Identities in a Genetic Network. *Cell* 176, 844-855.e15.

<https://doi.org/10.1016/j.cell.2019.01.007>

Rahimi, N., Averbukh, I., Haskel-Ittah, M., Degani, N., Schejter, E.D., Barkai, N., Shilo, B.-Z., 2016. A WntD-Dependent Integral Feedback Loop Attenuates Variability in *Drosophila* Toll Signaling. *Dev. Cell* 36, 401–414. <https://doi.org/10.1016/j.devcel.2016.01.023>

Rauzi, M., Lenne, P.-F., Lecuit, T., 2010. Planar polarized actomyosin contractile flows control epithelial junction remodelling. *Nature* 468, 1110–1114. <https://doi.org/10.1038/nature09566>

Reeves, G.T., Stathopoulos, A., 2009. Graded Dorsal and Differential Gene Regulation in the *Drosophila* Embryo. *Cold Spring Harb. Perspect. Biol.* 1, a000836–a000836.

<https://doi.org/10.1101/cshperspect.a000836>

Reuter, R., Grunewald, B., Leptin, M., 1993. A role for the mesoderm in endodermal migration and morphogenesis in *Drosophila*. *Dev. Camb. Engl.* 119, 1135–45.

<https://doi.org/10.1242/dev.119.4.1135>

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinforma. Oxf. Engl.* 26, 139–40.

<https://doi.org/10.1093/bioinformatics/btp616>

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., Regev, A., 2015. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* 33, 495–502. <https://doi.org/10.1038/nbt.3192>

Schaerlinger, B., Launay, J.M., Vonesch, J.I., Maroteaux, L., 2007. Gain of affinity point mutation in the serotonin receptor gene 5-HT 2Dro accelerates germband extension movements during *Drosophila* gastrulation. *Dev. Dyn.* 236, 991–999. <https://doi.org/10.1002/dvdy.21110>

Staller, M.V., Fowlkes, C.C., Bragdon, M.D.J., Wunderlich, Z., Estrada, J., DePace, A.H., 2015. A gene expression atlas of a bicoid -depleted *Drosophila* embryo reveals early canalization of cell fate. *Development* 142, 587–596. <https://doi.org/10.1242/dev.117796>

Stedden, C.G., Menegas, W., Zajac, A.L., Williams, A.M., Cheng, S., Özkan, E., Horne-Badovinac, S., 2019. Planar-Polarized Semaphorin-5c and Plexin A Promote the Collective Migration of Epithelial Cells in *Drosophila*. *Curr. Biol.* CB 29, 908-920.e6. <https://doi.org/10.1016/j.cub.2019.01.049>

Stern, T., Shvartsman, S.Y., Wieschaus, E.F., 2022. Deconstructing gastrulation at single-cell resolution. *Curr. Biol.* CB 32, 1861–1868. <https://doi.org/10.1016/j.cub.2022.02.059>

Stricker, S.H., Köferle, A., Beck, S., 2016. From profiles to function in epigenomics. *Nat. Rev. Genet.* 18, 51–66. <https://doi.org/10.1038/nrg.2016.138>

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., Satija, R., 2019. Comprehensive Integration of Single-Cell Data. *Cell* 177, 1888-1902.e21. <https://doi.org/10.1016/j.cell.2019.05.031>

Sweeton, D., Parks, S., Costa, M., Wieschaus, E., 1991. Gastrulation in *Drosophila*: The formation of the ventral furrow and posterior midgut invaginations. *Development* 112, 775–789.

Tanay, A., Regev, A., 2017. Scaling single-cell genomics from phenomenology to mechanism. *Nature* 541, 331–338. <https://doi.org/10.1038/nature21350>

Tetley, R.J., Blanchard, G.B., Fletcher, A.G., Adams, R.J., Sanson, B., 2016. Unipolar distributions of junctional myosin II identify cell stripe boundaries that drive cell intercalation throughout *drosophila* axis extension. *eLife* 5, 1–28. <https://doi.org/10.7554/eLife.12094>

Thisse, B., Stoetzel, C., Gorostiza-Thisse, C., Perrin-Schmitt, F., 1988. Sequence of the twist gene and nuclear localization of its protein in endomesodermal cells of early *Drosophila* embryos. *EMBO J.* 7, 2175–2183. <https://doi.org/10.1002/j.1460-2075.1988.tb03056.x>

Tomancak, P., Beaton, A., Weizmann, R., Kwan, E., Shu, S., Lewis, S.E., Richards, S., Ashburner, M., Hartenstein, V., Celniker, S.E., Rubin, G.M., 2002. Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.* 3, RESEARCH0088. <https://doi.org/10.1186/gb-2002-3-12-research0088>

Tomancak, P., Berman, B.P., Beaton, A., Weizmann, R., Kwan, E., Hartenstein, V., Celniker, S.E., Rubin, G.M., 2007. Global analysis of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.* 8, R145. <https://doi.org/10.1186/gb-2007-8-7-r145>

Vaughen, J., Igaki, T., 2016. Slit-Robo Repulsive Signaling Extrudes Tumorigenic Cells from Epithelia. *Dev. Cell* 39, 683–695. <https://doi.org/10.1016/j.devcel.2016.11.015>

Vincent, A., Blankenship, J.T., Wieschaus, E., 1997. Integration of the head and trunk segmentation systems controls cephalic furrow formation in *Drosophila*. *Development* 124, 3747–3754.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., Vijaykumar, A., Bardelli, A.P., Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C.N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D.A., Hagen, D.R., Pasechnik, D.V., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G.A., Ingold, G.-L., Allen, G.E., Lee, G.R., Audren, H., Probst, I., Dietrich, J.P., Silterra, J., Webber, J.T., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger, J.L., de Miranda Cardoso, J.V., Reimer, J., Harrington, J., Rodríguez, J.L.C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N.J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P.A., Lee, P., McGibbon, R.T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T.J., Robitaille, T.P., Spura, T., Jones, T.R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y.O., Vázquez-Baeza, Y., 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>

Wang, Y.C., Khan, Z., Kaschube, M., Wieschaus, E.F., 2012. Differential positioning of adherens junctions is associated with initiation of epithelial folding. *Nature* 484, 390–393. <https://doi.org/10.1038/nature10938>

Weigel, D., Jürgens, G., Klingler, M., Jäckle, H., 1990. Two gap genes mediate maternal terminal pattern information in *Drosophila*. *Science* 248, 495–498. <https://doi.org/10.1126/science.2158673>

Wu, Y.E., Pan, L., Zuo, Y., Li, X., Hong, W., 2017. Detecting Activated Cell Populations Using Single-Cell RNA-Seq. *Neuron* 96, 313-329.e6. <https://doi.org/10.1016/j.neuron.2017.09.026>

Yazdani, U., Terman, J.R., 2006. The semaphorins. *Genome Biol.* 7, 211. <https://doi.org/10.1186/gb-2006-7-3-211>

Yoo, S.K., Pascoe, H.G., Pereira, T., Kondo, S., Jacinto, A., Zhang, X., Hariharan, I.K., 2016. Plexins function in epithelial repair in both *Drosophila* and zebrafish. *Nat. Commun.* 7, 12282. <https://doi.org/10.1038/ncomms12282>

Zallen, J.A., Wieschaus, E., 2004. Patterned Gene Expression Directs Bipolar Planar Polarity in *Drosophila*. *Dev. Cell* 6, 343–355.

Zhang, M.J., Ntranos, V., Tse, D., 2020. Determining sequencing depth in a single-cell RNA-seq

experiment. Nat. Commun. 11, 774. <https://doi.org/10.1038/s41467-020-14482-y>

Zinzen, R.P., Cande, J., Ronshaugen, M., Papatsenko, D., Levine, M., 2006. Evolution of the Ventral Midline in Insect Embryos. Dev. Cell 11, 895–902. <https://doi.org/10.1016/j.devcel.2006.10.012>

注釈

本研究は複数の共同研究者とともに行った。本研究において、scRNA-seqデータの取得は近藤武史博士と共同で行った。また、バルクRNA-seqの実行と解析は近藤武史博士が行った。加えて、図 24のFISH実験と画像の取得は水野苑子さんが行った。

謝辞

本研究に関して、上村匡教授には私が学部生の頃から非常に熱心な指導をしていただきました。ラボセミナーやミーティングでの議論を介して、研究者としての姿勢を学ばせていただきました。近藤武史博士には、実験手法、データの取り扱い、文書の作成、研究発表のしかたなど研究活動を行うにあたって必要な多くの事項を指導していただきました。また、日々の研究の結果に関して、熱心かつ綿密な議論をしていただき、そのうえで研究計画について数々の助言をいただきました。碓井理夫博士と坪井有寿博士、春本敏之博士にはグループミーティングにおいて本研究に関する議論をしていただきました。二股真由美さん、三木雅代さんには、実験の補助をしていただきました。沖かなえさん、森口良子さん、今井博子さんには教務関連の補佐をしていただきました。また、他の上村研究室のメンバーの方々にも多くの有意義な助言をいただきました。以上の上村研究室の皆様の手助けがなければ、本研究を遂行することはできませんでした。心より感謝申し上げます。

本研究を行うにあたり、上村研究室以外の方々にもお世話になりましたのでお名前を上げさせていただきます。本研究で用いた空間再構成手法 Perler は広島大学の本田直樹教授と京都大学医学部附属病院の大河内康史医師との共同研究によるものであり、今回の適用

に関しても議論と助言をしていただきました。理化学研究所生命機能科学研究センターの種子島千春さん、西村理さん、門田満隆さんには scRNA-seq 手法の開発の手助けをしていただきました。井垣研究室の山銅ゆかりさんには、qPCR によるライブラリの定量方法、次世代シーケンサーの使用法に関して指導をしていただきました。Fluidigm 社の甲斐渉さんには、C1HT のプライマー設計について助言をいただきました。理化学研究所生命機能科学研究センターの Yu-Chiun Wang 博士にも本研究に関して助言をいただきました。皆様に厚くお礼申し上げます。

また、本研究は 2020 年 4 月から 2023 年 3 月までの期間、日本学術振興会の特別研究員 DC1 としての支援を受けて行いました。

最後に、支えてくださった家族や友人にも感謝の言葉を述べさせていただきます。本当にありがとうございました。

本学位論文は以下の学術論文の内容に基づいて書かれたものです。

Shunta Sakaguchi, Sonoko Mizuno, Yasushi Okochi, Chiharu Tanegashima, Osamu Nishimura, Tadashi Uemura, Mitsutaka Kadota, Honda Naoki and Takefumi Kondo

Single-cell transcriptome atlas of *Drosophila* gastrula 2.0

Cell Reports, in press, 2023

坂口 峻太