

---

Métodos Computacionais para Análise e  
Caracterização de Imagens de Embriões da  
*Drosophila melanogaster*

---

Daniela Justiniano de Sousa



UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
FACULDADE DE COMPUTAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Uberlândia  
2020

**Daniela Justiniano de Sousa**

**Métodos Computacionais para Análise e  
Caracterização de Imagens de Embriões da  
*Drosophila melanogaster***

Tese de doutorado apresentada ao Programa de Pós-graduação da Faculdade de Computação da Universidade Federal de Uberlândia como parte dos requisitos para a obtenção do título de Doutor em Ciência da Computação.

Área de concentração: Ciência da Computação

Orientador: Prof. Dr. Bruno Augusto Nassif Travençolo  
Coorientador: Prof. Dr. Francisco José Pereira Lopes

Uberlândia  
2020

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU  
com dados informados pelo(a) próprio(a) autor(a).

P436	Pereira, Daniela Justiniano de Sousa, 1985-
2020	Métodos Computacionais para Análise e Caracterização de Imagens de Embriões da <i>Drosophila melanogaster</i> [recurso eletrônico] / Daniela Justiniano de Sousa Pereira. - 2020.
<p>Orientador: Bruno Augusto Nassif Travençolo. Coorientador: Francisco José Pereira Lopes. Tese (Doutorado) - Universidade Federal de Uberlândia, Pós-graduação em Ciência da Computação. Modo de acesso: Internet. Disponível em: <a href="http://doi.org/10.14393/ufu.te.2021.1">http://doi.org/10.14393/ufu.te.2021.1</a> Inclui bibliografia. Inclui ilustrações.</p>	
<p>1. Computação. I. Travençolo, Bruno Augusto Nassif, 1981-, (Orient.). II. Lopes, Francisco José Pereira, 1968-, (Coorient.). III. Universidade Federal de Uberlândia. Pós-graduação em Ciência da Computação. IV. Título.</p>	
CDU: 681.3	

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:

Gizele Cristine Nunes do Couto - CRB6/2091



## UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Coordenação do Programa de Pós-Graduação em Ciência da Computação

Av. João Naves de Ávila, nº 2121, Bloco 1A, Sala 243 - Bairro Santa Mônica,  
Uberlândia-MG, CEP 38400-902  
Telefone: (34) 3239-4470 - [www.ppgco.facom.ufu.br](http://www.ppgco.facom.ufu.br) - [cpgfacom@ufu.br](mailto:cpgfacom@ufu.br)



### ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Ciência da Computação			
Defesa de:	Tese de doutorado, 36/2020, PPGCO			
Data:	15 de dezembro de 2020	Hora de início:	08:30	Hora de encerramento:
Matrícula do Discente:	11513CCP001			
Nome do Discente:	Daniela Justiniano de Sousa			
Título do Trabalho:	Métodos computacionais para análise e caracterização de imagens de embriões da <i>Drosophila melanogaster</i>			
Área de concentração:	Ciência da Computação			
Linha de pesquisa:	Ciência de Dados			
Projeto de Pesquisa de vinculação:	-			

Reuniu-se, por videoconferência, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Ciência da Computação, assim composta: Professores Doutores: André Ricardo Backes - FACOM/UFU, Marcelo Emílio Beletti - ICBIM/UFU, Leandro Alves Neves - IBILCE/UNESP, Andrea Gomes Campos Bianchi - DECOM/UFOP, Francisco José Pereira Lopes - UFRJ (coorientador) e Bruno Augusto Nassif Travençolo - FACOM/UFU, orientador da candidata.

Os examinadores participaram desde as seguintes localidades: Leandro Alves Neves - São José do Rio Preto/SP; Andrea Gomes Campos Bianchi - Ouro Preto/MG; Francisco José Pereira Lopes - Duque de Caxias/RJ; André Ricardo Backes, Marcelo Emílio Beletti, Bruno Augusto Nassif Travençolo - Uberlândia/MG. A discente participou da cidade de Uberlândia/MG.

Iniciando os trabalhos o presidente da mesa, Prof. Dr. Bruno Augusto Nassif Travençolo, apresentou a Comissão Examinadora e a candidata, agradeceu a presença do público, e concedeu à Discente a palavra para a exposição do seu trabalho. A duração da apresentação da Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir a candidata. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando a candidata:

**Aprovada.**

Esta defesa faz parte dos requisitos necessários à obtenção do título de Doutor.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Andrea Gomes Campos Bianchi**, **Usuário Externo**, em 15/12/2020, às 11:25, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **André Ricardo Backes**, **Professor(a) do Magistério Superior**, em 15/12/2020, às 11:25, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Francisco Jose Pereira Lopes**, **Usuário Externo**, em 15/12/2020, às 11:25, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Bruno Augusto Nassif Travençolo, Professor(a) do Magistério Superior**, em 15/12/2020, às 11:30, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Marcelo Emílio Beletti**, **Professor(a) do Magistério Superior**, em 15/12/2020, às 11:30, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Leandro Alves Neves, Usuário Externo**, em 15/12/2020, às 11:30, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [https://www.sei.ufu.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **2455308** e o código CRC **7EBB11B1**.

*À minha amada mãe Nilza, que apoiou-me incondicionalmente desde o início.  
Sem a sua ajuda não seria possível eu concluir esse trabalho.*

---

## Agradecimentos

Primeiramente a Deus, por ter me dado a graça de concluir o doutorado e realizar esse grande sonho. Por me guiar, proteger e iluminar em cada etapa desde o início. Obrigada Pai por fortalecer-me nos momentos difíceis e por colocar em meu caminho as melhores pessoas. O Senhor é bom e a Sua misericórdia dura para sempre. A Ti meu Senhor, toda honra e toda glória!

Ao meu orientador Prof. Dr. Bruno Augusto Nassif Travençolo, agradeço a orientação, ensino, confiança e paciência dedicada ao longo desses anos. Obrigado por acreditar em mim e por todo o respeito e empatia com que sempre me tratou. As lições que recebi de ti vão além do âmbito profissional. Tenho certeza que não chegaria neste ponto sem o seu apoio. Você é muito mais que orientador, para mim será sempre mestre e amigo.

Ao meu coorientador Prof. Dr. Francisco José Pereira Lopes, agradeço a orientação, dedicação e conhecimento compartilhado na interpretação e análise dos dados biológicos. Gratidão pela prestatividade, estímulo e simpatia sempre dispensada. Desejo poder contribuir à ciência e ao universo acadêmico com o mesmo entusiasmo que me transmitiu.

À equipe da UFRJ (Prof. Helena, Maíra, Paulo e Cláudio) agradeço por todo o conhecimento compartilhado, pela disponibilidade, interesse e atenção que sempre manifestaram. Em especial a Prof. Helena e a Maíra pela colaboração no desenvolvimento dos artigos e disponibilização das imagens biológicas.

Aos professores André Backes e Leandro Alves, pela leitura que fizeram da monografia e participação na minha banca de qualificação. Todos seus apontamentos foram primordiais para chegar a essa versão final da Tese.

Agradeço todos os professores e professoras que passaram por minha vida e contribuíram positivamente com a minha trajetória. Particularmente eu cito o Sr. Sílvio e o Prof. André Augusto, verdadeiros mestres que motivaram-me a acreditar em um futuro com possibilidades. Embora tenhamos traçados caminhos diferentes, ficou a admiração e carinho.

Carinhosamente, agradeço a TODOS os professores da FACOM/UFU com os quais eu tive o prazer de cursar disciplinas. Difícil citar o nome de todos, mas o profissionalismo,

dedicação e amor que sempre demonstraram ao exercerem as suas profissões marcaram-me positivamente. Com alguns, mais tarde eu pude dividir bancas, trocar materiais e compartilhar disciplinas. Um sonho realizado, quanto aprendizado e realização. De modo especial eu cito o Prof. Autran, que genuinamente ajudou-me nessa fase final tão decisiva. Sua colaboração foi fundamental para a concretização desse trabalho.

A minha amada mãe Nilza, minha referência e porto seguro. Agradeço por ter sido a primeira a acreditar em mim e apoiar-me incondicionalmente. A senhora que esteve presente desde a graduação e nos últimos anos deixou a sua casa para viver conosco. Não mediou esforços e se doou integralmente. Gratidão por cuidar tão bem de nós. Foi mãe para as minhas filhas, suporte para o meu esposo e o esteio do meu lar. Ouviu pacientemente as minhas angústias e sempre me acolheu com palavras de esperança. Sem a senhora essa vitória não seria possível. Formamos!

Ao meu esposo Lyncool, agradeço o amor, paciência e dedicação. Gratidão por compreender as minhas ausências e apoiar-me incondicionalmente. Foram intensos momentos compartilhados ao longo dessa trajetória, alguns difíceis e outros memoráveis, e em todos eu pude contar com a sua presença e carinho.

As principais razões de minha vida, minhas filhas Maria Clara e Maria Isabel agradeço pela compreensão e paciência que, embora não tivessem conhecimento disto, sempre demonstraram ter. Quantas vezes precisaram dividir o colo da mamãe com o notebook. Gratidão pelo amor que transbordaram em meus dias, tornando tudo mais doce e leve. Por serem a presença constante de Deus em minha vida e darem sentido à realização desse sonho. Essa conquista é por vocês.

Aos meus irmãos (Edney e Eliton), agradeço a torcida e carinho que, mesmo à distância, sempre me ofereceram. Obrigada por compreenderem a minha ausência e fazerem o possível para estarmos sempre juntos. A você Jacque, minha querida irmã, agradeço toda ajuda e oração. Gratidão por se fazer tão presente em minha vida, por abrir a sua casa e torná-la extensão da minha quando eu mais precisava. Obrigada por me ajudar com as Marias, por trazer alegria e leveza em seus dias tornando a minha ausência o mais confortável possível para elas.

Ao meu pai Admar, avó Nair e todos os meus familiares agradeço por compreenderem a minha ausência em vários momentos e principalmente pelo estímulo constante. Carinhosamente, agradeço a minha tia-avó Dica e a minha avó Lourdes (in memoriam) minhas primeiras incentivadoras. Agradeço às minhas mães (Ilza e Cida) por acompanharem mais de perto essa trajetória. Gratidão especial pelas vidas dos meus amados afilhados e sobrinhos, que sempre foram motivos de inspiração para eu perseverar na caminhada. Não imagino a minha vida sem vocês.

Ao meu amigo Rondinele, irmão de uma vida, parceiro desde a graduação agradeço a amizade, apoio e ajuda. Mesmo não sendo da mesma área de estudo esteve tão presente em todos os momentos, partilhando alegrias, tristezas, desafios, vitórias e sonhos. Quan-

tas conversas, viagens, desabafos, exames e conquistas. A sua companhia tornou essa trajetória mais leve e feliz. Verdadeiramente, "...um amigo fiel é uma poderosa proteção: quem o achou, descobriu um tesouro..." (Eclo 6:14).

A todos os colegas da FACOM, particularmente ao Regis, Sara e Jean, que direta ou indiretamente estiveram presentes e colaboraram comigo nessa jornada acadêmica.

Não poderia deixar de agradecer a família GPR São José Operário e RCC Uberlândia. Gratidão por ouvirem minhas partilhas e sempre torcerem pelo meu sucesso. As orações de cada um foram fundamentais para fortalecer-me nos dias mais tensos.

A todos os meus amigos, de perto e de longe, com os quais eu dividi alegrias e incertezas, agradeço o carinho e a amizade. Um agradecimento especial às amigas mais próximas (Nayara, Laiz e Claudinha) pelo apoio, bons conselhos e palavras positivas. E aos amigos que sempre atenderam os meus pedidos de oração (Lili, Inez, Marcinha, Maurinho, Cris, Cláudia e Mônica) e se colocaram disponíveis para todo tipo de ajuda.

Agradeço antecipadamente aos membros da banca examinadora, que gentilmente aceitaram participar e colaborar com esta tese.

À FAPEMIG pela concessão da bolsa durante o desenvolvimento da pesquisa.

*“Se cheguei até aqui foi porque me apoiei nos ombros de gigantes.”*  
*(Bernard of Chartres)*

---

## Resumo

No presente trabalho é proposto uma metodologia computacional para o processamento e análise de imagens relativas a embriões da *Drosophila melanogaster*. Investigações básicas sobre a expressão gênica desse organismo possibilitam conhecimentos fundamentais sobre processos biológicos importantes, como o desenvolvimento e a diferenciação celular. A metodologia apresentada é constituída por seis módulos computacionais desenvolvidos a partir de técnicas de Processamento Digital de Imagens que podem ser aplicados sequencialmente ou individualmente, a saber: (1) Isolamento do embrião, (2) Padronização, (3) Segmentação dos núcleos, (4) Representação dos dados nucleares, (5) Representação dos dados do citoplasma e (6) Quantificação da expressão. Os módulos incluem soluções para padronização dos dados, segmentação de imagens, representação dos dados de expressão gênica e extração de dados quantitativos – sendo útil para análises biológicas que vão desde a identificação do embrião principal na imagem até a visualização dos padrões de expressão gênica. Em conjunto, é oferecido uma solução genérica para o tratamento da complexidade dos dados desse tipo de imagem. Para validação da proposta, foram usadas imagens da superfície dos embriões e das seções sagital e transversa. Essas imagens foram obtidas de bases de imagens públicas (BDGP e FlyEx) e também de bases próprias. Os resultados comprovam que a metodologia proposta é flexível e robusta, uma vez que é capaz de tratar uma ampla variedade de imagens desse domínio. Parte dos métodos propostos tiveram desempenho superiores aos encontrados na literatura. Também são apresentadas interpretações biológicas realizadas a partir dos dados obtidos com a aplicação da metodologia.

**Palavras-chave:** Processamento de imagens. Análise de imagens. Núcleos celulares. Expressão gênica. Microscopia confocal. *Drosophila melanogaster*.

---

## Abstract

In this thesis, it is proposed a computational methodology for processing and analysis of *Drosophila melanogaster* embryos images. Basic investigations of the gene expression in this organism provide fundamental knowledge about important biological processes, such as cell development and differentiation. The methodology consists of six computational modules developed from Digital Image Processing techniques that can be applied sequentially or individually, namely: (1) Embryo isolation, (2) Standardization, (3) Nuclear segmentation, (4) Representation of nuclear data, (5) Representation of cytoplasm data and (6) Expression quantification. The modules include solutions for data standardization, image segmentation, representation of gene expression data and extraction of quantitative gene expression data – being useful for biological analysis ranging from the identification of the main embryo in the image to the visualization of the gene expression patterns. Together, it is offered a generic solution for the treatment of the data complexity of this image type. The proposal was validated using embryo surface images and sagittal and transverse sections. These images were obtained from public image databases (BDGP and FlyEx) and also from our own databases. The results show that the proposed methodology is flexible and robust, as it handles a wide variety of images in this domain. Part of the proposed methods performed better than those found in the literature. In addition, it is presented the biological interpretations made from the data obtained with the application of the methodology.

**Keywords:** Image processing. Image analysis. Cell nuclei. Gene expression. Confocal microscopy. *Drosophila melanogaster*.

---

## Listas de ilustrações

Figura 1 – Ciclo de vida da <i>Drosophila melanogaster</i> . . . . .	24
Figura 2 – Número de clivagens sofridas por um embrião de <i>Drosophila</i> . . . . .	26
Figura 3 – Cascata de interações entre os genes envolvidos no processo de segmentação do corpo de embriões da <i>Drosophila melanogaster</i> . . . . .	28
Figura 4 – Padrão de expressão gênica dos genes pair-rule . . . . .	29
Figura 5 – Visões ortogonais de um embrião de <i>Drosophila</i> . . . . .	30
Figura 6 – Tipos de imagens de embriões de <i>Drosophila</i> . . . . .	31
Figura 7 – Geometria dos lados do embrião de <i>Drosophila</i> . . . . .	32
Figura 8 – Visão geral da abordagem proposta por (TRISNADI et al., 2013). . . . .	47
Figura 9 – Etapas na geração da máscara binária de (COPPEY et al., 2008) . . . . .	48
Figura 10 – Estrutura geral da metodologia computacional proposta . . . . .	50
Figura 11 – Configuração de marcador no módulo Isolamento do embrião principal	52
Figura 12 – Etapas para detecção do embrião de interesse no módulo Isolamento do embrião principal . . . . .	53
Figura 13 – Variações de direção e orientação dos embriões de <i>Drosophila</i> . . . . .	54
Figura 14 – Eixos e setores em um embrião de <i>Drosophila</i> . . . . .	55
Figura 15 – Esquema de padronização do embrião de <i>Drosophila</i> . . . . .	58
Figura 16 – Resultado qualitativo da aplicação do método de segmentação . . . . .	60
Figura 17 – Correspondência entre os pontos de mínimos locais . . . . .	61
Figura 18 – Definição do ponto central em imagens da seção transversa . . . . .	66
Figura 19 – Abertura das faixas nuclear e citoplasmática em imagem de embrião na seção transversa . . . . .	68
Figura 20 – Processo de quantificação da região nuclear - plano frontal . . . . .	70
Figura 21 – Processo de quantificação em uma imagem arbitrária da base FlyEx .	71
Figura 22 – Representação gráfica da faixa central. . . . .	71
Figura 23 – Comparação gráfica das faixas central e mediatriz. . . . .	72
Figura 24 – Processo de quantificação da região nuclear - seção sagital. . . . .	72
Figura 25 – Aplicação da transformada distância em uma máscara binária nuclear .	74

Figura 26 – Diferentes tamanhos da transformada distância na delimitação das regiões do citoplasma. . . . .	74
Figura 27 – Determinação das regiões dos citoplasmas apical e basal. . . . .	75
Figura 28 – Procedimento para definição da máscara nuclear rotulada. . . . .	75
Figura 29 – Imagens das bases SAGIBASE e SURFBASE . . . . .	78
Figura 30 – Exemplos de imagens contidas na base TRANSBASE . . . . .	80
Figura 31 – Exemplos de imagens tratadas pela base FlyEx . . . . .	81
Figura 32 – Exemplos de imagens tratadas pela base <i>Berkeley Drosophila Genome Project</i> (BDGP) . . . . .	81
Figura 33 – Pré-processamento em embriões da base BDGP . . . . .	82
Figura 34 – Resultado qualitativo do método de segmentação do embrião - base SAGIBASE . . . . .	83
Figura 35 – Resultado qualitativo do método de segmentação do embrião - base SURFBASE . . . . .	83
Figura 36 – Isolamento do embrião principal - Base SURFBASE . . . . .	84
Figura 37 – Desafios das imagens de embriões da <i>Drosophila</i> . . . . .	85
Figura 38 – Resultado qualitativo do método de segmentação do embrião - base BDGP . . . . .	85
Figura 39 – Resultado qualitativo do módulo Padronização - base SURFBASE . . . .	87
Figura 40 – Resultado qualitativo do módulo Padronização - base SAGIBASE . . . .	88
Figura 41 – Resultado qualitativo do módulo Padronização - base BDGP . . . . .	89
Figura 42 – Resultado qualitativo do módulo Padronização - base FlyEx . . . . .	90
Figura 43 – Efeitos do pré-processamento em imagens do plano frontal . . . . .	91
Figura 44 – Efeitos do pré-processamento em imagens da seção sagital . . . . .	92
Figura 45 – Variação de características nas imagens de embriões . . . . .	93
Figura 46 – Resultado qualitativo do método de segmentação nuclear sagital - base TRANSBASE . . . . .	94
Figura 47 – Resultado qualitativo do método de segmentação nuclear sagital - base SAGIBASE . . . . .	95
Figura 48 – Erros do método de segmentação sagital . . . . .	96
Figura 49 – Resultado qualitativo do método de segmentação nuclear frontal - base SURFBASE . . . . .	97
Figura 50 – Desafios das imagens de embriões da <i>Drosophila</i> . . . . .	98
Figura 51 – Representação dos dados nucleares - SAGIBASE . . . . .	99
Figura 52 – Representação dados nucleares - lado dorsal . . . . .	100
Figura 53 – Representação dados nucleares - lado ventral . . . . .	101
Figura 54 – Representação dados nucleares - TRANSBASE . . . . .	102
Figura 55 – Representação dos dados citoplasmáticos - Região Basal . . . . .	103
Figura 56 – Representação dos dados citoplasmáticos - Região Apical . . . . .	104

Figura 57 – Imagens do processo de quantificação gênica em embriões de <i>Drosophila</i>	105
Figura 58 – Sobreposição do contorno das regiões de interesse da seção transversa.	106
Figura 59 – Perfis de expressão gênica - Faixas Central e Mediatriz	107
Figura 60 – Comparação gráfica das faixas central e mediatriz.	108
Figura 61 – Perfis de expressão gênica - Região nuclear (meio embrião)	109
Figura 62 – Comparação gráfica das regiões dorsal e ventral.	110
Figura 63 – Resultado individual do perfil de expressão gênica em uma imagem de <i>Drosophila</i> na seção transversa.	111
Figura 64 – Perfil da intensidade média das regiões nuclear, apical e basal para todos os embriões selvagens.	112
Figura 65 – Perfil da intensidade média das regiões nuclear, apical e basal para todos os embriões mutantes.	113
Figura 66 – Comparação entre os perfis de intensidade média das regiões nuclear, apical e basal para os embriões mutantes e selvagens.	115

---

## **Lista de tabelas**

Tabela 1 – Lista de trabalhos correlatos. . . . .	36
Tabela 2 – Lista de bases de imagens correlatas . . . . .	37
Tabela 3 – Experimentos e Bases de imagens tratadas neste trabalho . . . . .	77
Tabela 5 – Número de imagens considerada para cada lâmina nas bases SAGI-BASE e SURFBASE. . . . .	78
Tabela 6 – Número de imagens considerada para cada genótipo. . . . .	79
Tabela 7 – Taxa de sucesso do módulo Padronização - SAGIBASE e SURFBASE	86
Tabela 8 – Taxa de sucesso do módulo Padronização - base FlyEx. . . . .	86
Tabela 9 – Taxa de sucesso do módulo Padronização - base BDGP. . . . .	86

---

## **Lista de siglas**

**2D** Bidimensional

**3D** Tridimensional

**AP** Anterior-Posterior

**BDGP** *Berkeley Drosophila Genome Project*

**CP** Ponto Central (*Central Point*)

**DAPI** (4',6-diamidino-2-phenylindole), um marcador de DNA

**DNA** Ácido desoxirribonucleico (*Desoxyribonucleic acid*)

**DV** Dorso-Ventral

**EP** Embrião principal

**ES** Elemento estruturante

**Eq.** Equação

**Eqs.** Equações

**Fig.** Figura

**Figs.** Figuras

**PCA** Análise de componentes principais (*Principal Component Analysis*)

**PDI** Processamento Digital de Imagens

**RNA** Ácido ribonucleico (*Ribonucleic acid*)

**Sec.** Seção

**TIFF** Formato de arquivo de imagem marcado (*Tagged Image File Format*)

---

# Sumário

1	INTRODUÇÃO . . . . .	17
1.1	Motivação . . . . .	18
1.2	Objetivos . . . . .	19
1.3	Justificativa . . . . .	19
1.4	Principais Contribuições . . . . .	20
1.5	Organização da tese . . . . .	21
2	CONCEITOS BIOLÓGICOS . . . . .	23
2.1	<i>Drosophila melanogaster</i> . . . . .	23
2.2	Desenvolvimento embrionário da <i>Drosophila</i> . . . . .	24
2.3	Expressão Gênica . . . . .	25
2.4	Imagens de Embriões de <i>Drosophila</i> . . . . .	29
2.5	Imagens de expressão gênica . . . . .	32
2.5.1	Distribuição espacial da expressão gênica . . . . .	32
2.5.2	Microscopia confocal . . . . .	33
3	TRABALHOS CORRELATOS . . . . .	35
3.1	Trabalhos correlatos . . . . .	35
3.2	Bases de imagens . . . . .	37
3.2.1	Fly-FISH . . . . .	37
3.2.2	FlyExpress 7 . . . . .	38
3.2.3	FlyEx . . . . .	38
3.2.4	Berkeley Drosophila Transcription Network Project . . . . .	39
3.2.5	Berkeley <i>Drosophila</i> Genome Project . . . . .	40
3.2.6	FlyBase . . . . .	40
3.3	Padronização das imagens . . . . .	41
3.4	Pré-processamento . . . . .	43
3.5	Segmentação . . . . .	44

<b>3.6</b>	<b>Representação e quantificação dos dados . . . . .</b>	<b>46</b>
<b>4</b>	<b>METODOLOGIA COMPUTACIONAL PROPOSTA . . . . .</b>	<b>49</b>
<b>4.1</b>	<b>Visão Geral . . . . .</b>	<b>49</b>
<b>4.2</b>	<b>Métodos . . . . .</b>	<b>50</b>
<b>4.3</b>	<b>Módulo 1 - Isolamento do embrião principal . . . . .</b>	<b>51</b>
<b>4.4</b>	<b>Módulo 2 - Padronização dos embriões . . . . .</b>	<b>52</b>
4.4.1	Direção e orientação . . . . .	54
4.4.2	Eixos e setores do embrião . . . . .	55
4.4.3	Rotação, espelhamento e corte . . . . .	56
<b>4.5</b>	<b>Módulo 3 - Segmentação dos núcleos . . . . .</b>	<b>57</b>
4.5.1	Seções Sagital e Transversa . . . . .	59
4.5.2	Imagens da Superfície . . . . .	62
<b>4.6</b>	<b>Módulo 4 - Representação dos dados nucleares . . . . .</b>	<b>63</b>
4.6.1	Visão linear dos núcleos celulares . . . . .	64
<b>4.7</b>	<b>Módulo 5 - Representação dos dados do citoplasma . . . . .</b>	<b>67</b>
<b>4.8</b>	<b>Módulo 6 - Quantificação dos dados da expressão . . . . .</b>	<b>69</b>
4.8.1	Quantificação da expressão nos núcleos . . . . .	69
4.8.2	Quantificação nas faixas Central e Mediatriz . . . . .	69
4.8.3	Quantificação das Seções Sagital e Transversa . . . . .	71
4.8.4	Quantificação do citoplasma . . . . .	73
<b>5</b>	<b>EXPERIMENTOS E RESULTADOS . . . . .</b>	<b>76</b>
<b>5.1</b>	<b>Bancos de imagens . . . . .</b>	<b>76</b>
5.1.1	SURFBASE e SAGIBASE . . . . .	78
5.1.2	TRANSBASE . . . . .	79
5.1.3	FlyEx . . . . .	79
5.1.4	BDGP . . . . .	80
<b>5.2</b>	<b>Resultados . . . . .</b>	<b>82</b>
5.2.1	Módulo 1 – Isolamento do embrião . . . . .	82
5.2.2	Módulo 2 – Padronização do embrião . . . . .	84
5.2.3	Módulo 3 – Segmentação dos núcleos: pré-processamento . . . . .	88
5.2.4	Módulo 3 – Segmentação dos núcleos . . . . .	90
5.2.5	Módulo 4 – Representação dos dados nucleares . . . . .	92
5.2.6	Módulo 5 – Representação dos dados do citoplasma . . . . .	94
5.2.7	Módulo 6 – Quantificação nuclear e citoplasmática . . . . .	96
<b>5.3</b>	<b>Resultados biológicos . . . . .</b>	<b>98</b>
5.3.1	Imagens da superfície dos embriões . . . . .	99
5.3.2	Seção Sagital . . . . .	100
5.3.3	Seção transversa . . . . .	103

6	CONCLUSÃO . . . . .	117
6.1	Limitações . . . . .	117
6.2	Trabalhos Futuros . . . . .	118
6.3	Contribuições bibliográficas . . . . .	119
	REFERÊNCIAS . . . . .	120

# CAPÍTULO 1

## Introdução

Desde as investigações de Thomas Hunt Morgan no início do século passado (1908), a mosca *Drosophila melanogaster* tem sido considerada um dos principais modelos de estudos usados em biomedicina para investigar uma ampla gama de fenômenos biológicos (TAORMINA et al., 2019). É vasto o campo de estudo de aplicabilidade dessa mosca, como por exemplo, em biologia do desenvolvimento, embriologia e pesquisa genômica. Temáticas como regulação gênica, diferenciação celular, morfogênese, crescimento, reprodução, evolução e a genética se inserem naturalmente nesse contexto (TOLWINSKI, 2017). Ao longo dos últimos anos observou-se um crescimento vertiginoso dos bancos de dados (públicos ou privados) e equipamentos de captura de imagens relativos à *Drosophila* (PRUTEANU-MALINICI; MAJOROS; OHLER, 2013).

A *Drosophila* é um organismo modelo, referência na biologia do desenvolvimento. Seus padrões de expressão gênica são amplamente estudados na forma de imagens digitais (WUNDERLICH et al., 2019; SURKOVA et al., 2019). Esses padrões têm o potencial de fornecer *insights* significativos sobre as funções e interações gênicas, podendo gerar percepções sobre as complexas redes regulatórias que governam o desenvolvimento embrionário (JENNINGS, 2011; SURKOVA et al., 2008; JI et al., 2008). Investigações básicas sobre a expressão gênica da *Drosophila* fornecem informações fundamentais sobre propriedades (biológicas, fisiológicas, neurológicas) e processos biológicos diversos (e.g., desenvolvimento, crescimento, diferenciação) (PANDEY; NICHOLS, 2011). Além disso, é importante mencionar que os conhecimentos obtidos com os padrões da *Drosophila* podem ser estendidas a outros animais, como os mamíferos. Isso porque, seus genes possuem um alto grau de identidade e conservação com os genes de vários organismos multicelulares (HOSKINS et al., 2015).

Todavia, isoladamente, as imagens dos padrões de expressão gênica não são suficientes para responder muitas questões. Por exemplo, quais os princípios que controlam o desenvolvimento embrionário. Ou ainda, como as células percebem suas posições dentro do tecido em desenvolvimento (KNOWLES, 2012; GILBERT, 2003). Nesse contexto, somado à observação visual dos padrões de expressão, torna-se necessário extrair e avaliar

as informações quantitativas contidas nessas imagens. Entretanto, essa tarefa não é considerada trivial, uma vez que bancos de imagens podem conter dados para centenas de genes (JAEGER; MANU; REINITZ, 2012). Dependendo do tamanho, a inspeção manual dessa tarefa torna-se ainda mais difícil, ou até mesmo inviável de ser realizada.

O cenário descrito pode ser caracterizado como um desafio computacional da área de Processamento Digital de Imagens (PDI). Nele existe uma demanda por soluções capazes de prover a realização automática de diversas tarefas relativas ao campo de pesquisa abordado. Até o momento é mais comum encontrar soluções que tratam problemas específicos a partir de um único tipo de imagem. Sendo assim, é necessário o desenvolvimento de soluções que agreguem diferentes propostas de maneira unificada.

## 1.1 Motivação

A compreensão dos mecanismos moleculares envolvidos no desenvolvimento embrionário da *Drosophila melanogaster* é de interesse em diversas pesquisas (ZHOU et al., 2019; CHONG; AMOURDA; SAUNDERS, 2018; WUNDERLICH et al., 2019). A caracterização da distribuição proteica e de RNAs dos genes de desenvolvimento permite que se estabeleça uma ligação entre os acontecimentos genéticos e as várias etapas do seu desenvolvimento. Embora os mecanismos moleculares de regulação gênica sejam bem conhecidos, os mecanismos que levam formação de padrões de expressão (resultado de complexas redes de interações gênicas) não são bem caracterizados. A produção de dados experimentais para a caracterização dessas redes de sinalização possibilita avanços no entendimento dos mecanismos de sinalização celular em tecidos embrionários e adultos. Além disso, a compreensão desses mecanismos tem o potencial de contribuir para o entendimento de diversas doenças disparadas por sinais celulares como a autoimunidade, diabetes e diversas formas de melanomas (GARCIA; FERNÁNDEZ, 2012; GILBERT, 2003).

O cenário descrito apresenta-se como um potencial motivador para a comunidade de Bioinformática. Do ponto de vista computacional, a produção de métodos ou ferramentas automáticas para esse âmbito é desafiadora, uma vez que as imagens analisadas não são triviais. Geralmente elas incluem um elevado número de informações e dados heterogêneos (e.g., diferentes ciclos celulares, posições e planos de captura, padrões de expressão de vários genes). Encontrar técnicas que, quando aplicadas isoladamente, sejam suficientes para tratá-las é improvável. Sendo assim, buscar avanços científicos para esse contexto incipiente, como a produção de ferramentas automáticas mais genéricas e que lidam com diferentes tipos de imagens, é outro ponto motivador dessa pesquisa.

## 1.2 Objetivos

O objetivo principal deste trabalho é o desenvolvimento de uma metodologia computacional para o processamento e análise de bioimagens relativas à morfologia e padrões de expressão gênica em embriões da *Drosophila melanogaster*. Visando atender o objetivo geral, são traçados os seguintes objetivos específicos:

- ❑ Desenvolver um conjunto de métodos computacionais que permitam a análise e caracterização morfológica do desenvolvimento de embriões da *Drosophila melanogaster* utilizando técnicas de processamento digital de imagens.
- ❑ Aplicar o conjunto de métodos a diferentes tipos de imagens de embriões de *Drosophila*, visando ilustrar o espectro de aplicação das propostas.
- ❑ Solucionar os problemas específicos ainda não resolvidos do cenário da *Drosophila*, como por exemplo, os relacionados aos desafios de segmentação, padronização e representação das imagens de embriões.
- ❑ Favorecer a compreensão dos mecanismos de formação e deslocamento dos padrões de expressão gênica da *Drosophila melanogaster* por meio do fornecimento de dados quantitativos do níveis de Ácido ribonucleico (*Ribonucleic acid*) (RNA) e proteínas nos embriões.
- ❑ Contribuir com investigações biológicas envolvidas na morfogênese da *Drosophila*, tais como estudos sobre a evolução espacial e temporal do processo de padronização, redes de regulação gênica e processos biológicos.

## 1.3 Justificativa

Com o avanço das técnicas de visão computacional para aplicações biológicas, a tarefa de processamento e análise automática de imagens de embriões da *Drosophila* recebeu grande atenção. Entretanto, a maioria das abordagens propostas ao longo dos últimos anos referem-se à soluções desenvolvidas para uma finalidade e campo específico. Por exemplo, a pesquisa de (LI; GONG, 2017) que apresenta um método para a extração automática do contorno de embriões em imagens da *Drosophila*. Apesar de recente, é limitada, uma vez que restringe a aplicação em imagens coletadas da superfície do embrião. Outro exemplo é o trabalho de (JANSSENS et al., 2005), o qual propõe um método de segmentação voltado especificamente para o plano frontal dos embriões de *Drosophila*. Seguindo esse raciocínio, citam-se as pesquisas de (AIZENBERG et al., 2002) e (MACE et al., 2010) que propõem, respectivamente, soluções peculiares para a classificação dos padrões de expressão gênica contidos em imagens de embriões obtidas do nível de resolução subcelular e celular. Aliado a esses fatores, cita-se a importância que os métodos e

ferramentas computacionais representam para o campo de análise de imagens, haja visto o rápido crescimento das bases de dados biológicas. A maioria aborda imagens complexas, no que diz respeito às várias configurações que os objetos de interesse podem apresentar (e.g., tamanhos distintos dos embriões, diferentes planos de captura, imagens com baixo contraste). Essa realidade evidencia a demanda por soluções capazes de resolver tarefas biológicas automaticamente, ou que, minimamente possibilitem acelerar os processos realizados manualmente por especialistas humanos.

Nesse contexto, a metodologia computacional proposta apresenta-se como solução adequada para processamento e análise de imagens relativas à embriões da *Drosophila*. Essa abordagem sugere um conjunto de rotinas computacionais desenvolvidas especificamente para suportar a diversidade de características e configurações das imagens tratadas. Ela representa uma estratégia inovadora, uma vez que trabalha com uma ampla variedade de domínios de imagens da *Drosophila*, tais como, diferentes imagens, níveis de resolução, planos, posições e eixos anatômicos.

## 1.4 Principais Contribuições

É possível citar pontos relevantes para ambas as áreas abordadas, computação e biologia.

### ❑ Métodos computacionais

Apresentação de um amplo conjunto de métodos computacionais para o processamento e análise de imagens embriões da *Drosophila*, desenvolvido para extrair dados quantitativos a partir dos padrões de expressão gênica e capaz de trabalhar com a ampla heterogeneidade e complexidade dos tipos de dados inerentes a esse cenário.

### ❑ Padronização de imagens

Proposição de um algoritmo não-supervisionado para a padronização de imagens de embriões da *Drosophila*. Essa solução considera propriedades relevantes dos dados, até então não tratadas de forma automática pela literatura, como a direção e orientação. Os resultados obtidos nessa análise estão publicados no artigo *Automated standardization of images of Drosophila embryos* publicado na revista *Journal of Visual Communication and Image Representation* (SOUSA et al., 2020).

### ❑ Seção transversa

Proposição de uma nova abordagem computacional para a quantificação e análise de imagens de embriões de *Drosophila* no plano transverso. O conjunto de passos e técnicas abordadas nesse trabalho são descritos no artigo *Computational Image Analysis from the Transverse Plane of Drosophila embryos* submetido à revista IEEE/ACM Transactions on Computational Biology and Bioinformatics.

□ Gradiente de Expressão Gênica

Apresentação de uma análise quantitativa dos perfis de expressão gênica da proteína Dorsal relativa às regiões dos núcleos, citoplasma apical e basal. Os resultados exibiram uma análise comparativa entre imagens de embriões selvagens e mutantes, para cada genótipo e região de interesse e descrevem resultados biológicos ainda não observados com o mesmo nível de detalhamento deste trabalho.

□ Bases de Imagens

Apresentação de novas base de imagens de embriões da *Drosophila* em diferentes configurações.

## 1.5 Organização da tese

No primeiro capítulo é apresentada a contextualização do tema da tese, a motivação e a justificativa para o trabalho. Também são apresentados os objetivos geral e específicos.

No segundo capítulo são descritos os principais conceitos biológicos relativos aos tópicos discutidos nessa tese, tais como a *Drosophila melanogaster*, a expressão gênica e a captura de dados por meio de imagens. É destacada a importância da *Drosophila* como organismo modelo, utilizado em pesquisas biomédicas para o estudo de uma ampla gama de investigações biológicas.

No terceiro capítulo são apresentados alguns trabalhos correlatos propostos ao longo dos últimos anos. O objetivo foi investigar na literatura a aplicação de métodos e técnicas computacionais para obter dados quantitativos de expressão gênica em imagens de embriões da *Drosophila*. O levantamento reforçou a importância do trabalho proposto como solução flexível e modular para o estudo da dinâmica dos padrões de expressão gênica em embriões de *Drosophila*.

No quarto capítulo é descrita a metodologia computacional proposta, composta por seis módulos, para processar e analisar as imagens (confocais ou não-confocais) de embriões da *Drosophila melanogaster*. Os módulos propõem soluções automáticas que contemplam tarefas que vão desde a identificação do tipo de embrião até a extração de características e análise biológica dos dados. Na descrição de cada módulo são detalhados os materiais e técnicas de processamento digital de imagens que foram utilizados. A contribuição principal consiste na proposição do conjunto completo, o qual permite abordar diferentes tipos de imagens de embriões da *Drosophila*, tratando de modo particular as características intrínsecas e a complexidade de cada tipo.

No quinto capítulo são descritos os experimentos realizados e os resultados alcançados. Do ponto de vista computacional, destaca-se a aplicação da metodologia em embriões de diferentes planos, posições e eixos anatômicos. Com relação ao campo biológico, são analisados padrões de formação dos gradientes *Hunchback* e Dorsal, duas proteínas relevantes no processo de desenvolvimento, nas regiões nucleares e citoplasmáticas.

Por fim, no Capítulo 6 são apresentadas as conclusões e o resumo das principais contribuições e limitações deste trabalho. Além disso, são descritos alguns temas relevantes para serem tratados em trabalhos futuros.

# CAPÍTULO 2

## Conceitos biológicos

Neste capítulo são descritos os conceitos biológicos necessários para os principais tópicos discutidos nesta tese, tais como informações sobre a *Drosophila*, de expressão gênica e captura de dados de expressão por meio de imagens.

### 2.1 *Drosophila melanogaster*

A *Drosophila melanogaster* é uma pequena mosca naturalmente encontrada ao redor de frutas em estado de putrefação, e por este hábito, é também conhecida como a “mosca da fruta” ou “mosca do vinagre”. Trata-se de um organismo multicelular amplamente utilizado na literatura para examinar fenômenos biológicos complexos. Os primeiros movimentos do uso da *Drosophila* datam de 1901, quando William Ernest Castle (um professor na Universidade de Harvard) passou a disseminar a ideia da utilização desse inseto para zoólogos experimentais. Mais tarde (1910), Dr. Thomas Hunt Morgan usou *Drosophila* para refinar a teoria que estabelece os mecanismos da herança mendeliana, proposta inicialmente por Gregor Mendel. Os benefícios em usar a *Drosophila* como organismo modelo são diversos, como facilidade de manutenção, ciclo de vida relativamente curto (10–12 dias), manipulações ambientais e genéticas que alteram o tempo de vida, sequência do genoma completo, estrutura sólida para a descrição anatômica dos estágios embrionários e métodos robustos para avaliação da expressão gênica (HELPAND; ROGINA, 2003; TOMANCAK et al., 2007; JOHNSTON; NUSSLEIN-VOLHARD, 1992).

Ao longo do último século a *Drosophila melanogaster* tem desempenhado um papel significativo em estudos sobre a relação entre o Ácido desoxirribonucleico (*Desoxyribonucleic acid*) (DNA), RNA e a síntese de proteínas, nos processos de regulação gênica e na formação dos padrões de expressão gênica (GILBERT, 2003; GARCIA; FERNÁNDEZ, 2012). Seu genoma possui um alto grau de conservação com o dos mamíferos, o que permite que ambos compartilhem muitas propriedades fundamentais. Estima-se que cerca de 75% dos genes associados a doenças humanas possuam um homólogo nos genes da *Drosophila* (MIRZOYAN et al., 2019; YUAN et al., 2013). Essa proximidade genética

significa que experimentos com a *Drosophila* podem ser traduzidos de maneira efetiva para humanos. Assim, a compreensão do seu desenvolvimento celular pode estender experiências a organismos mais complexos (MILLER et al., 2013). Esse é o caso, por exemplo, do estudo de distúrbios como, Parkinson, Mal de Alzheimer e Huntington (MIRZOYAN et al., 2019; BAGATINI et al., 2011; STEPHENSON; METCALFE, 2013; MICHNO, 2009; CHEN; CROWTHER, 2012). Sua utilização é também observada na descoberta de compostos farmacêuticos (HIRTH, 2010), na investigação do desenvolvimento de óvulos (ARIAS, 2008), e no estudo da tumorigênese, com foco na compreensão dos processos fundamentais do câncer em seres humanos (MIRZOYAN et al., 2019). É ainda importante mencionar a utilização de *Drosophila* em estudos relativos à investigação do vício do alcoolismo, distúrbios do sono e comportamentos de aprendizagem (BAGATINI et al., 2011). De modo geral, cada fase do desenvolvimento (sendo embrião, larva, pupa e adulto) apresenta vantagens para estudos específicos (GILBERT, 2003).

## 2.2 Desenvolvimento embrionário da *Drosophila*

O desenvolvimento da *Drosophila*, do ovo fertilizado até o organismo adulto, consiste basicamente de uma sequência de eventos modulados controlada por sinais ambientais e mecanismos intracelulares. Ao longo do ciclo ocorrem diferentes estágios: ovo, blastoderme sincicial, embrião, larvais, pupa e adulto (Fig. 1).

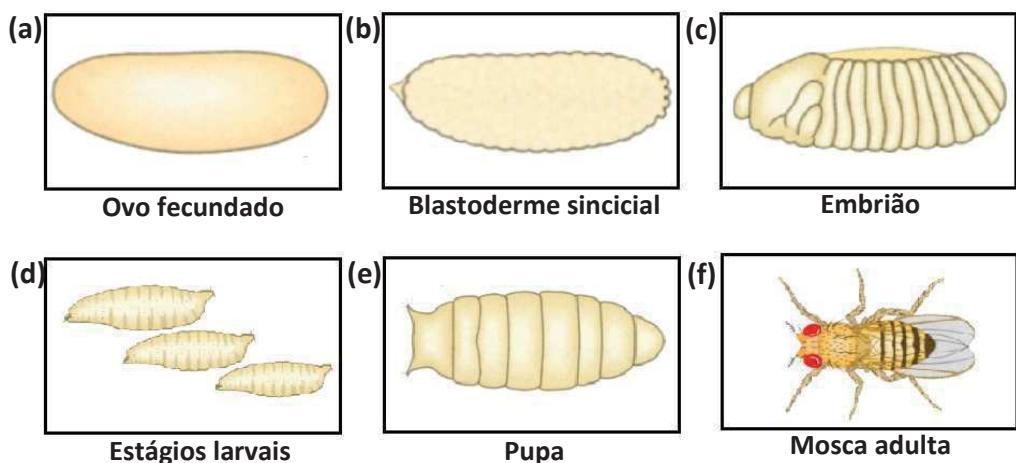


Figura 1 – Ciclo de vida da *Drosophila melanogaster*. Após a fertilização (a) a divisão nuclear começa. Nove ciclos de divisão nuclear promovem a blastoderme sincicial (b). E depois de mais algumas divisões, o estágio de blastoderme celular é concluído. Cerca de 10 horas após a fertilização, é formado um embrião com segmentos morfológicamente distintos (c). Nas fases finais, o embrião se desenvolve em uma larva que passa por três estágios (d). O estado subsequente, de pupa (e), dura cerca de cinco dias. E após esse período de metamorfose, se origina um indivíduo adulto. Figura adaptada de (MORGAN, 2007)

Nesse contexto, a embriogênese (fase de interesse desse trabalho) refere-se aos processos de formação e desenvolvimento dos embriões de *Drosophila*. Ela compreende os estágios da fecundação até a eclosão da larva (GILBERT, 2003), tal como destacado na Fig. 1(a-c). De modo simplificado, um único espermatozóide penetra no óvulo maduro e os núcleos haploides do espermatozoide e do óvulo fundem-se para formar um núcleo zigótico diploide. Então, o núcleo zigótico sofre uma série de divisões sem que ocorra a divisão do citoplasma e formação de células. Na décima divisão os núcleos migram progressivamente para a periferia do embrião (córTEX) e ocorrem mais quatro divisões (ciclos mitóticos) parando na interfase do ciclo 14. Entre os ciclos 10 – 13 os núcleos resultantes das sucessivas divisões partilham um citoplasma comum, designando-se o embrião de blastoderme sincial. Por fim, durante a interfase do 14º ciclo mitótico todos os núcleos são envolvidos por membranas plasmáticas. Trata-se do processo de invaginação, em que a membrana celular do oócito dobra-se sobre os núcleos e intercala-se entre estes, dando origem a uma camada única de células, a blastoderme celular.

Completados os 14 ciclos mitóticos (aproximadamente 4,5 horas após a fecundação) o embrião é uma única célula contendo cerca de  $2^{13}$  núcleos. Esses núcleos, morfologicamente idênticos, são imersos em diferentes concentrações de proteínas cujos RNAs são depositados maternalmente, antes da fecundação. Sendo assim, os núcleos e suas ilhas citoplasmáticas são diferentes ao nível molecular, dado a expressão gênica dos muitos genes e fatores de transcrição do organismo (GARCIA; FERNÁNDEZ, 2012). Na Fig. 2 é ilustrado os ciclos mitóticos (clivagens) sofridas pelo embrião de *Drosophila* durante a embriogênese. Em particular, esta tese aborda imagens de embriões com sinais de expressão gênica, normalmente utilizadas em estudos relativos à formação dos padrões, a determinação do destino celular e à organogênese (PANDEY; NICHOLS, 2011).

## 2.3 Expressão Gênica

Nas primeiras horas do desenvolvimento embrionário (entre fertilização e gastrulação) a expressão gênica da *Drosophila* é altamente dinâmica. Duas categorias de transcritos marcam esse período do ciclo, sendo (i) aqueles depositadas no ovo durante a oogênese (produzida por genes maternos) e (ii) aqueles que são expressos somente após a fertilização (produzida por genes zigóticos) (ARBEITMAN et al., 2002). Até o 7º ciclo o desenvolvimento embrionário é controlado pelos genes maternos, isto é, o embrião não transcreve o seu genoma. Os primeiros genes zigóticos são transcritos no 8º ciclo, e na interfase do 14º ciclo a transcrição zigótica aumenta em grande escala com o processo de celularização (GARCIA; FERNÁNDEZ, 2012; WATSON; BAKER; BELL, 2015).

A expressão gênica é o processo pelo qual a informação hereditária contida em um gene é processada em um produto gênico. Ela consiste na transcrição do DNA para o mRNA (RNA mensageiro) e na tradução do mRNA em uma proteína (WATSON; BAKER; BELL,

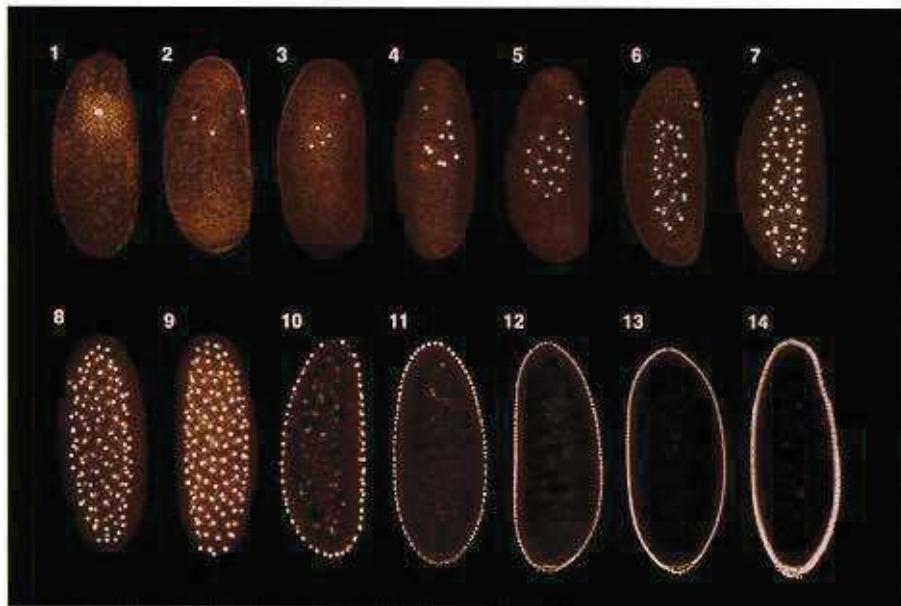


Figura 2 – Número de clivagens sofridas por um embrião de *Drosophila*. As primeiras divisões ocorrem centralmente. Os números no topo de cada embrião identificam o ciclo mitótico. No décimo ciclo os núcleos e suas ilhas citoplasmáticas migram para a periferia do embrião, criando a blastoderme sincicial. Após o 13º ciclo, as membranas celulares começam a se formar, invaginando da superfície externa do embrião em direção ao centro, formando o blastoderma celular. Figura obtida de (GILBERT, 2003).

2015). Esse é o processo no qual a informação contida nos genes (sequência de DNA) gera produtos gênicos, que são as moléculas de RNA (na etapa de transcrição gênica) e as proteínas (na etapa de tradução gênica). Essa segunda etapa consiste na síntese de proteínas, que são os produtos finais e os objetivos reais da expressão gênica. Por simplificação, muitos outros detalhes referentes às etapas de transcrição e tradução foram omitidos nesta tese. Uma leitura avançada pode ser encontrada na literatura (GILBERT, 2003).

A princípio todas as células de um organismo compartilham o mesmo DNA, e por isso, possuem a mesma informação genética. No entanto, graças às mudanças na expressão gênica que ocorrem ao longo do desenvolvimento, são geradas células diferentes com relação à função e estrutura. Por exemplo, os tecidos da pele e do coração, possuem o mesmo genoma, porém diferenciam-se quanto à morfologia e função. Esse processo em que as células de um mesmo organismo se especializam em determinada função é definido como diferenciação celular. Resumidamente, ela compreende os mecanismos responsáveis por determinar a especificidade, o destino e a função das células dentro de um organismo. A diferenciação celular explica a grande diversidade de estruturas, órgãos e sistemas existentes em um organismo, os quais interagem entre si e desempenham diversas funções necessárias à sua sobrevivência. A realidade de diferentes grupos de células agindo de maneira cooperativa determina a existência dos organismos multicelulares. A diferença

fundamental entre esses organismos está na inibição ou na ativação de determinados grupos de genes. Portanto, a expressão diferenciada acontece porque existe um controle sobre a expressão gênica (WATSON; BAKER; BELL, 2015; GILBERT, 2003).

O mecanismo que define a interação e regulação entre os genes durante a expressão gênica denomina-se rede de regulação gênica. Ela define como os genes influenciam uns aos outros, de maneira ativadora ou inibidora. Nem todos os genes presentes no núcleo se expressam ao mesmo tempo. Alguns genes são ativados em determinados tecidos ou em períodos específicos do desenvolvimento. Alguns genes ativados podem atuar na ativação ou inativação de outros genes, isto é, tratam-se dos genes com função reguladora. Ao conjunto de todos os genes expressos em um dado momento denomina-se padrão de expressão gênica. De modo geral, o padrão de expressão gênica é determinado pelo código genético e também por informações externas à célula, tal como sinais químicos de outras células ou níveis de nutrientes (GARCIA; FERNÁNDEZ, 2012).

Os complexos detalhes de diferenciação celular, desenvolvidos a partir do zigoto até a embriogênese, induzem o estudo dos padrões de expressão gênica. É sabido que os mecanismos envolvidos nesse processo não ocorrem somente durante o desenvolvimento embrionário, pois mudanças na expressão gênica continuam ativas em cada indivíduo durante toda a sua vida. No entanto, a elevada atividade mitótica e constantes mudanças no destino das células fazem deste período o preferido para a realização de investigações diversas (GILBERT, 2003).

O interesse desta tese foca-se na análise computacional de imagens relativas aos padrões de expressão gênica de embriões da *Drosophila melanogaster*. Estudos da expressão gênica em grande escala nesse organismo contribui para a compreensão de importantes processos e fenômenos biológicos, tais como desenvolvimento e diferenciação celular (ARBEITMAN et al., 2002). Além disso, esta abordagem constitui um passo essencial para esclarecer a complexidade do processo de regulação gênica. Um exemplo relevante desse contexto é o estabelecimento dos padrões de expressão em um caso bastante importante, a formação do plano corporal dos embriões de *Drosophila*.

Simplificadamente, diferentes conjuntos gênicos atuam em cascata para o estabelecer o plano corporal dos embriões de *Drosophila melanogaster*. Os atores iniciais dessa cascata são os genes de efeito materno, com localização nas células polares do embrião. Seus produtos gênicos determinam a formação do eixo antero-posterior e criam informações que induzem a ativação do próximo grupo de genes, os genes *gap*. Os produtos dos genes *gap* interagem entre si para definir regiões grandes do corpo do embrião. A sua expressão se dá em faixas amplas. O gradiente de expressão dos genes *gap* leva à ativação subsequente dos genes *pair-rule*. Esse grupo de genes divide o embrião em sete faixas verticais perpendiculares ao eixo anterior-posterior (Anterior-Posterior (AP)), com largura de aproximadamente dois segmentos. Eles são responsáveis pela formação dos futuros segmentos (parasegmentos). A ação do grupo de genes *pair-rule* ativa o grupo de genes

polaridade segmentar, responsáveis por assegurar a divisão dos segmentos. A ativação desses genes divide o embrião em quatorze faixas com a largura dos segmentos corporais. E ainda, divide cada segmento em metade anterior e posterior. Esse processo é ilustrado na Fig. 3, em que é possível observar a cascata de interações entre os genes envolvidos no processo de segmentação do corpo de embriões da *Drosophila melanogaster*.

A caráter de exemplificação é contextualizado a ação do grupo de genes de efeito materno, fundamentais para a organização básica e a polaridade do eixo AP dos embriões de *Drosophila*. Aqui citam-se o *bicoid* e *hunchback* (como necessários para a formação da cabeça e do tórax do embrião) e *nanos* e *caudal* (como essenciais para a formação dos segmentos abdominais). O gene *bicoid*, por exemplo, atua especificando o desenvolvimento da região anterior do ovócito. Depois da fecundação, o mRNA de *bicoid* é traduzido e a proteína Bicoid forma um gradiente de concentração que se estende da região anterior (onde é mais concentrada) até a região posterior (onde sua concentração é menor). O gene *caudal*, por sua vez, atua padronizando a região posterior do embrião. Antes da fecundação o mRNA de *caudal* é sintetizado homogeneamente por todo o ovócito. No entanto, a alta concentração de Bicoid na porção anterior do embrião inibe a tradução do mRNA do gene *caudal* nesta região. Assim, a máxima concentração da proteína Caudal encontra-se na região posterior do embrião. O mRNA do gene *nanos* possui maior concentração na região posterior embrião. O seu gradiente de expressão estende-se da região posterior para a anterior, mas somente é traduzido na porção posterior do embrião. Em geral, o *hunchback* atua na repressão de sua transcrição (GARCIA; FERNÁNDEZ, 2012).

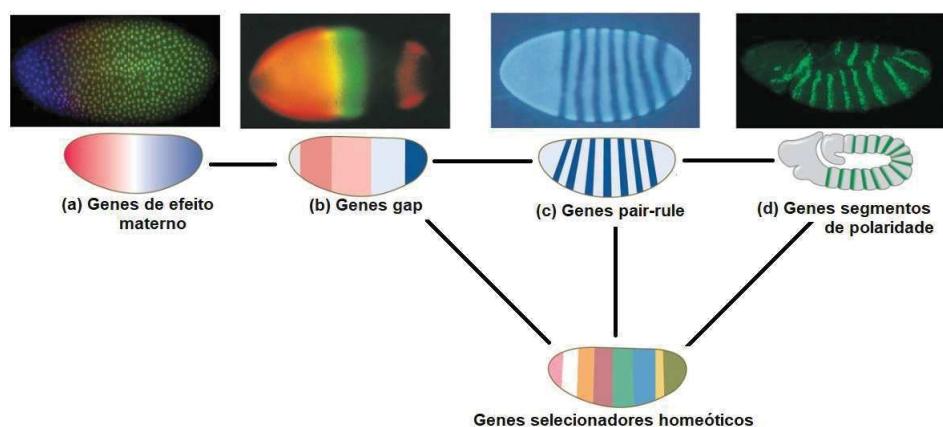


Figura 3 – Cascata de interações entre os genes envolvidos no processo de segmentação do corpo de embriões da *Drosophila melanogaster*. Figura adaptada de (HILLIS et al., 2012)

Uma explicação detalhada de todo o contexto biológico deste processo foge ao escopo desta tese. Porém, uma leitura avançada pode ser encontrada na literatura (GILBERT, 2003; GARCIA; FERNÁNDEZ, 2012). Na Fig. 4 é exibido o padrão de expressão do *Even-skipped* (vermelho), *Hunchback* (azul) e *Kruppel* (verde) em um embrião no 14º

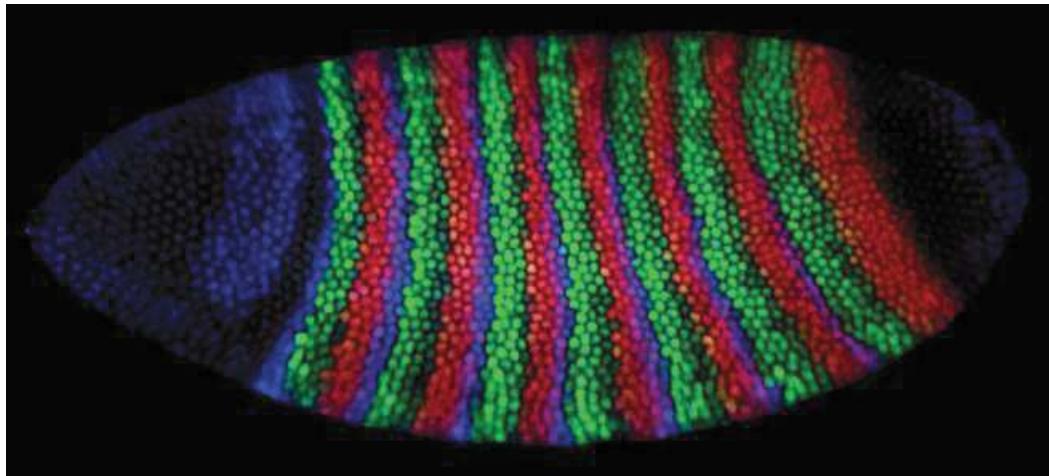


Figura 4 – Padrão de expressão gênica dos genes *pair-rule*. *Even-skipped* (vermelho), *Hunchback* (azul) e *Kruppel* (verde) em um embrião de *Drosophila* no ciclo de clivagem 14A e classe temporal 8. *Even-skipped* é caracterizado por um padrão de sete faixas transversais ao longo do eixo antero-posterior da blastoderme do embrião. Imagem obtida da base de dados online FlyEx (POUSTELNIKOVA et al., 2004).

ciclo de desenvolvimento e classe temporal 8. Os diferentes níveis de expressão de seus genes são decorrência de estarem submersos em diferentes concentrações das proteínas reguladoras da expressão gênica.

## 2.4 Imagens de Embriões de *Drosophila*

Os equipamentos de captura de imagens existentes (e.g., microscopia confocal) podem coletar amostras de embriões da *Drosophila* em diferentes posições: lateral, dorsal e ventral. Além disso, as imagens dos embriões podem ser coletadas em diferentes planos: superficial, sagital ou transversal.

Em grandes bases de dados, para interpretar os padrões de expressão e identificar a função dos respectivos genes, torna-se necessário separar as centenas de embriões em seus respectivos estágios de desenvolvimento (KAZMAR et al., 2013). Nesse sentido, compreender a geometria e anatomia dos embriões é fundamental. A análise inclui algumas propriedades principais, como (a) eixos corporais, (b) visões, (c) planos e (d) seções dos embriões.

De modo geral, o embrião possui um formato elíptico que se estende ao longo do eixo horizontal (ALLENA; AUBRY, 2011; ALBERTS et al., 2002). Nesse contexto, os eixos embrionários definem a estrutura corporal do futuro animal. No caso de embriões da *Drosophila melanogaster* citam-se dois eixos, o anterior-posterior (AP) e o dorso-ventral (DV), como ilustrado na Fig. 7. O eixo AP é definido por uma linha imaginária que se estende da extremidade anterior (cabeça do embrião) até a porção posterior (cauda do

embrião). Por outro lado, o eixo Dorso-Ventral (DV) define a linha imaginária perpendicular ao eixo AP, a qual se estende verticalmente do lado dorsal (parte superior do embrião) ao ventral (parte inferior) (GARCIA; FERNÁNDEZ, 2012).

No que diz respeito à propriedade (b), é possível dizer que os embriões de *Drosophila* podem ser coletados sob dois tipos de visualização principais, a lateral e a não-lateral (que inclui os lados dorsal e ventral). O tipo de visualização diz respeito à visão utilizada para capturar o embrião (KARAISKOS et al., 2017). A ampla maioria dos trabalhos disponíveis na literatura abordam a visualização lateral dos embriões (LI; GONG, 2017). Por meio da Fig. 5 é possível conhecer as seis projeções dos embriões de *Drosophila*.

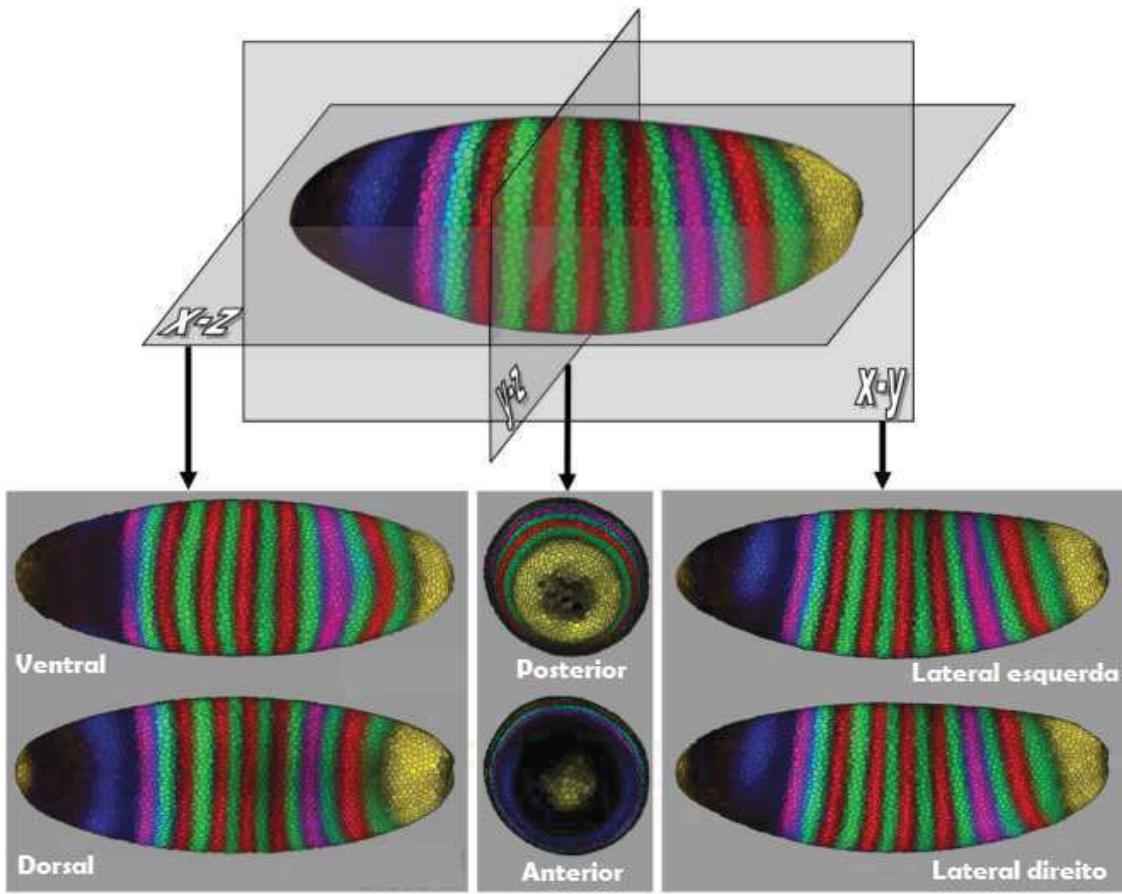


Figura 5 – Visões ortogonais de um embrião de *Drosophila* ao longo de seus eixos corporais. As projeções ilustradas são da visão ventral, dorsal, anterior, posterior, esquerda e direita. Figura adaptada de (WEBER et al., 2009).

Outra propriedade importante relativa à geometria refere-se ao plano de captura dos embriões de *Drosophila* (propriedade (c)). Neste trabalho, são capturadas imagens referentes à cortes no embrião, sendo as seções (i) sagital (Fig. 6(a)) e transversa (Fig. 6(b)) e à (ii) superfície do embrião (Fig. 6(c)). As imagens obtidas de (i) são geradas por técnicas de seccionamento virtual, e isto as permite detectar planos focais internos ao embrião, gerando a impressão de um seccionamento físico das amostras. Em contrapartida, as imagens obtidas de (ii) capturam os núcleos e sinais de expressão a partir de uma visão

ortogonal da superfície do embrião, por meio do achatamento físico do mesmo. Com isso, esses planos possibilitam a visualização de células e tecidos, localização de genes e proteínas, bem como, a visualização da dinâmica de expressão gênica em embriões fixados ou *in vivo* (GRIMM; COPPEY; WIESCHAUS, 2010). Em suma, as imagens das seções sagital e transversa apresentam vantagens importantes, às quais têm atraído o interesse da comunidade biológica. Por exemplo, (1) a captura de sinais de expressão de RNAs e proteínas em diferentes planos focais, (2) a preservação do embrião, uma vez que esse tipo de imagem é obtido por equipamentos e técnicas de captura de imagem capazes de capturar o plano interno das amostras sem achatar fisicamente os embriões, e (3) o estudo do eixo dorso-ventral, possível somente a partir desse plano da imagem (SOUSA et al., 2013).

Além disso, cita-se uma observação adicional sobre as seções sagital e transversa. No caso de animais com simetria bilateral, a seção sagital (longitudinal) é paralela ao eixo AP. Por outro lado, a seção transversa é perpendicular a esse eixo. Pelo fato dos embriões de *Drosophila* apresentarem forma elíptica, a seção transversa gera um formato aproximadamente circular, enquanto que a sagital mantém o formato elíptico. Na Fig. 6 é ilustrado os dois tipos descritos.

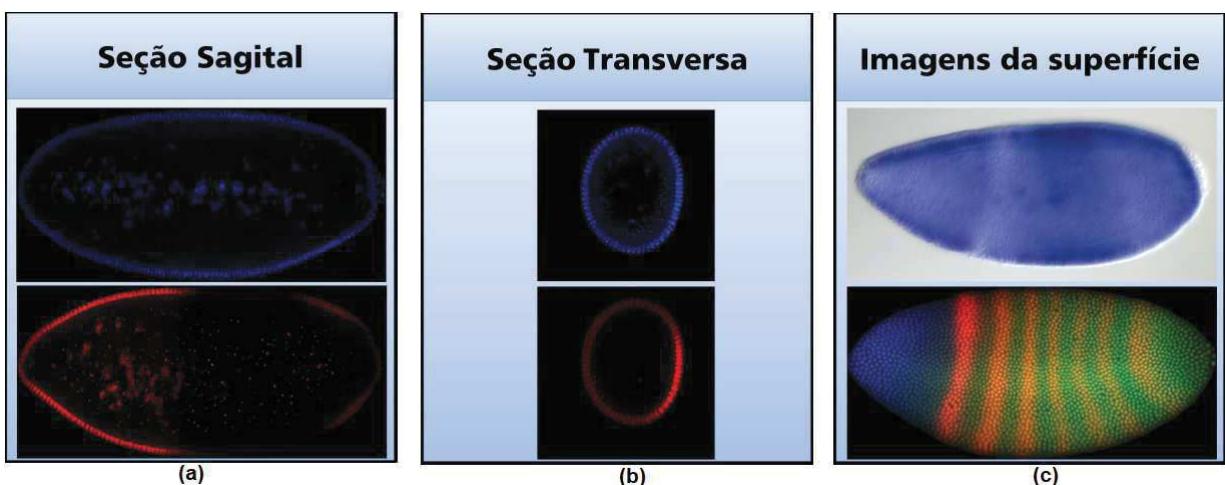


Figura 6 – Tipos de imagens de embriões de *Drosophila*. (a) e (b) cortes do embrião nas seções, respectivamente, sagital e transversa. Em ambas as figuras, no canal azul é ilustrado a marcação DAPI e no canal vermelho é expresso o sinal da proteína. (c) Imagem de embrião coletada da superfície. Na segunda figura é destacado a resolução dos núcleos.

Os embriões de *Drosophila*, apesar de apresentarem um formato próximo a uma elipsóide, não possuem os lados (dorsal e ventral) simétricos (ALLENA; AUBRY, 2011). Ele diferenciam-se com relação ao tamanho e curvatura, sendo que, o lado dorsal é menor e menos curvo (HE et al., 2010). Na Fig. 7 é destacado esse caso. Nela é possível perceber que o tamanho do contorno medido ao longo do perímetro do lado dorsal (cor azul) é

menor em relação ao tamanho do contorno medido ao longo do perímetro do lado ventral (cor vermelho).

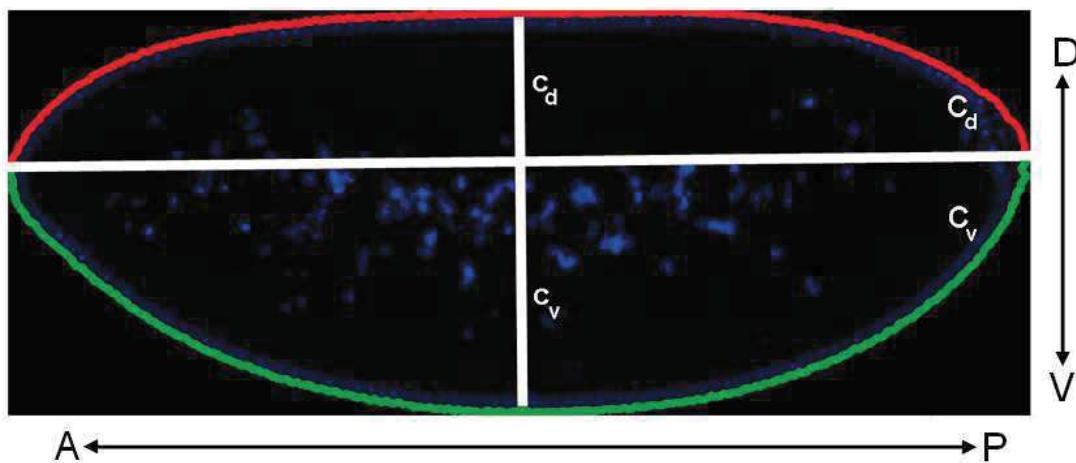


Figura 7 – Geometria dos lados do embrião de *Drosophila* na seção sagital (longitudinal). Em vermelho é delimitado o lado dorsal e em verde lado ventral. É nítida a diferença dos respectivos perímetros.

## 2.5 Imagens de expressão gênica

A localização e quantificação da expressão gênica em embriões da *Drosophila melanogaster* é definida a partir da distribuição espacial e temporal de seus RNAs e proteínas. Esse processo é realizado por meio de três métodos principais da biologia molecular: (i) Hibridização *in situ* (ISH), (ii) hibridização fluorescente *in situ* (FISH), (iii) imunolocalização e (iv) imagiamento *in vivo*. Esses métodos, quando associados à técnicas de captura de imagens (e. g., microscopia confocal), geram imagens 2D ou 3D da expressão gênica e também da anatomia (TRAVENÇOLO, 2007). Na Seção 2.5.1 são descritas, resumidamente, as técnicas (ii) e (iii), e na Seção 2.5.2 é apresentada a técnica de captura Microscopia confocal, amplamente usada no estudo da *Drosophila*.

### 2.5.1 Distribuição espacial da expressão gênica

A FISH é uma técnica utilizada para determinar a localização citológica de moléculas de RNA ou DNA a partir de sondas marcadas por fluorescência (NIZAMI; LIU; GALL, 2015). As sondas podem ser definidas como sequências de nucleotídeos complementares desenvolvidas a partir de segmentos conhecidos do DNA ou RNA que se deseja identificar. Sendo assim, a FISH baseia-se na propriedade da complementaridade, inerente às moléculas de DNA e RNA. De modo bem resumido, ela consiste em determinar quais moléculas de mRNA de um determinado gene estão presentes em uma célula ou tecido em um dado momento. É importante mencionar que existem três tipos de hibridização,

do tipo RNA/RNA, RNA/DNA e DNA/DNA (GILBERT, 2003). O procedimento completo aborda algumas etapas principais, que inclui desde a (i) fixação da amostra; (ii) preparação do material biológico, que pode incluir ou não passos de permeabilização; (iii) hibridização com sondas específicas usadas na detecção das sequências (determinados alvos em solução de hibridização); (iv) lavagem da amostra que removem a hibridização inespecífica, até a (v) montagem, visualização e documentação.

A FISH é considerada uma abordagem para examinar sistematicamente os padrões de expressão gênicos durante o desenvolvimento, e um ensaio central no estabelecimento de relações regulatórias entre os fatores de transcrição e os elementos reguladores (HAMMONDS et al., 2013). No cenário da *Drosophila melanogaster*, ela tem sido utilizada para determinar perfis de expressão, com potencial de revelar aspectos espaciais e temporais dos padrões gênicos durante o desenvolvimento embrionário. No contexto da presente tese, trata-se da técnica abordada pela base de imagens BDGP (TOMANCAK et al., 2002).

Outra técnica para a detecção de proteínas expressas nas células é a imunolocalização. De modo geral, essa técnica usa anticorpos específicos para localizar macromoléculas (proteínas, polissacarídeos) dentro do material biológico (frações subcelulares, células, tecidos, biofilmes) (HOPPERT; WREDE, 2011). A imunolocalização é baseada nos princípios das reações antígeno-anticorpo. Dessa forma, para determinar a distribuição espacial de uma proteína em um conjunto de células, utiliza-se um anticorpo associado a um corante (fluorocromo) específico para a proteína desejada (TRAVENÇOLO, 2007). As células devem ser fixadas e permeabilizadas para facilitar o acesso dos anticorpos ao citoplasma e ao núcleo (PAULA; TECHIO, 2014). Utilizando mais de um corante é possível visualizar diferentes proteínas simultaneamente. Esse processo é ilustrado na Fig. 4. Em resumo, a utilização da técnica de imunolocalização combinada com estudos de biologia molecular é uma ferramenta indispensável e tem contribuído significativamente no cenário da *Drosophila* para investigar a distribuição espacial ou temporal de uma determinada proteína no núcleo ou no citoplasma.

### 2.5.2 Microscopia confocal

Avanços em equipamento de captura de imagens microscópicas criaram novas oportunidades para estudar interações gênicas em *Drosophila*, por exemplo, permitindo a descoberta de padrões genéticos em diferentes estágios do desenvolvimento embrionário. Nesse contexto, a microscopia confocal apresenta-se como uma ferramenta poderosa para a geração de imagens da expressão gênica de embriões em desenvolvimento.

A microscopia confocal baseia-se na iluminação por varredura a laser, ponto-a-ponto, de um espécime biológico normalmente tratado ou marcado com compostos fluorescentes. De modo simplificado, os microscópios permitem a aquisição de imagens digitalizadas de diferentes planos focais de uma amostra biológica, ao passo que elimina a informação

fora do foco. Isto é alcançado por meio de uma abertura circular (*pinhole*) localizada em frente ao plano da imagem, a qual atua como um filtro espacial e permite somente a passagem de luz proveniente do plano focal (BOROVEC; KYBIC; NAVA, 2017). Essa ferramenta apresenta algumas vantagens que justificam sua popularidade e maior utilização, por exemplo: (i) capacidade de produzir imagens tridimensionais a partir de amostras espessas, (ii) capacidade de obter imagens de planos específicos sem a necessidade de submeter a amostra a seccionamentos físicos, (iii) eliminação de artefatos na imagem tridimensional final, (iv) redução do tempo de processamento da amostra até a obtenção da imagem, (v) possibilidade da observação de amostras vivas. Em relação aos microscópios de campo convencionais pode-se dizer que a microscopia confocal oferece várias vantagens: controla a profundidade de campo, reduz a informação periférica ao plano focal, aumenta o contraste, melhora a qualidade das imagens e possibilita maior resolução nos componentes axiais (lateral e vertical) (CROIX; SHAND; WATKINS, 2005).

Em *Drosophila*, o processamento de imagens obtidas por varredura confocal de embriões corados é uma abordagem muito utilizada para a aquisição de dados quantitativos de expressão gênica na resolução de uma única célula. A principal vantagem dessa ferramenta sobre outras abordagens é a capacidade de medir os níveis de expressão ao longo do embrião, espaço que fornece campos de expressão 2D e possibilita a reconstrução 3D de um embrião inteiro. Como desvantagem, só é possível medir simultaneamente um número relativamente pequeno de produtos gênicos de uma determinada amostra (BAKER, 2010; BOROVEC; KYBIC; NAVA, 2017). Isso significa que um mapa de expressão espacial de todos os genes de segmentação deve ser sintetizado a partir de muitas observações, cada uma feita em um embrião separadamente. Embora os dados possam ser obtidos de diferentes combinações dos produtos gênicos, esses padrões não podem ser diretamente sobrepostos devido às distorções causadas por erros de observação e diferenças individuais implícitas entre os embriões tratados (SPIROV et al., 2001).

# CAPÍTULO 3

## Trabalhos correlatos

Neste capítulo são apresentados alguns trabalhos correlatos, com foco no processamento e análise de imagens com dados de morfologia ou expressão gênica da *Drosophila*. São discutidas algumas bases de dados e as técnicas para processamento dessas imagens.

### 3.1 Trabalhos correlatos

Os padrões de expressão gênica relativos à *Drosophila*, analisados ao longo do desenvolvimento embrionário, oferecem informações importantes a respeito das funções e do papel regulador dos genes. Na literatura é possível encontrar vários trabalhos envolvendo o processamento computacional desse tipo de imagem, notadamente com interesse no estudo das expressões gênicas estabelecidas nos estágios iniciais do desenvolvimento embrionário e também do estágio larval. Os trabalhos encontrados reforçam a importância desta área de pesquisa, e evidenciam algumas lacunas que ainda permanecem abertas.

Nesse contexto, na Tabela 1 são listados diversos trabalhos propostos ao longo dos últimos quinze anos, que apresentam propostas de técnicas e algoritmos para o processamento e análise de imagens com dados de morfologia ou expressão gênica. Os resultados desses trabalhos descrevem importantes descobertas relacionadas aos processos biológicos e a complexa cascata de regulação gênica durante o desenvolvimento.

O levantamento da literatura correlata mostra que nenhum dos trabalhos apresentados propõe um sistema mais genérico para a extração de informações quantitativas a partir de imagens de embriões da *Drosophila*, enfatizando a originalidade da pesquisa ora apresentada. Neste capítulo, bases de dados que abordam especificamente imagens de expressão gênica dos embriões de *Drosophila* (Seção 3.2) e trabalhos correlatos (Seções 3.3– 3.6) são brevemente descritos, a fim de exemplificar soluções computacionais para a análise e processamento de padrões de expressão gênica baseados em imagens bidimensionais de embriões da *Drosophila*. Os trabalhos apresentados foram selecionados por terem sido utilizados para comparação com os módulos propostos nesta tese.

Tabela 1 – Lista de trabalhos correlatos.

Referência	Foco do trabalho	Características das imagens
(SURKOVA et al., 2019)	Análise quantitativa	Superfície subcelular
(BOROVEC; KYBIC; NAVA, 2017)	Análise gênica global Oogenesis	Superfície subcelular
(KUMAR et al., 2017)	Ferramenta automática	Superfície celular
(LI; GONG, 2017)	Isolamento embrião principal	Superfície celular
(WILK et al., 2016)	Análise gênica Embriogênese e Larval	Superfície subcelular
(GAO et al., 2015)	Anotação automática	Superfície celular
(CICIN-SAIN et al., 2015)	Base de dados da <i>Drosophila</i>	Expressão gênica
(JANSSENS et al., 2014)	Análise quantitativa em <i>C. albipunctata</i>	Superfície subcelular
(LI; SILVEIRA; GONG, 2014)	Detecção da expressão gênica	Superfície celular
(LI; ANANTA, 2014)	Extração do contorno de embriões	Superfície celular
(ROBINSON et al., 2013)	Base de dados da <i>Drosophila</i>	Expressão gênica
(PIERRE et al., 2013)	Base de dados da <i>Drosophila</i>	Genes e genomas
(SURKOVA et al., 2013)	Análise gênica embriogênese	Superfície subcelular
(YUAN et al., 2013)	Anotação automática	Superfície celular
(GARCIA et al., 2013)	Análise quantitativa	Seção transversa
(KAZMAR et al., 2013)	Detecção do embrião	Superfície celular
(KONIKOFF et al., 2012)	Padronização supervisionada de embriões	Superfície celular
(GOLYANDINA et al., 2012)	Segmentação nuclear	Superfície subcelular
(CROMBACH et al., 2012)	Análise quantitativa da faixa central	Superfície celular
(KNOWLES, 2012)	Análise gênica embriogênese	Superfície subcelular
(REEVES et al., 2012)	Análise quantitativa	Seção transversa
(LI; KAMBHAMETTU, 2011)	Extração do contorno de embriões	Superfície celular
(MACE et al., 2010)	Extração do contorno de embriões	Superfície celular
(PUNIYANI; FALOUTSOS; XING, 2010)	Pipeline para expressão gênica	Superfície celular
(WEBER et al., 2009)	Análise gênica 3D	Superfície subcelular
(PISAREV et al., 2009)	Segmentação nuclear	Superfície subcelular
(LI; XING; SHINDHELM, 2008)	Padronização parcial de embriões	Superfície celular
(LI; XING; SHINDHELM, 2008)	Padronização de imagens	Superfície celular
(TOMANCAK et al., 2007)	Análise gênica global Embriogênese	Superfície celular

## 3.2 Bases de imagens

Padrões espaciais de expressão gênica são a base para estudos da biologia molecular. Milhares de imagens contendo esse tipo de dados foram publicadas na literatura (KUMAR et al., 2002; WILK et al., 2016; TOMANCAK et al., 2002). Com o aumento do volume desse tipo de informação, as bases se tornaram cada vez mais úteis, uma vez que proporcionam ambientes de alto desempenho para o armazenamento, controle e análise dos dados de expressão gênica espaço-temporal. É possível encontrar diferentes bases na literatura que disponibilizam imagens de expressão gênica dos embriões de *Drosophila*. A maioria é disponível gratuitamente online. Na Tabela 2 são descritas algumas dessas bases, e as subseções seguintes descrevem as bases que estão mais relacionadas a esta tese.

Tabela 2 – Lista de bases de imagens correlatas.

Bases de imagens		
Nome	Descrição	Referência
BDGP	Imagens 2D da expressão gênica dos embriões de <i>Drosophila</i>	(TOMANCAK et al., 2007)
BDTNP	Imagens 3D da expressão gênica dos embriões de <i>Drosophila</i>	(HENDRIKS et al., 2006)
FlyEx	Imagens 2D da expressão gênica dos embriões de <i>Drosophila</i>	(POUSTELNIKOVA et al., 2004)
FlyExpress	Imagens 2D da expressão gênica dos embriões de <i>Drosophila</i>	(KUMAR et al., 2011)
FlyBase	Dados genéticos e genômicos do organismo modelo <i>Drosophila melanogaster</i> e toda a família de insetos <i>Drosophilidae</i>	(GRUMBLING; STRELETS, 2006)
Fly-FISH	Imagens 2D da expressão gênica dos embriões de <i>Drosophila</i>	(LECUYER et al., 2007)
FlyAtlas	Dados de expressão gênica nos tecidos de <i>Drosophila melanogaster</i> .	(ROBINSON et al., 2013)
SuperFly	Dados de expressão espaço-temporal quantificados de genes de segmentação em diferentes espécies de insetos dípteros	(CICIN-SAIN et al., 2015)
FlyMine	Dados genômicos e proteômicos para <i>Drosophila</i> e outros insetos	(LYNE et al., 2007)

### 3.2.1 Fly-FISH

A base de dados Fly-FISH documenta padrões de expressão e localização de mRNAs de *Drosophila* ao nível celular e subcelular durante a embriogênese e terceira fase do estágio larval. Um método de detecção de fluorescência de alta resolução e alta produtividade é usado para detectar mRNAs expressos (WILK et al., 2016). Os achados gerais e as implicações do trabalho realizado estão resumidos em (LECUYER et al., 2007) e (WILK

et al., 2016). Os dados podem ser acessados pesquisando as categorias de localização, pesquisando genes específicos ou percorrendo a lista de genes analisados.

### 3.2.2 FlyExpress 7

FlyExpress<sup>1</sup> é uma base de dados online composta por uma série de imagens relativas à padrões de expressão gênica de embriões da *Drosophila melanogaster* coletadas durante a embriogênese. As imagens provêm de bases como o BDGP (TOMANCAK et al., 2002), Fly-FISH (LECUYER et al., 2007) e publicações de outros trabalhos (PubMed), contendo atualmente mais de 100.000 imagens de expressão de quase 5.000 genes derivados de dois estudos de hibridização *in situ* de alto rendimento (TOMANCAK et al., 2002; LECUYER et al., 2007), e composta ainda, por mais de 30.000 imagens extraídas de publicações correlatas.

As imagens disponíveis no FlyExpress são apresentadas como GEMs (*Genomewide Expression Maps*), isto é, mapas de expressão de todo o genoma, que resumem os padrões espaciais dos genes expressos em um determinado estágio do desenvolvimento da mosca. Esses mapas não apenas fornecem uma pista visual de onde os genes são expressos, como também quantos deles são expressos na mesma vizinhança, sendo que as regiões mais escuras do mapa de calor se correlacionam com uma maior quantidade de genes expressos. Eles são construídos pela agregação e normalização de padrões de expressão espacial que compartilham o mesmo estágio de desenvolvimento e a mesma orientação (ver métodos em (KONIKOFF et al., 2012)). O foco principal do FlyExpress é a proposição de um mecanismo de busca para encontrar genes co-expressos e pesquisas científicas relacionadas. Seu objetivo é facilitar a mineração analítica e visual dos dados de entrada. Para isso, o FlyExpress propõe um mecanismo de busca que se baseia em objetos gráficos, i.e., imagens dos padrões de expressão gênica, diferentemente da metodologia tradicional de vocabulário controlado. Simplificadamente, a abordagem permite ao usuário clicar em locais específicos das imagens de embriões da *Drosophila*, em diferentes estágios do desenvolvimento e projeções visuais, e com isso produzir uma lista de genes e publicações correlatas. Em geral, o FlyExpress propõe uma nova metodologia para resumir os dados de imagens dos embriões de *Drosophila*, visando facilitar a geração de hipóteses sobre interações gênicas e acelerar o processo de novas descobertas biológicas (KUMAR et al., 2017).

### 3.2.3 FlyEx

A FlyEx<sup>2</sup> é uma base de dados online voltada o armazenamento e análise de imagens 2D relativas à padrões de expressão gênica dos embriões de *Drosophila melanogaster*.

<sup>1</sup> <<http://www.flyexpress.net>>

<sup>2</sup> <<http://urchin.spbcas.ru/flyex/>>

Ela possibilita ao usuário monitorar e analisar a dinâmica de formação dos domínios de expressão gênica dos embriões durante todo o período de determinação do segmento, o que equivale a 1,5 h de desenvolvimento. Atualmente essa base armazena cerca de 4716 imagens de 14 padrões de expressão de genes de segmentação obtidos de 1579 embriões e 9.500.000 registros de dados quantitativos. Dados de referência estão disponíveis para todos os genes de segmentação nos ciclos 11-13 e todas as classes temporais do ciclo 14A (PISAREV et al., 2009).

A FlyEx suporta operações básicas nas imagens da base, como por exemplo, visualização da máscara de segmentação, dos padrões (em níveis de cinza e nos diferentes canais RGB) e *download* das imagens em diferentes resoluções. O banco de dados pode ainda ser usado para examinar a qualidade dos dados, analisar a dinâmica de formação dos domínios de expressão gênica, bem como, estimar a variabilidade dos padrões de expressão gênica (POUSTELNIKOVA et al., 2004). Trata-se de uma das bases de imagens mais populares disponíveis online, sendo amplamente utilizada (JANSSENS et al., 2013; JANSSENS et al., 2005; SURKOVA et al., 2011).

### 3.2.4 Berkeley Drosophila Transcription Network Project

A Berkeley Drosophila Transcription Network Project (BDTNP)<sup>3</sup> é uma base online de imagens de expressão gênica obtidas por microscopia confocal, em três dimensões (3D) (FOWLKES et al., 2008). Essa base faz parte de um projeto científico focado no estudo e análise da rede de regulação gênica da embriogênese da *Drosophila melanogaster*. Os autores mantêm a base de imagem, bem como, suportam as ferramentas e métodos computacionais necessários para a manipulação de seus dados (FOWLKES et al., 2008).

Atualmente o projeto analisa 37 fatores de transcrição materno e zigótico como os principais reguladores da cascata transcrecional de formação do padrão embrionário. O projeto tem como objetivo desenvolver técnicas para decifrar informações do processo de regulação transcrecional codificadas no genoma e modelar quantitativamente como os padrões espaciais e temporais de expressão gênica se desenvolvem dentro da rede (HEN-DRIKS et al., 2006).

De maneira geral, a BDTNP permite a busca e recuperação de três tipos de arquivos que descrevem a expressão gênica e a morfologia dos embriões *Drosophila*, a saber *PointCloud*, *VirtualEmbryo* e *Correspondence*. *PointCloud* descreve os dados de um embrião individual, tipicamente a expressão de dois genes. Cada arquivo *VirtualEmbryo* descreve a média dos dados para muitos genes, embriões e estágios de desenvolvimento. E cada arquivo *Correspondence* descreve a melhor correspondência entre cada célula de um *PointCloud* e a(s) célula(s) associada no *VirtualEmbryo*.

Por fim, é relevante mencionar que a BDTNP propõe trabalhos na literatura que descrevem detalhadamente a sua base de imagens (FOWLKES et al., 2008), bem como,

<sup>3</sup> <<http://bdtnp.lbl.gov/Fly-Net/>>

o pipeline de análise tridimensional desenvolvido para analisar a morfologia e a expressão gênica dos embriões de *Drosophila* (HENDRIKS et al., 2006; KERANEN et al., 2006).

### 3.2.5 Berkeley *Drosophila* Genome Project

O Berkeley *Drosophila* Genome Project (BDGP)<sup>4</sup> representa um projeto de pesquisa abrangente, focado no estudo e investigação do genoma da *Drosophila melanogaster*. Trata-se de um consórcio do *Drosophila* Genome Center financiado pelo Instituto Nacional de Pesquisa do Genoma Humano, Instituto Nacional do Câncer e Instituto Médico Howard Hughes. O BDGP (TOMANCAK et al., 2007) suporta o desenvolvimento de pesquisas diversas, tais como, sequenciamento genômico (HOSKINS et al., 2015), sequenciamento de EST'S (Etiquetas de Sequência Expressa) (STAPLETON et al., 2002) e análise de expressão gênica (TOMANCAK et al., 2002; TOMANCAK et al., 2007). Seus objetivos principais se concentram em (a) concluir a sequência do genoma eucromático da *Drosophila melanogaster* em alta qualidade, bem como, gerar e manter anotações biológicas dessa sequência. Além do sequenciamento genômico, o BDGP visa (b) produzir disruptões gênicas, (c) caracterizar a sequência e expressão de cDNAs e (d) desenvolver ferramentas de informática capazes de suportar o processo experimental, identificar características das sequências de DNA e de apresentar informações atualizadas sobre anotações de sequências (TOMANCAK et al., 2007). Os trabalhos do BDGP mais relevantes para o contexto da presente pesquisa dizem respeito às investigações dos padrões de expressão gênica da *Drosophila* (TOMANCAK et al., 2002; PENG et al., 2006) e ao desenvolvimento de ferramentas computacionais voltadas ao processamento e análise automática desse tipo de dados (BOROVEC; KYBIC; NAVA, 2017; YUAN et al., 2013).

O banco de dados BDGP conta com mais de 138.564 imagens digitais de expressão genética de mais de 8.450 genes. Tratam-se de imagens bidimensionais do embrião obtidas pela técnica de hibridização *in situ*. As imagens na BDGP são agrupadas por estágios de desenvolvimento, sendo 1–3, 4–6, 7–8, 9–10, 11–12 e 13–17 (TOMANCAK et al., 2002; TOMANCAK et al., 2007).

### 3.2.6 FlyBase

A FlyBase<sup>5</sup> é uma base online de dados genéticos e genômicos do organismo modelo *Drosophila melanogaster* e toda a família de insetos *Drosophilidae*. O Grupo FlyBase organiza, anota, integra e mantém uma grande variedade de dados dentro deste domínio, os quais, originam-se de uma variedade de fontes, desde projetos de genoma em larga escala (BDGP) até a literatura de pesquisa primária. Os tipos de dados disponíveis na FlyBase incluem mapas nos níveis citológico, genético e de sequência, genes e alelos incluindo seus

<sup>4</sup> <<http://www.fruitfly.org/>>

<sup>5</sup> <http://flybase.org/>

produtos, funções, padrões de expressão, fenótipos mutantes e interações gênicas, bem como, cromossomos aberrantes, genomas anotados, coleções genéticas, construções transgênicas e inserções, anatomia e imagens, dados bibliográficos e informações de contato da comunidade (GRUMBLING; STRELETS, 2006).

A FlyBase é coordenada pelo BDGP (TOMANCAK et al., 2007) (descrito na Seção 3.2.5) e fornece acesso aos dados por meio de interfaces gráficas e textuais adaptadas aos tipos específicos dos dados.

### 3.3 Padronização das imagens

O resultado da captura de imagens de embriões da *Drosophila melanogaster* dependem de dois fatores principais, (a) configuração dos microscópios utilizados e (b) preparação humana das amostras biológicas (MACE et al., 2010). As diferentes técnicas e procedimentos utilizados contribuíram para a geração de diferentes repositórios de dados (TOMANCAK et al., 2007; POUSTELNIKOVA et al., 2004; HENDRIKS et al., 2006; KUMAR et al., 2011; LECUYER et al., 2007), os quais geralmente disponibilizam imagens sem normalização (YUAN et al., 2013). Sendo assim, a maioria apresenta imagens em diferentes tamanhos, orientações, direções, posições, escalas e condições de iluminação e textura (LI; GONG, 2017; KUMAR et al., 2002; NUTAKKI, 2010). Essas variações podem ser reunidas em dois grupos distintos, a saber: (a) eixo anatômico dos embriões e (b) condições gerais da imagem. No caso (a), as variações referem-se à propriedades que caracterizam o eixo anatômico dos embriões de *Drosophila*, tais como orientação, direção e posição. No caso (b), é possível citar por exemplo, variações de textura, iluminação, tamanho e escala. Ainda segundo o autor, as inconsistências mais comuns e que prejudicam severamente os resultados das análises biológicas resultam do grupo (a). No trabalho de (TURNER et al., 2011) é destacado que algumas propriedades variam mesmo com embriões de espécies iguais, e podem ainda, acentuar-se em diferentes estágios do ciclo do desenvolvimento embrionário (KAZMAR et al., 2013).

A variabilidade das propriedades descritas, quando não controlada, altera significativamente o resultado dos procedimentos automáticos, tais como comparação (MACE et al., 2010; PUNIYANI; XING, 2013), anotação (KAZMAR et al., 2013; YUAN et al., 2013), classificação (HEFFEL et al., 2009), agrupamento (PENG et al., 2006), mineração (PAN et al., 2006) e reconhecimento (NUTAKKI, 2010). A padronização objetiva uniformizar um conjunto de imagens, a fim de prepará-las para que as análises biológicas posteriores sejam realizadas de forma efetiva. Variações das imagens de embriões da *Drosophila* incluem mudanças na orientação, escala, tamanho e direção (PAN et al., 2006; KONIKOFF et al., 2012; LI; GONG, 2017), e ainda, presença de embriões parciais ou completos (MACE et al., 2010). Mais seriamente, a aparência dos embriões e dos padrões de expressão gênica, também podem variar significativamente devido a diversidade dos

ciclos de desenvolvimento embrionário (LI; XING; SHINDHELM, 2008). Essas variações desafiam a realização automática de análises biológicas que necessitam de dados uniformes, tais como recuperação, registro e classificação de imagens de expressão gênica (PAN et al., 2006; MACE et al., 2010; KONIKOFF et al., 2012; YUAN et al., 2013; HEFFEL et al., 2009). Na literatura é possível encontrar alguns trabalhos relativos à padronização de imagens dos embriões de *Drosophila* (PAN et al., 2006; KONIKOFF et al., 2012; LI; XING; SHINDHELM, 2008; KUMAR et al., 2002; GARGESHA et al., 2005; PENG; MYERS, 2004; PAN et al., 2006).

Algumas abordagens são supervisionadas, sendo assim, em alguma etapa da execução requerem manipulação manual significativa (KUMAR et al., 2002). Outras, por sua vez, tratam apenas um subconjunto de variações da imagem (PENG; MYERS, 2004), como é o caso da pesquisa de (PUNIYANI; FALOUTSOS; XING, 2010) que propuseram um método para ajuste da orientação dos embriões de *Drosophila* com base na seleção de padrões de expressão gênica semelhantes. Nem todas as propostas são eficientes para os casos de imagens em que há a presença de dois ou mais embriões que se tocam. Assim, é possível encontrar soluções específicas para detectar e isolar o embrião principal das imagens (PAN et al., 2006; LI; XING; SHINDHELM, 2008). Por fim, é importante mencionar que alguns trabalhos encontrados na literatura propõem soluções específicas para a padronização de imagens dos embriões de *Drosophila* (PUNIYANI; FALOUTSOS; XING, 2010), ao passo que outros abordam esta tarefa superficialmente, descrevendo como uma das fases incluídas em suas metodologias propostas (JANSSENS et al., 2005; SURKOVA et al., 2008). Por sua relevância e destaque, será descrito simplificadamente a proposta de (KONIKOFF et al., 2012). Nessa proposta, os autores apresentam um método semi-supervisionado para a padronização das propriedades orientação, tamanho e direção das imagens de embriões de *Drosophila*. O método padroniza ainda as imagens com relação à quantidade de embriões, configurando apenas um embrião por imagem. Essa última tarefa é realizada manualmente, em que o objetivo é isolar o embrião principal de outros embriões secundários ou partes de amostras indesejadas.

O método consiste de uma série de rotinas desenvolvidas em MATLAB que são executadas sequencialmente. Primeiramente um filtro passa-baixa é aplicado sobre a imagem em tons de cinza, visando com isso suavizar e manipular os dados para a etapa posterior. Consequentemente, o detector de bordas *Canny* (CANNY, 1986) é utilizado para capturar a borda do embrião. Após isso, uma série de operações morfológicas (dilatação, preenchimento, abertura, suavização e erosão) é aplicada sobre a imagem visando definir o contorno do embrião. Por fim, os pixels não pertencentes ao embrião são configurados para branco. Após separar o embrião do fundo da imagem, o método padroniza a direção dos embriões. Esta tarefa é realizada a partir da função *imrotate* do Matlab. De modo simplificado, essa função gira a imagem por um dado ângulo no sentido anti-horário em torno do seu ponto central. O ângulo é um escalar que especifica o grau existente entre

o eixo  $X$  da imagem e o eixo principal do objeto (embrião). A função *imrotate* usa interpolação de vizinho mais próximo para definir os pixels da imagem de saída que situam-se fora da imagem rotacionada para 0 (zero). Dessa forma, a direção dos embriões é padronizada horizontalmente. Entretanto, a orientação não é definida automaticamente, ficando a cargo de supervisão humana. Por consistência, os autores seguem a orientação comumente adotada pela literatura, i.e., região anterior à esquerda e lado dorsal para topo no caso de visualização lateral. E região anterior à esquerda para as outras visualizações, como dorsal e ventral (KUMAR et al., 2002). Assim, quando necessário, a orientação é ajustada manualmente por meio da função *flipdim* do Matlab. Por fim, o tamanho de todas as imagens é padronizado com o uso da função *imresize*. Os autores finalizam o trabalho descrevendo que as tarefas de revisão e atribuição dos ciclos de desenvolvimento e da visão dos embriões também não é automatizada.

### 3.4 Pré-processamento

O pré-processamento é um passo importante dos sistemas de análise automática. Ele visa melhorar e adequar a imagem para as etapas posteriores de um sistema. As imagens de embriões da *Drosophila* geralmente apresentam uma textura rugosa (nas regiões dos núcleos celulares). E ainda, costumam apresentar detalhes espúrios na região do plano de fundo (i.e, a solução aquosa ao qual o embrião é envolvido). No caso dos sistemas automáticos da *Drosophila*, o pré-processamento é essencial para destacar o embrião principal ou áreas de interesse (como os núcleos celulares), bem como, reduzir ruídos de aquisição característicos desse tipo de imagem.

Na literatura, as propostas de pré-processamento encontradas geralmente são descritas superficialmente e, comumente, abordam técnicas tradicionais existentes, tais como filtros de suavização e operadores morfológicos. Um dos trabalhos propostos na literatura que descreve detalhadamente a etapa de pré-processamento é proposto por (JANSSENS et al., 2005). Esse trabalho propõe um conjunto de rotinas automáticas para o processamento e análise de imagens bidimensionais relativas à embriões da *Drosophila melanogaster*. O objetivo principal é a aquisição de dados quantitativos a partir de imagens confocais dos padrões de expressão gênica da *Drosophila*. Para isso, os autores propuseram três etapas principais, sendo (a) pré-processamento, (b) segmentação e (c) análise. Em (a), o foco consiste na utilização das técnicas binarização, equalização de histograma, operações morfológicas (dilatação, erosão) e filtros de suavização (mediana) (GONZALEZ; WOODS, 2010). Primeiramente, uma limiarização simples é aplicada ao canal de intensidade máxima da imagem original. O objetivo é obter a máscara do embrião, isto é, uma imagem binária em que os pixels pertencentes ao embrião são ajustados para 1 (um) e os pixels do fundo da imagem (região externa do embrião) são assinalados para 0 (zero). O resultado dessa etapa pode ser corrigido com a aplicação da operação morfológica preenchimento.

Posteriormente, a borda da máscara é suavizada com um filtro mediano, que substitui cada pixel pelo valor mediano dos pixels vizinhos. Na sequência, operações morfológicas (erosão e dilatação) são aplicadas a fim de remover detalhes espúrios das bordas e reduzir buracos da máscara do embrião, e ainda, eliminar possíveis pixels encontrados fora do embrião. A máscara obtida é então utilizada para rotacionar e cortar o embrião principal. Após a imagem ser rotacionada e centralizada, os autores aplicam uma equalização de histograma, filtro de suavização mediano e realiza uma inversão. A combinação dessas técnicas permite melhorar o contraste, reduzir o ruído e suavizar a borda dos núcleos celulares. Com isso, pode-se iniciar a etapa de segmentação. Os autores sugerem uma combinação dos algoritmos *Watershed* (GONZALEZ; WOODS, 2010) e detector de bordas Shen–Castan (SHEN J.; CASTAN, 1992). Como último passo, é proposto a aplicação da operação morfológica erosão seguida de reconstrução mínima e um passo final de erosão, a fim de remover pequenos buracos e objetos incompletos que podem permanecer na borda do embrião.

### 3.5 Segmentação

A segmentação de imagens é uma tarefa chave em sistemas de Processamento Digital de Imagens (PDI), com uma vasta gama de aplicações em sistemas de biologia computacional relativos ao contexto da *Drosophila melanogaster*, que inclui segmentação dos embriões (LI; GONG, 2017), dos núcleos celulares (JANSSENS et al., 2014), dos padrões de expressão (PUNIYANI; FALOUTSOS; XING, 2010), das asas (NETO; PAULA, 2017) e dos ovos (JIA et al., 2016). A segmentação de imagens relativas a embriões da *Drosophila* enfrenta desafios quando realizada pela aplicação isolada de técnicas tradicionais da literatura. As dificuldades encontradas normalmente referem-se a ruídos existentes nas imagens e à variabilidade dos dados. É comum encontrar imagens com baixo contraste, iluminação irregular e ruído.

Os métodos de segmentação geralmente são menos genéricos que os métodos relativos à outras etapas, como pré-processamento. Eles costumam tratar tipos específicos de imagens e tarefas particulares (POUSTELNIKOVA et al., 2004; KOSMAN; REINITZ; SHARP, 1999). Nesse sentido, pode-se encontrar na literatura trabalhos propostos para os diferentes planos anatômicos (sagital e frontal) e eixos dos embriões de *Drosophila* (AP e DV). Dentro deste contexto, existem soluções para (a) segmentação de núcleos (SURKOVA et al., 2008; SOUSA et al., 2013), (b) detecção da expressão gênica (TOMANCAK et al., 2007) e (c) localização do embrião principal em imagens de *Drosophila* contendo vários embriões imersos em solução aquosa (LI, 2014; LI; ANANTA, 2014; LI; GONG, 2017).

A localização de embriões em imagens de *Drosophila* é um passo fundamental no desenvolvimento de sistemas computacionais relativos ao processamento e análise de padrões

de expressão gênica. A automatização dessa tarefa é um desafio devido à variações diversas desse tipo de imagem, por exemplo (i) o tamanho, orientação e forma do embrião de interesse; (ii) a vizinhança do embrião principal (que toca, ou não, na amostra desejada) e (iii) características da iluminação. Dentro deste contexto, a aplicação direta de técnicas tradicionais da literatura não consegue obter resultados desejáveis na detecção do contorno de embriões da *Drosophila* (PENG; MYERS, 2004; PAN et al., 2006; LI; KAMBHAMETTU, 2011). Diversas abordagens têm sido propostas ao longo dos anos para contornar esta questão (GARGESHA et al., 2005; PENG; MYERS, 2004; PAN et al., 2006). (PENG; MYERS, 2004) propuseram uma abordagem que calcula o desvio padrão de janelas locais sobrepostas sob a imagem do embrião para caracterizar os pixels como pixels de primeiro ou segundo plano, e aplica um método de crescimento de região para localizar o contorno do embrião. (PAN et al., 2006) propuseram uma variante do algoritmo de Marquardt-Levenberg para calcular uma transformação afim ótima, visando registrar embriões localizados em uma região elipsoidal. (FRISE; HAMMONDS; CELNIKER, 2009) estenderam o método de (PENG; MYERS, 2004) adicionando três operações em uma imagem binária, sendo i) remoção de pixels isolados, ii) dilatação e iii) processamento majoritário. Além disso, (FRISE; HAMMONDS; CELNIKER, 2009) propuseram um algoritmo heurístico para separar o embrião de interesse de múltiplos embriões adjacentes, partindo da suposição de que o centro do embrião de interesse coincide com o centro da imagem. (MACE et al., 2010) propuseram um método para extraer o contorno dos embriões, em que foi utilizado um otimizador de enxame de partículas para reduzir o custo computacional na busca de parâmetros ótimos. (PUNIYANI; FALOUTSOS; XING, 2010) propuseram um método baseado na detecção de bordas que envolve um conjunto de restrições heurísticas, incluindo tamanho do objeto, convexidade, características da forma (por exemplo, proporção do eixo maior sobre o menor) e a porcentagem da região sobreposta. (LI; KAMBHAMETTU, 2011) propuseram um modelo de curva quadrática para inicializar o contorno do embrião de interesse baseado nos pixels da borda, e aplicaram um modelo de contorno ativo para refinar os contornos do embrião. (BESSINGER; XING; LI, 2012) propuseram critérios para selecionar pixels de borda do componente conectado ideal no espaço de escala da imagem de entrada. (LI, 2014) propôs algoritmos para detectar e restaurar deficiências e falhas de *tokens* primários de esboço que ocorrem quando um objeto de interesse é cercado por um fundo complexo.

Dentre os trabalhos mais recentes que propõem soluções para a localização de embriões em imagens da *Drosophila*, é descrito nessa seção o método proposto por (LI; GONG, 2017). Neste, os autores apresentam um método para extração do contorno de embriões da *Drosophila*, tendo a extração de K pontos dominantes como chave do método. Em trabalho anterior (LI, 2014), os mesmos autores propuseram a extração de 3 pontos dominantes para analisar a estrutura geométrica de formas retangulares. Nesse trabalho a proposição consiste em generalizar a extração de 3 pontos dominantes para a

extração de  $K$  pontos dominantes, visando adaptar o método para a estrutura geométrica complexa do contorno de um embrião de *Drosophila*. A complexidade dessas estruturas reside principalmente em dois aspectos, (i) o contorno de um embrião de *Drosophila* pode ser côncavo e (ii) o contorno de dois embriões que se tocam pode ser côncavo. Simplificadamente, o método consiste em aproximar os pixels de borda do componente conectado por um polígono que pode ser convexo ou côncavo. E com isso ele é capaz de segmentar embriões que se tocam, característica esta que pode não ser atendida por outros métodos.

### 3.6 Representação e quantificação dos dados

A representação e quantificação dos padrões de expressão gênica diz respeito à extração de informações quantitativas a partir das imagens de embriões da *Drosophila melanogaster* em diferentes fases do ciclo de seu desenvolvimento. De modo geral, essa etapa envolve as tarefas de identificação das regiões de interesse (ROIs), definição do formato de representação, remoção do ruído de fundo (*background*) e extração dos níveis médio de expressão gênica. Nesta seção são descritas duas abordagens relativas a embriões de *Drosophila* capturados da seção transversal, as propostas de (COPPEY et al., 2008) e (TRISNADI et al., 2013).

No trabalho de (TRISNADI et al., 2013) são propostos métodos computacionais para medir o gradiente de expressão de proteínas ao longo do eixo dorso-ventral no embrião de *Drosophila*. A abordagem consiste em cinco passos, (i) detectar a borda do embrião, (ii) calcular as intensidades médias citoplasmáticas, (iii) localizar núcleos, (iv) calcular as intensidades médias nucleares e (v) medir pontos nucleares. A principal diferença está na forma como a região nuclear e citoplasmática são determinadas. Primeiro, presume-se que o centro da imagem resida dentro do embrião. E assim, a imagem é dividida em 60 domínios irradiando do centro da imagem (vide Fig. 8). Resumidamente, a borda do embrião é encontrada computacionalmente, e então os núcleos são segmentados usando um protocolo de limiarização local. Para encontrar os núcleos, os autores usam 60 pontos na borda do embrião - a periferia externa da seção transversa do embrião (até 20 micrônus de profundidade) é desenrolada em uma longa faixa, que se transforma em uma máscara nuclear binária. Em seguida, o algoritmo *Watershed* é usado para definir os limites entre os núcleos e determinar a localização das regiões citoplasmáticas. Essas localizações 1D são então mapeadas de volta na faixa desenrolada original para definir regiões retangulares, onde dentro de cada retângulo há exatamente um núcleo. Uma vez determinada a máscara nuclear, a intensidade de proteína em cada núcleo segmentado foi calculada como a razão entre a intensidade no canal de proteína dividida pela intensidade no canal nuclear. A intensidade da expressão de mRNA foi calculada como a intensidade média dentro de um anel com cerca de  $18\mu\text{m}$  de largura em torno do perímetro do embrião.

(TRISNADI et al., 2013) não propuseram a detecção individual das regiões citoplasmáticas.

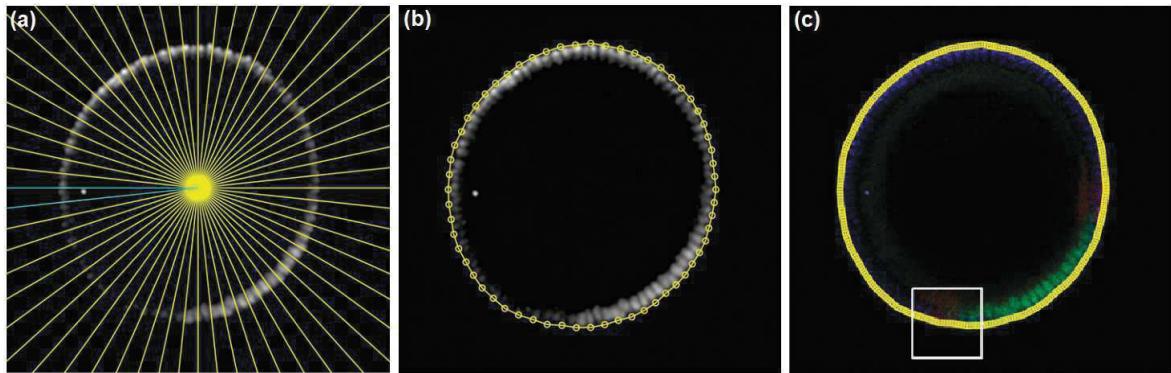


Figura 8 – (a) Divisão da imagem em 60 domínios que irradiam a partir do centro da imagem (linhas amarelas). (b) Determinação dos pontos onde a borda do embrião é localizada. (c) Imagem RGB de um embrião para o qual a fronteira foi calculada. O anel amarelo que envolve o embrião é a malha densa de 300 pontos. Figura adaptada de (TRISNADI et al., 2013).

máticas (apicais e basais) para cada núcleo dos embriões. A estratégia de detecção das regiões de interesse é limitada a formas com formato circular. E ainda, eles adotam uma estratégia de segmentação baseada nos algoritmos Watershed e Otsu para detectar o núcleo individual de cada retângulo.

Em relação ao trabalho de (COPPEY et al., 2008), é proposto um conjunto de métodos para a quantificação de gradientes nucleares em imagens de embriões de *Drosophila* na seção sagital. Na metodologia, o perímetro do embrião é identificado e, em seguida, a média de intensidades é calculada através de uma linha normal ao limite, até que o sinal caia para duas vezes o valor do fundo. Isso foi feito para 1000 pontos uniformemente espaçados em torno do perímetro do embrião (vide Fig. 9). Na sequência, são quantificados os gradientes nucleares, gerando o perfil ao longo do comprimento do arco. O tamanho da janela média deve ser maior que a distância internuclear, mas menor que o comprimento do decaimento do gradiente. E assim, o gradiente nuclear pode ser ajustado a uma função linear do gradiente citoplasmático.

Os autores estenderam o esquema anterior para quantificar ao mesmo tempo os gradientes nuclear e citoplasmático. Com base no sinal GFP, a máscara nuclear binária é construída. É importante dizer que a quantificação dos gradientes nucleares foi feita nos embriões do ciclo 13, pois neste estágio os núcleos estão claramente separados e o gradiente nuclear resultante é uma função espacialmente modulada, que é zero entre os núcleos. Essa estratégia não é capaz de detectar as regiões citoplasmáticas isoladamente. Além disso, não envolve imagens do plano transveso.

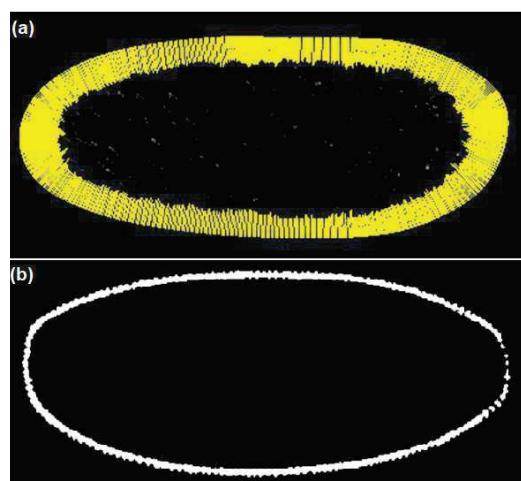


Figura 9 – Etapas na geração da máscara binária (b) correspondente a região nuclear. As linhas amarelas (a) traçam os intervalos definidos pela regra sobre os quais a concentração é integrada. Figura adaptada de (COPPEY et al., 2008).

# CAPÍTULO 4

---

## Metodologia computacional proposta

Neste capítulo é descrito a metodologia computacional proposta para processar e analisar imagens a fim de caracterizar a morfologia e os padrões de expressão gênica em embriões da *Drosophila melanogaster*.

### 4.1 Visão Geral

A *Drosophila melanogaster* tem sido extensivamente usada como um organismo modelo para estudar investigações biológicas diversas, tal como apresentado na Seção 2.1. Nesse contexto, biólogos frequentemente precisam conhecer o conjunto de genes expressos no embrião em desenvolvimento, isto é, necessitam analisar os padrões de expressão gênica. No entanto, mesmo após décadas de pesquisa ainda não é possível encontrar uma proposta computacional capaz de tratar as diferentes características desse tipo de imagem (e.g., orientação, tamanho, direção, posição e alinhamento dos embriões). As propostas encontradas normalmente oferecem soluções para um cenário específico. A fim de preencher esta lacuna, é proposta nesta tese uma metodologia computacional mais genérica, isto é, que oferece uma solução que engloba um espectro maior de imagens e análises.

Nesse contexto, a metodologia proposta constitui-se de seis módulos computacionais desenvolvidos a partir de técnicas de PDI para o processamento e análise de imagens confocais e não-confocais de embriões da *Drosophila melanogaster*. Os módulos podem ser aplicados sequencialmente ou individualmente, e ainda, em diferentes combinações com outros métodos computacionais disponíveis na literatura. A saber, citam-se: (1) Isolamento, (2) Padronização, (3) Segmentação, (4) Representação dos dados nucleares, (5) Representação dos dados do citoplasma e (6) Quantificação dos dados.

Alguns dos módulos representam soluções inovadoras, enquanto outros correspondem à estratégias desenvolvidas tanto durante o desenvolvimento da presente tese quanto em trabalhos anteriores. Propõe-se um conjunto de módulos que contemplam tarefas que vão desde a identificação do tipo de embrião até a extração de características e análise biológica dos dados. Abordagens com essa amplitude, até então, não foram encontradas

na literatura. As técnicas e a sequência de passos adotada em cada módulo mostrou-se útil e necessária para manter o caráter genérico a todos os tipos de imagens. Nas seções seguintes e na Fig. 10 são apresentados os módulos definidos na metodologia proposta.

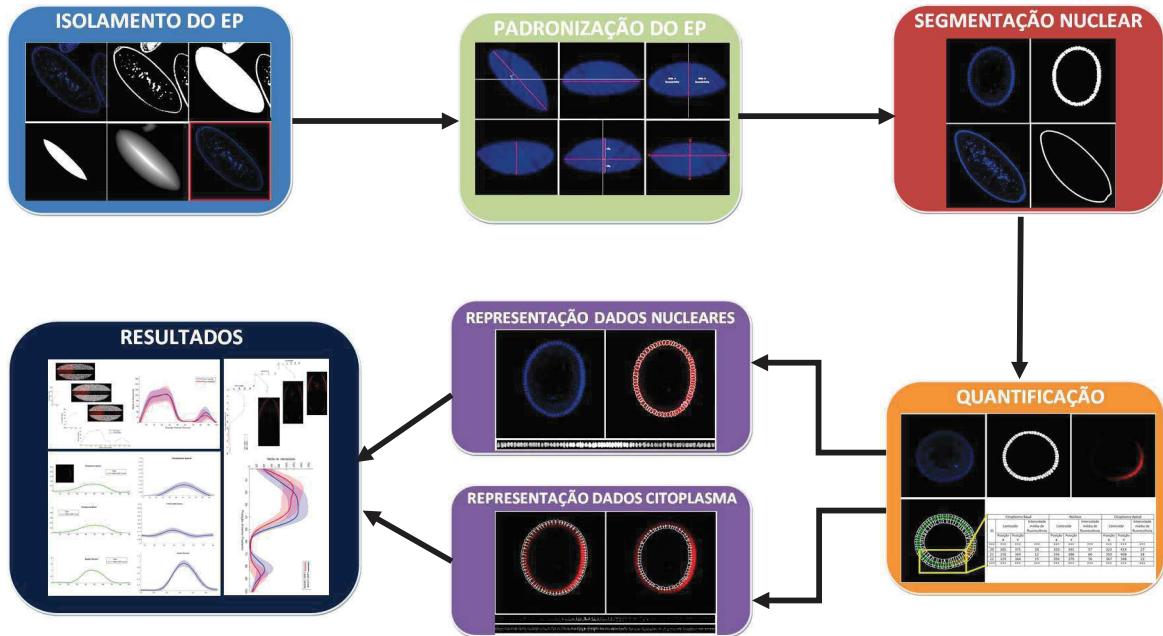


Figura 10 – Estrutura da metodologia computacional proposta. Cada retângulo destaca um módulo da metodologia. Cada módulo engloba um conjunto de métodos computacionais, desenvolvidos a partir de técnicas e algoritmos de PDI. Eles podem ser aplicados na sequência em destaque ou isoladamente, conforme a necessidade. (EP) Embrião Principal.

## 4.2 Métodos

Nas seções seguintes (4.3 à 4.8) cada módulo da metodologia computacional proposta é descrito detalhadamente. Todos os métodos inclusos nesta tese foram desenvolvidos em MATLAB (The Math Works, Natick, MA). O MATLAB é uma linguagem de alto nível que possui um ambiente interativo amigável para computação numérica, visualização e programação. As funções e bibliotecas integradas no MATLAB permitem ao usuário explorar vários problemas e encontrar soluções mais rapidamente. Ele pode ser usado para uma ampla gama de aplicações, incluindo processamento de imagens e biologia computacional. O seu *toolbox* de processamento de imagem fornece um conjunto de algoritmos e ferramentas gráficas para o processamento, análise, visualização e desenvolvimento de métodos automáticos.

### 4.3 Módulo 1 - Isolamento do embrião principal

As imagens obtidas por microscopia geralmente apresentam múltiplos embriões em cada amostra, sendo o embrião de interesse normalmente localizado no centro da imagem (PAN et al., 2006). Além disso, podem ser encontrados artefatos ou partes incompletas de embriões que devem ser eliminados. Assim, antes de começar a análise computacional de uma imagem de *Drosophila* é fundamental a identificação do embrião principal. Nesse sentido, é possível encontrar diferentes técnicas propostas na literatura para a localização do embrião principal em imagens de *Drosophila* (GARGESHA et al., 2005; PENG; MYERS, 2004; PAN et al., 2006; LI; GONG, 2017). No módulo apresentado propõe-se uma sequência de passos que visa criar a máscara binária dos embriões ou partes dos embrião presentes na imagem analisada e em seguida identificar o embrião principal, tal como descrito em (POUSTELNIKOVA et al., 2004). Os parâmetros usados nesses passos são adaptados de acordo com o tipo de imagem em análise.

Primeiramente realiza-se uma binarização na imagem original, a fim de obter o resultado inicial da máscara do embrião. Uma sequência de operações típicas de PDI é capaz de gerar o resultado esperado. A sequência inclui os seguintes passos (a) binarização, (b) filtragem por mediana, (c) dilatação, (d) preenchimento e (e) operação fechamento. O passo (a) é realizado por meio de um algoritmo de limiarização simples. Um limiar definido por Otsu é capaz de binarizar a imagem conservando o maior número de pixels relativos ao embrião. Em imagens da seção sagital/transversa, o resultado da binarização é um aglomerado de núcleos celulares. A saída do passo anterior é posteriormente tratado com filtragem por mediana, a fim de remover pixels aleatórios remanescentes no contorno do embrião. Nesse caso, a imagem binária (anteriormente obtida) é convoluída usando uma máscara de dimensões  $3 \times 3$ . Após isso, aplica-se o passo (c). A operação dilatação é aplicada buscando unificar todos os componentes obtidos da limiarização. Essa tarefa é realizada iterativamente e incrementalmente (o número de operações varia de acordo com o tipo de imagem), usando para isso um elemento estruturante do tipo “quadrado” e tamanho  $3 \times 3$ . Entretanto, somente esses passos não são suficientes para gerar a máscara completa do embrião. Sendo assim, posteriormente aplica-se dois passos finais, (d) e (e), que visam respectivamente, preencher as lacunas que permanecem abertas após a dilatação e suavizar a borda do embrião.

Após a correta identificação dos embriões na imagem, para isolar o embrião principal aplica-se a transformada de distância (ROSENFIELD; PFALTZ, 1968). O resultado dessa operação é o mapa de distâncias, isto é, uma imagem semelhante à original, em que os níveis de cinza dentro dos objetos são alterados para identificar a menor distância de cada ponto ao contorno (perímetro) da forma. O objetivo principal é utilizar esse resultado para definir um marcador que seja capaz de identificar somente o embrião principal das imagens (Fig. 11). Para definir o marcador, o mapa de distâncias é filtrado para exibir somente os pixels localizados a uma distância ( $D$ ) menor que 150 (valor definido empiricamente).

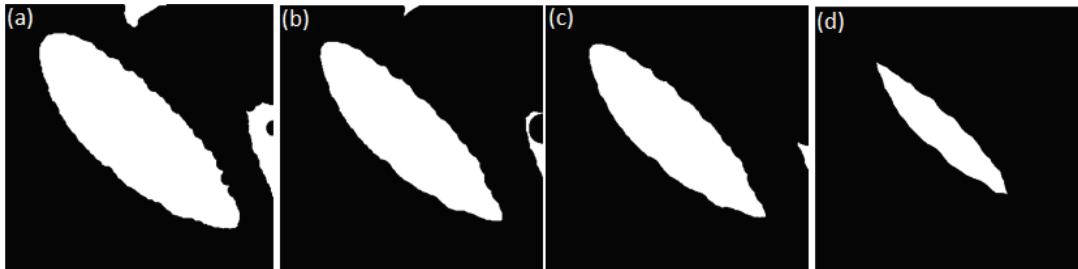


Figura 11 – Configuração de marcador para isolar o embrião principal das imagens de *Drosophila*. (a)–(d) Marcadores obtidos a partir de diferentes valores definidos em  $D$ . Em (d) é possível visualizar um marcador obtido a partir de  $D = 150$ , o qual é útil para marcar apenas o embrião principal.

Após isso, aplica-se a operação morfológica Reconstrução (GONZALEZ; WOODS, 2010), em que é executada dilatações recursivas na imagem marcador (usando um elemento estruturante (ES) “quadrado”, tamanho 3), até que seu contorno preencha completamente a região do embrião principal. De modo simplificado, a região do marcador cresce até a área específica delimitada pela imagem do embrião previamente binarizada. Na Fig. 12 são ilustradas as etapas do processo descrito.

#### 4.4 Módulo 2 - Padronização dos embriões

A padronização das imagens de embriões da *Drosophila* é crucial para o sucesso dos resultados de análise biológica (YUAN et al., 2013). Até o momento, as propostas computacionais desenvolvidas são robustas para os tipos de variações mais comuns, como tamanho e condições de iluminação (KUMAR et al., 2002; YE et al., 2008). Entretanto, as principais inconsistências encontradas na literatura resultam das propriedades relativas ao eixo anatômico dos embriões, como orientação e direção (MACE et al., 2010). Essa questão ainda permanece um desafio pois as soluções propostas na literatura geralmente são abordagens semi-supervisionados (KONIKOFF et al., 2012), que não garantem robustez e eficácia nos resultados. Além dessas, existem as técnicas supervisionadas que requerem a configuração de muitos parâmetros, os quais são obtidos por meio de processos tediosos e custosos (GARGESHA et al., 2005). Outros trabalhos apresentam algoritmos que não requerem supervisão humana, porém que baseiam-se em heurísticas que restringem o seu espectro de aplicação. Esse é caso, por exemplo, da proposta descrita por (PUNIYANI; XING, 2013), em que é apresentado um método para a detecção da orientação dos embriões de *Drosophila* a partir da ideia de que imagens relativas ao mesmo gene e no mesmo estágio de desenvolvimento apresentam padrões de expressão semelhantes.

Em virtude do cenário descrito, foi proposto o *Embystandar* como primeiro método computacional capaz de ajustar as propriedades Orientação e Direção dos embriões de *Drosophila*. Baseado em uma sequência de operações automáticas, *Embystandar* converte

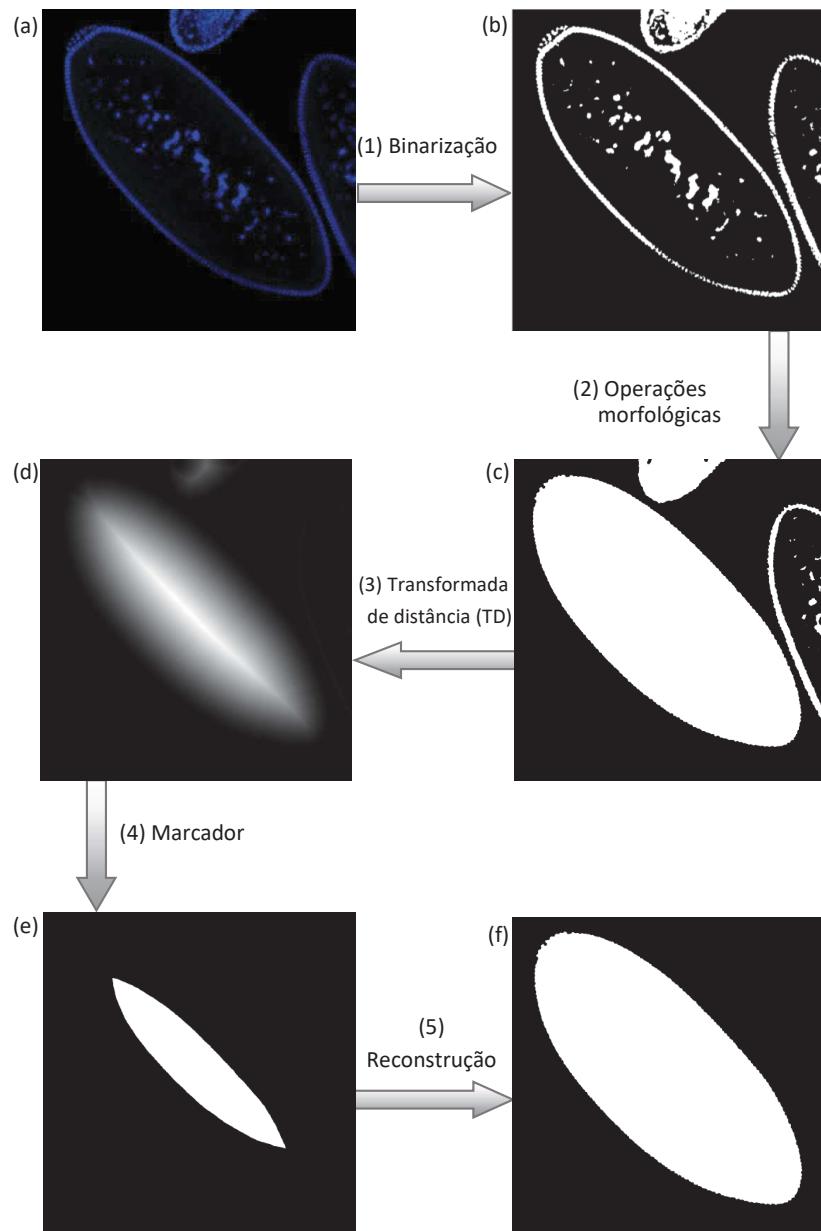


Figura 12 – Etapas para detecção do embrião principal. Este exemplo ilustra a aplicação do método em uma imagem de embrião da seção sagital. (a) Imagem original com artefatos em torno do embrião principal. (b) Resultado da binarização. (c) Máscara binária suavizada após remoção de detalhes espúrios e aplicação da operação preenchimento. (d) Resultado da Transformada Distância, que posteriormente é processada para definir o marcador. (e) Marcador. (f) Máscara obtida após operação reconstrução a partir da dilatação restrita, resultando na detecção do embrião principal.

os embriões de *Drosophila* para uma posição padrão, adequada para o processamento e análise em larga escala.

Em resumo, o método possui três estágios principais: (i) ajuste das propriedades Orientação e Direção, (ii) definição dos eixos e setores do embrião e (iii) rotação, espelhamento e corte.

#### 4.4.1 Direção e orientação

A direção refere-se ao posicionamento do embrião durante o processo de aquisição da imagem, e ela pode ser horizontal, vertical e diagonal (em diferentes rotações). A orientação refere-se ao sentido dos eixos, e esta pode ser anterior-posterior (AP) ou posterior-anterior (PA), e dorsal-ventral (DV) ou ventral-dorsal (VD). Na Fig. 13 são ilustrados embriões de *Drosophila* em orientações e direções diferentes.

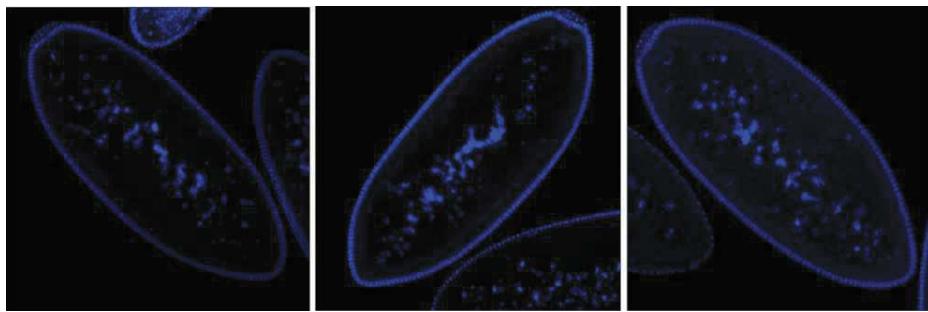


Figura 13 – Exemplo de diferentes imagens em que é possível observar as variações de direção e orientação dos embriões capturados.

A padronização das propriedades Direção e Orientação é essencial para o sucesso da maioria dos procedimentos automáticos. Além disso, muitas análises biológicas, como anotação automática dos estágios de desenvolvimento dos embriões, só são significativas com a utilização dos dados uniformizados (YUAN et al., 2013). Nesse contexto, o objetivo geral do *Embryostandar* é alinhar os embriões de *Drosophila* ao longo do mesmo eixo anatômico. O alinhamento adotado como padrão é direção horizontal e orientação AP e DV (KUMAR et al., 2002; JANSSENS et al., 2005). Em conformidade, o método visa garantir que a extremidade anterior (cabeça do embrião) seja voltada para o lado esquerdo e a extremidade posterior (cauda do embrião) seja voltada para a direita. E ainda, que a região dorsal esteja na parte superior (topo) do embrião e a região ventral na parte inferior (base). Essa premissa garante a padronização dos embriões de *Drosophila*, e consequentemente, colabora para a uniformização dos repositórios de dados.

*Embryostandar* ajusta as propriedades Direção e Orientação com base na análise automática da forma dos embriões de *Drosophila*. Primeiro, o método define os eixos (maior e menor) da elipse ajustado ao embrião e os quatro setores (ilustrado na Fig. 14): AD (anterior-dorsal), AV (anterior-ventral), PD (posterior-dorsal) e PV (posterior-ventral). Em processamento digital de imagens, o Análise de componentes principais (*Principal Component Analysis*) (PCA) é frequentemente usado para determinar o eixo principal (maior), definido como a linha que passa pelo centroide do objeto de interesse e se estende até seus limites anterior e posterior. Entretanto, devido a geometria dos embriões de *Drosophila* (semelhante a uma elipsoide com os lados dorsal e ventral assimétricos (HE et al., 2010)), o eixo principal foi definido como a maior linha horizontal que pode ser traçada entre dois pontos do contorno do embrião – essa linha é assumida como o eixo

AP. De modo semelhante, o eixo secundário (menor) é definido como a linha perpendicular que passa pelo centro do eixo principal do embrião e se estende até as bordas das regiões dorsal e ventral, definindo o eixo DV. A determinação dos eixos (principal e secundário) possibilita identificar os quatros setores (AD, PD, AV, PV) nos embriões, tal como ilustrado na Fig. 14.

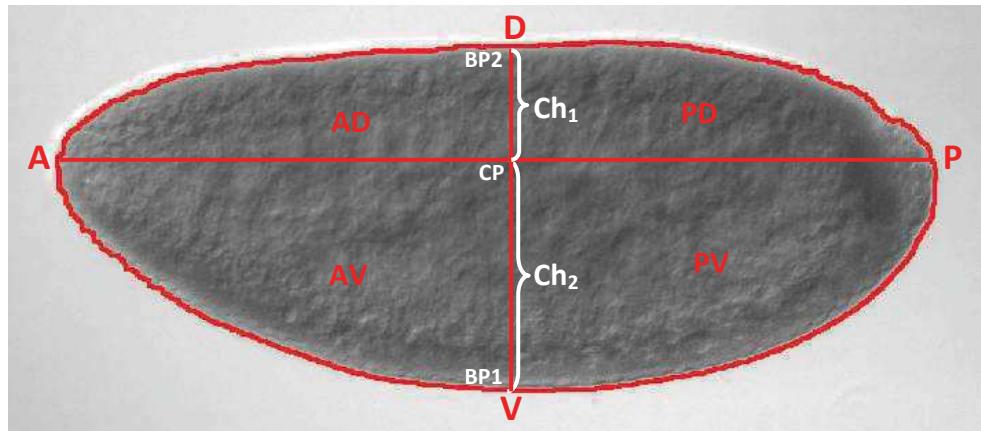


Figura 14 – Eixos e setores em um embrião de *Drosophila*. O eixo AP é definido como a maior linha horizontal que pode ser desenhada entre dois pontos do contorno do embrião (linha horizontal AP). O eixo DV é definido como a linha vertical perpendicular ao eixo AP traçado entre os pontos D e V. A extremidade anterior é definida pelos setores anterior dorsal (AD) e anterior ventral (AV). A extremidade posterior é composta pelos setores posterior dorsal (PD) e posterior ventral (PV).  $CP(x_c, y_c)$  é o ponto central,  $BP_1(x_1, y_1)$  e  $BP_2(x_2, y_2)$  são os dois pontos de borda usados para definir a corda. Neste exemplo  $Ch_1$  é menor que  $Ch_2$ , então  $Ch_1$  define o lado dorsal, ao passo que  $Ch_2$  define o lado ventral. Imagem de embrião obtida da base de dados BDGP (TOMANCAK et al., 2002).

#### 4.4.2 Eixos e setores do embrião

Após a etapa anterior, o próximo passo é caracterizar a posição dos eixos e setores do embrião de *Drosophila*. Observando, por exemplo, se a extremidade anterior e o lado dorsal do embrião coincidem com a posição definida como padrão. Este estágio é realizado com base em duas regras relativas à morfologia dos embriões:

- ❑ O lado dorsal do embrião é menos curvo que o lado ventral.
- ❑ A extremidade anterior do embrião é mais estreita que a extremidade posterior.

Para caracterizar as regiões do embrião (ANTERIOR, POSTERIOR, DORSAL e VENTRAL) o método utiliza duas propriedades: excentricidade ( $E$ ) e corda ( $Ch$ ). A excentricidade é a medida de circularidade da elipse e pode ser calculada pela Eq. 1,

$$E = \sqrt{1 - \frac{b^2}{a^2}}, \quad (1)$$

em que  $a$  determina o semi-eixo maior e  $b$  o semi-eixo menor do contorno da elipse definida sobre a região de interesse. A excentricidade varia entre 0 e 1. Valores mais próximos de 0 (zero) indicam formas mais circulares, ao passo que, valores próximos de 1 (um) supõem formas mais achatadas (WEISSTEIN, 2020b).

A excentricidade é calculada separadamente em cada parte do embrião. Dois objetos são gerados, um com a extremidade anterior do embrião e outro com a extremidade posterior. Então, a excentricidade é calculada em cada uma das partes. O lado com maior valor de excentricidade é caracterizado como lado anterior. A segunda propriedade, a corda ( $Ch$ ), é calculada separadamente nas partes superior e inferior do embrião (Fig. 14), a partir das Eqs. 2 e 3, visando caracterizar as regiões como dorsal e ventral. A corda pode ser descrita como qualquer segmento de linha que se estende de um ponto a outro em uma figura geométrica (WEISSTEIN, 2020a), podendo ser obtida dessa forma:

$$Ch_1 = \sqrt{(x_1 - x_c)^2 + (y_1 - y_c)^2}; \quad (2)$$

$$Ch_2 = \sqrt{(x_2 - x_c)^2 + (y_2 - y_c)^2}. \quad (3)$$

$Ch_1$  e  $Ch_2$  são definidos pela distância entre o ponto central do eixo AP ( $CP(x_c, y_c)$ ) e dois pontos específicos da borda:  $BP_1(x_1, y_1)$  e  $BP_2(x_2, y_2)$ , respectivamente. Esses pontos são definidos pela interseção do eixo secundário com a borda do embrião. Na Fig. 14 são apresentados os locais onde esses pontos são definidos no embrião.

O lado do embrião com maior  $Ch$  é caracterizado como ventral, pois corresponde a parte de maior perímetro entre as duas regiões analisadas. Em contrapartida, o lado com menor corda é caracterizado como dorsal.

#### 4.4.3 Rotação, espelhamento e corte

Após caracterizar os eixos e setores no embrião de *Drosophila*, o método verifica se as propriedades Direção e Orientação precisam ser ajustadas. O objetivo é garantir o alinhamento do embrião à posição definida como padrão, isto é, direção horizontal, lado anterior para a esquerda e região dorsal para o topo (KONIKOFF et al., 2012). Quando for necessário ajustar a direção e orientação do embrião, um esquema de três passos é adotado: (1) rotação; (2) espelhamento anterior-posterior; e (3) espelhamento dorsal-ventral.

No primeiro passo, ilustrado na Fig. 15(a-c), é usada uma função de rotação para alinhar o embrião na direção horizontal, isso requer três parâmetros de entrada: o embrião em análise, o ângulo de rotação e o método de interpolação. O ângulo de rotação é um escalar que especifica o menor ângulo entre o eixo  $x$  (eixo horizontal em um sistema de coordenadas retangulares) e o eixo principal do embrião. O ângulo varia de -90 a 90 graus. O embrião é girado no sentido anti-horário. O método de interpolação usado para

a rotação do embrião é o “vizinho mais próximo”, que usa o valor do pixel mais próximo para atribuir o valor do pixel de saída. Ele foi utilizado por ser um dos métodos de interpolação mais simples e rápidos (HAN, 2013).

A função de rotação apenas alinha o embrião horizontalmente, ela não garante a orientação correta do eixo. Se necessário, uma segunda operação, baseada na excentricidade, é adotada. Isto é, após alinhar o embrião na direção horizontal (AP), os valores de excentricidade nas partes anterior e posterior do embrião são avaliados, e caso não representem a posição padrão (AP) um espelhamento horizontal é aplicado, como mostrado na Fig. 15(d-e). No terceiro passo, a propriedade *corda* é utilizada para verificar a posição dos lados ventral e dorsal. De modo semelhante, quando a orientação vertical do embrião não coincide com a padrão, um espelhamento é realizado, garantindo que o lado dorsal do embrião fique voltado para o topo (Fig. 15(f-h)).

Por fim, após a padronização do embrião de *Drosophila*, um recorte é realizado na imagem. O recorte é feito com base em um retângulo que delimita uma região mínima da imagem, de tal modo que comporta todos os pixels do embrião principal. Um esquema resumido do *Embrystandar* é detalhado a seguir.

1. O embrião é rotacionado para se alinhar ao eixo *x*.
2. A orientação do eixo AP é definida pela propriedade excentricidade (*E*), calculada em cada extremidade do embrião. A metade do embrião com a maior excentricidade define a região anterior do embrião. Se necessário, o embrião é espelhado (invertido) horizontalmente.
3. A orientação do eixo DV é definida pela propriedade *corda*, calculada em cada lado do embrião. Se necessário, o embrião é invertido verticalmente.
4. Recorte da imagem.

## 4.5 Módulo 3 - Segmentação dos núcleos

Na metodologia proposta, a tarefa de segmentação possui dois objetivos: (a) identificação dos embriões de *Drosophila* e (b) detecção dos núcleos celulares. O resultado da tarefa (a) é a máscara do embrião e o da tarefa (b) é a máscara nuclear. Para a tarefa (a) pode-se adotar a solução apresentada no Módulo 1. Por outro lado, tarefa (b) requer técnicas mais sofisticadas, pois envolve uma tarefa mais complexa.

O objetivo da máscara nuclear é obter a forma individual dos núcleos celulares, importante característica fenotípica, frequentemente usada em estudos biológicos sobre sociologia celular e processos intracelulares (MEIJERING, 2012). Essa imagem é primordial para a realização de várias análises biológicas, tais como, quantificação do nível de expressão gênica (PAN et al., 2006), análise da dinâmica de formação dos padrões de expressão

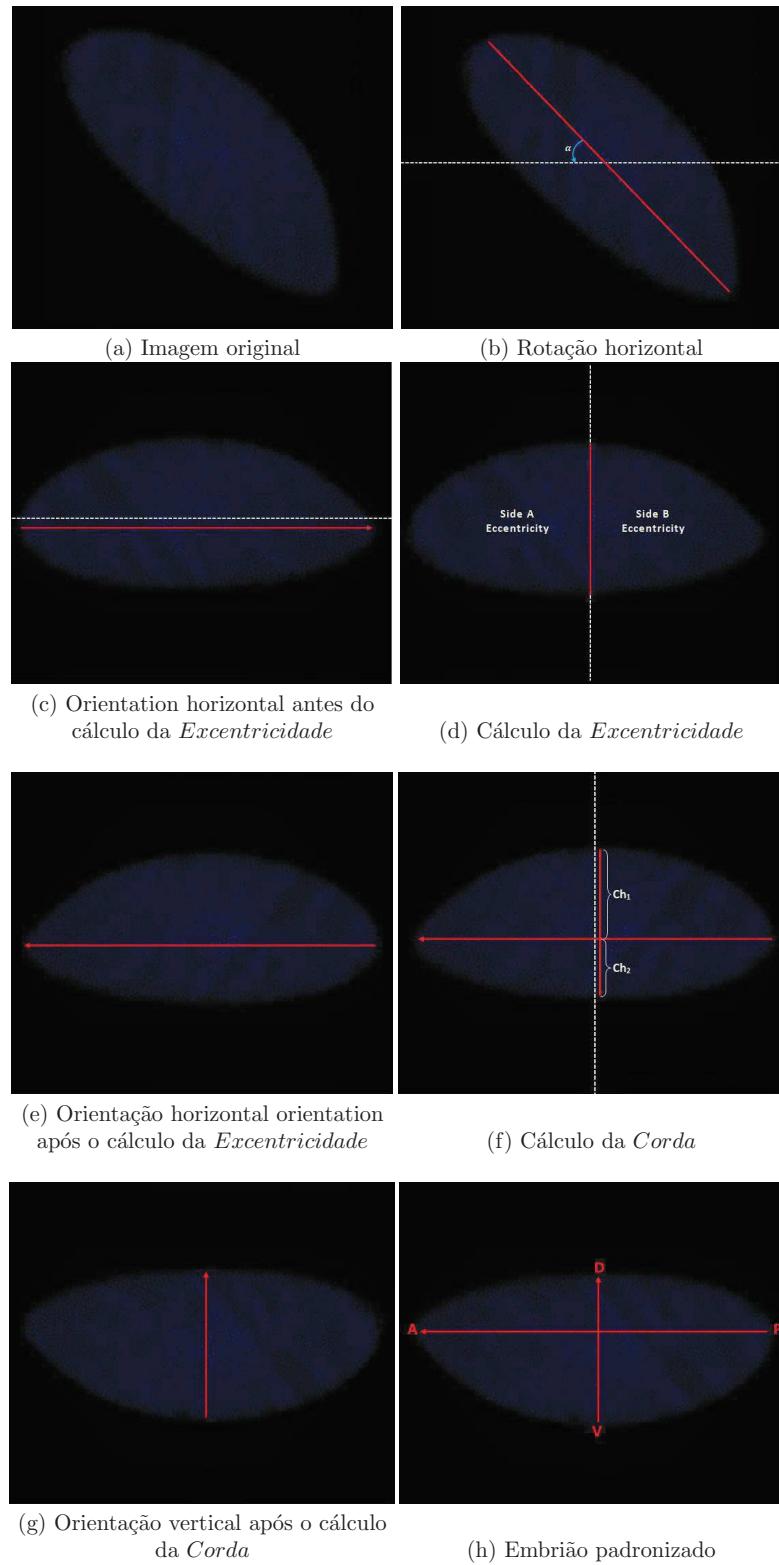


Figura 15 – Esquema de padronização do embrião de *Drosophila*. (a-c) Alinhamento do embrião em relação ao eixo  $x$ . (d-e) Padronização da orientação horizontal do embrião. (f-g) A medida corda é usada para padronizar a orientação vertical. (h) A saída esperada é um embrião com o lado anterior à esquerda e a região dorsal para cima.

gênica (POUSTELNIKOVA et al., 2004), análise das redes de regulação gênica (FOWLKES et al., 2008), comparação dos padrões de expressão gênica (KUMAR et al., 2002), entre outros.

Os núcleos celulares dos embriões de *Drosophila* podem ser identificados a partir das imagens de superfície e das seções sagital e transversa. Como essas imagens são morfológicamente distintas, são propostos dois métodos de segmentação diferentes para a geração da máscara binária nuclear. Para cada caso, são discutidas técnicas de pré-processamento e os métodos de segmentação adotados.

#### 4.5.1 Seções Sagital e Transversa

Antes da segmentação é feito um pré-processamento de imagens confocais obtidas nas seções sagital e transversa. Como descrito na Seção 2.4, nesse tipo de imagem os núcleos são visualizados lateralmente e ocorre maior presença de ruídos nas ilhas citoplasmáticas. Além disso, o baixo contraste entre os núcleos celulares e a iluminação irregular das imagens prejudicam os resultados de etapas posteriores. Sendo assim, o método aborda três passos básicos, sendo: (a) filtragem Gaussiana, (b) realce linear e (c) filtragem morfológica. A escolha dessas técnicas baseou-se em trabalhos anteriores (SOUSA et al., 2013). No próximo parágrafo são descritos os parâmetros usados por essas técnicas. Esses parâmetros foram definidos de forma empírica.

A filtragem Gaussiana corretamente ajustada é capaz de suavizar o ruído, e ainda, conservar os contornos dos núcleos celulares. Usualmente, o tamanho da máscara depende do tamanho dos objetos em estudo. Para o caso em questão, definiu-se um filtro de tamanho  $5 \times 5$ , desvio padrão igual a 2 e valores da máscara determinados pela função Gaussiana discreta (GONZALEZ; WOODS, 2010). São selecionadas para o passo (b) as imagens suavizadas pela filtragem Gaussiana. Esse passo envolve a aplicação da técnica realce linear, que por meio da função de mapeamento linear realça sobretudo a região do citoplasma circundante aos núcleos celulares (FILHO; NETO, 1999). O valor ajustado para o brilho é inicializado com  $b = -10$ . No passo (c) aplica-se a filtragem morfológica, a fim de ajustar o contraste e realçar as regiões de interesse da imagem. Nesse caso, o operador abordado é o *Tophat*, com elemento estruturante no formato “disco” e raio de tamanho 15. Em geral, essa técnica permite destacar as regiões de picos (núcleos) e isolar as regiões de mínimo local (ilhas citoplasmáticas). Casos específicos de imagens sagitais com difícil identificação das formas nucleares – decorrente de maior variabilidade na distribuição de intensidade dos pixels – é aplicada uma função de suavização por spline (GARCIA, 2010b; GARCIA, 2010a) com parâmetro de suavização  $S = 5$ . Em tal caso, um escalar positivo pode ser subtraído da saída da suavização para remover os pixels arbitrários de baixa intensidade que permanecem na região circundante aos núcleos celulares.

Após o pré-processamento, é feita a segmentação da seção sagital e transversa. A abordagem em questão é dividida em dois passos principais, sendo (a) binarização da imagem e (b) pós-processamento (SOUZA et al., 2013). A etapa (a) consiste em segmentar os núcleos celulares identificados da imagem original do embrião utilizando o algoritmo de Otsu. Entretanto, tem-se como resultado uma parcela de núcleos agregados (mesclados), tal como pode ser visto na Fig. 16 pelos núcleos destacados na cor amarelo.

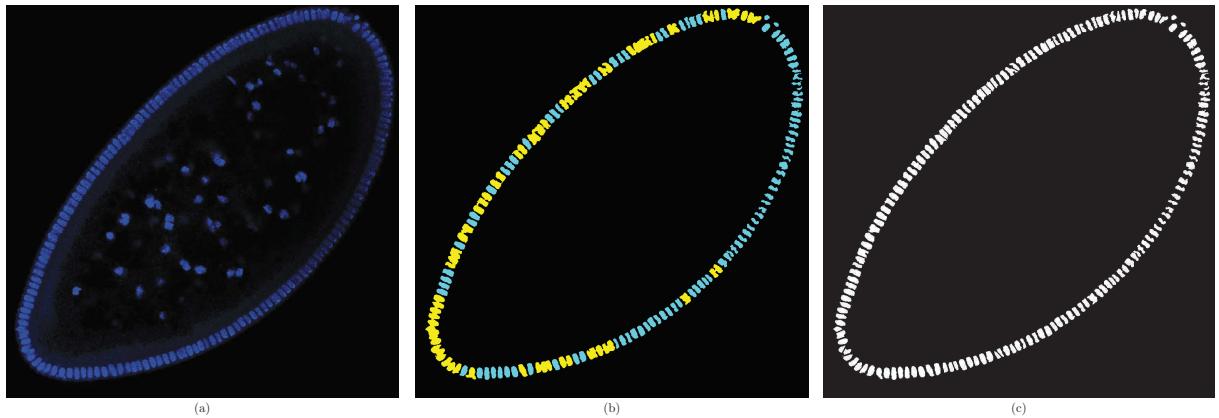


Figura 16 – Resultado qualitativo da aplicação do método de segmentação proposto para imagens das seções sagital e transversa. (a) Embrião de *Drosophila* na seção sagital obtida por microscopia confocal. (b) Resultados da binarização Otsu. Os núcleos mesclados (permanecem agrupados após a binarização) são identificados em amarelo, enquanto os núcleos corretamente segmentados são exibidos na cor azul ciano. (c) Resultado obtido após a aplicação da metodologia proposta.

Visando contornar esse problema, aplica-se um pós-processamento. A ideia principal dessa etapa consiste em estimar a curvatura suavizada para todo componente binário com dois ou mais núcleos mesclados e, a partir disso, detectar pontos extremos de mínimos locais, isto é, regiões candidatas à separação dos núcleos. Para isso, a estratégia baseia-se na análise de curvatura dos objetos binarizados. Resumidamente a abordagem envolve os seguintes passos:

1. Extração do contorno paramétrico;
2. Cálculo da curvatura;
3. Suavização da curvatura obtida;
4. Localização dos pontos extremos de mínimos locais;
5. Definição dos pares de pontos correspondentes;
6. Operação linear para separação dos núcleos agregados.

No primeiro passo obtém-se o contorno paramétrico para cada componente conexo de núcleos agregados, de maneira parametrizada ou ordenada, na forma  $C(t) = (x(t), y(t))$ . A estimativa da curvatura suavizada é realizada para todo componente conexo com dois ou mais núcleos agregados. Após isso, busca-se pelos pontos de mínimos locais (ou mínimos relativos), os quais caracterizaram-se por determinar posições candidatas à divisões dos componentes agregados. Na Fig. 17 são ilustrados seis desses pontos ( $P_1$  à  $P_6$ ). A ideia consiste em traçar segmentos de linha entre pares de mínimos locais correspondentes, formando assim fronteiras de divisão entre os núcleos mesclados. A correspondência entre pontos de mínimos opostos é definido com base na análise do ângulo formado entre (a) *vetor normal* do ponto atual (por exemplo,  $\vec{V}_n$  na Fig. 17) em relação aos (b) *vetores entre pontos* (por exemplo,  $\vec{V}_{p_i}$ ), os quais são definidos considerando a diferença entre o ponto atual e os pontos opostos. O menor ângulo  $\theta_i$  formado entre os vetores (a) e (b), de cada ponto contido no lado escolhido, define os pares correspondentes.

Por fim, após estabelecer a correspondência entre os mínimos relativos procede-se a separação dos componentes agregados. Os pares de pontos correspondentes são unidos através de segmentos de retas traçados na imagem. Isto é, os pixels sobre o segmento são ajustados para zero, gerando uma fronteira de divisão entre os núcleos. Esse procedimento é repetido para todos os pontos localizados no mesmo lado do primeiro mínimo analisado.

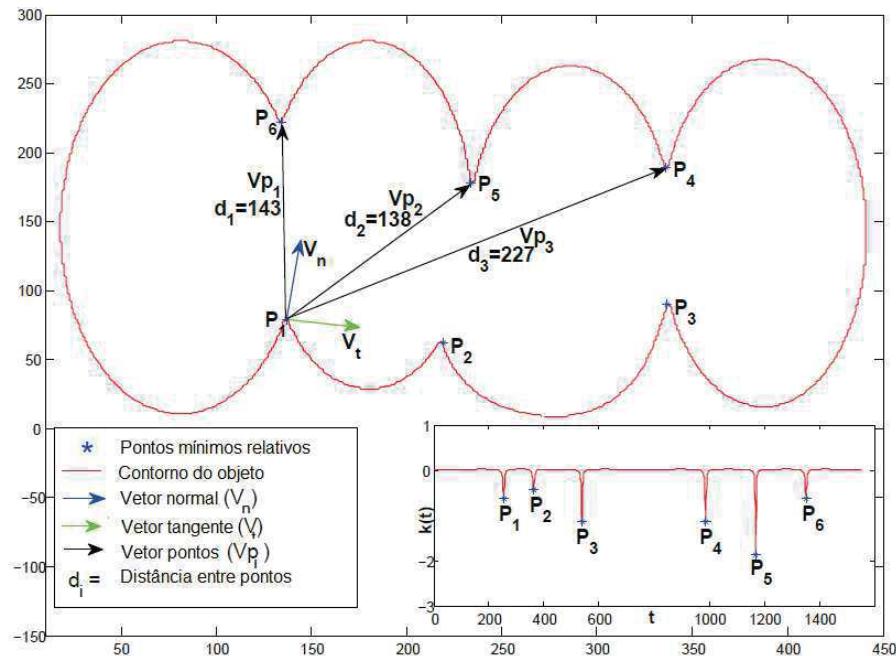


Figura 17 – Correspondência entre os pontos de mínimos locais. Os pontos  $P_1$  à  $P_6$  representam os mínimos locais da curvatura  $k(t)$ . Os valores  $d_1$  à  $d_3$  definem a distância entre os pontos opostos. O gráfico na parte inferior à direita exibe a função de curvatura  $k(t)$  do contorno analisado, em que são destacados os respectivos mínimos locais. Aqui  $P_6$  é identificado como o ponto correspondente à  $P_1$ , tendo em vista que o menor  $\theta_i$  encontra-se entre o vetor  $\vec{V}_n$  do ponto  $P_1$  e o  $\vec{V}_{p_1}$ . Figura obtida de (Sousa et al., 2013).

Na Fig. 16 é ilustrado um resultado da aplicação do método de segmentação descrito. A primeira coluna exemplifica imagens de embriões relativas à seção sagital. A segunda coluna enfatiza o resultado da primeira etapa do método, isto é, binarização Otsu. E por fim, a terceira coluna destaca resultados obtidos após a aplicação da metodologia proposta. A precisão da segmentação nuclear é avaliada a partir de um conjunto de imagens de referências marcado manualmente. Maiores detalhes sobre o módulo em questão pode ser encontrado em (SOUZA et al., 2013).

#### 4.5.2 Imagens da Superfície

Imagens de embriões da *Drosophila* obtidas da superfície são mais comuns de serem encontradas na literatura. O pré-processamento desenvolvido para esse tipo de imagem baseia-se na aplicação de técnicas tradicionais da literatura de PDI. A sequência proposta aborda quatro passos básicos, sendo que os três primeiros (a–c) são os mesmos passos descritos no caso de imagens da seção sagital (Sec. 4.5.1), e um passo final (d) em que é aplicada uma filtragem por mediana. Esse último passo adicional torna-se necessário no plano frontal considerando que a ocorrência de pixels aleatórios de alta frequência (aqui considerados como ruído) dentro dos núcleos celulares e nas regiões do citoplasma é maior nesse tipo de imagem. Esse ruído caracteriza-se por um padrão ou textura granulada aleatória, de modo que impede a visualização de pequenos detalhes na imagem, e ainda prejudica a qualidade dos resultados de etapas posteriores (e.g., segmentação). A filtragem por mediana foi a estratégia para suavizar as regiões que permanecem ruidosas após os passos anteriores e, ao mesmo tempo, preservar as bordas dos núcleos celulares. A aplicação desse conjunto de técnicas torna as imagens de embriões mais apropriadas para as etapas posteriores.

O método de segmentação usado para a imagem de superfície é voltado exclusivamente para imagens confocais (devido à resolução e marcação, as imagens ópticas usadas neste trabalho não permitem identificar os núcleos separadamente). Essas imagens exibem os núcleos celulares distribuídos ao longo do embrião e concentrados em uma região elíptica. As características geométricas e anatômicas dos núcleos são relativamente parecidas, uma vez que eles são capturados da superfície do embrião. Basicamente, a abordagem usada baseia-se na proposta de (POUSTELNIKOVA et al., 2004) que utiliza as estratégias (a) segmentação por região e (b) detecção de bordas. Ambas são aplicadas no resultado de um pré-processamento descrito no parágrafo anterior.

A sequência adota inicialmente a técnica de segmentação *Watershed* controlada por marcadores. Outras técnicas também foram testadas. No entanto, no caso de imagens nucleares essa técnica é a mais eficiente para contornar as configurações de complexos aglomerados (clusters) binários normalmente gerados pelo processo de segmentação (SKOBEL et al., 2019). Como o operador *Watershed* detecta picos de alta intensidade, e como os núcleos nas imagens apresentam baixa intensidade, primeiramente gera-se o negativo da

imagem pré-processada. O próximo passo consiste basicamente em duas etapas, (i) selecionar os marcadores para os objetos de interesse e (ii) modificar o inverso da imagem original a partir da imposição dos marcadores. Em ambos os casos são utilizadas as funções *imextendedmin* e *imimposemin* do MATLAB. A primeira função detecta na imagem de entrada os vales de intensidade mais profundos que um limiar global (obtido a partir método de Otsu), enquanto a função *imimposemin* modifica a imagem para conter apenas os vales encontrados pela função *imextendedmin*. Em suma, elas buscam, respectivamente, criar vales maiores entre os núcleos e impor os mínimos locais do operador *Watershed* na região dos marcadores. Essa sequência, apesar de apropriada, não mostrou-se suficiente para detectar individualmente todos os núcleos celulares dispostos no embrião de *Drosophila*.

Posto isso, o método aborda um passo final (b), isto é, detecção dos contornos nucleares. O objetivo aqui é aproveitar o resultado do *Watershed* (passo (a)) para gerar uma imagem em que os núcleos celulares sejam mais definidos (destacados), e com isso, detectados de forma eficaz. O procedimento utilizado para enfatizar as regiões dos núcleos foi multiplicar a imagem do embrião previamente tratada (pelo módulo Pré-processamento) pela saída da técnica *Watershed*. O resultado do *Watershed* é antes erodido pelo operador morfológico Erosão (em que é utilizado um Elemento estruturante (ES) do tipo “quadrado” e dimensão  $3 \times 3$ ). Consecutivamente, a saída dessa multiplicação é então suavizada utilizando a filtragem por mediana (de dimensão  $5 \times 5$ ), a fim de eliminar detalhes espúrios remanescentes. Enfim, o passo (b) é realizado a partir da aplicação do filtro de bordas Shen-Castan. Esse filtro aplicado com os parâmetros ( $A1=0.5$ ,  $A2=0.7$ ,  $ME_L = 1$ ,  $MA_L = 12$ ,  $JAN=11$ ,  $SEG=10$ ) mostrou-se capaz de encontrar, com maior precisão, os pixels de borda relativos aos núcleos celulares. Maiores detalhes sobre o detector de bordas Shen-Castan, bem como, sobre os respectivos parâmetros, podem ser encontrados em (SHEN J.; CASTAN, 1992).

## 4.6 Módulo 4 - Representação dos dados nucleares

A representação mais comum de imagens 2D de embriões da *Drosophila* é a sua forma original, isto é, o embrião organizado em uma configuração elíptica ao longo dos eixos antero-posterior e dorso-ventral. Esse é o caso principalmente de imagens obtidas da superfície dos embriões. Em geral, as propostas encontradas buscam padronizar as imagens de embriões da *Drosophila* quanto ao alinhamento ou tamanho dos embriões (KUMAR et al., 2002), conforme descrito na Seção 4.4. No entanto, muitas vezes essa tarefa não é suficiente para que as análises biológicas subsequentes sejam realizadas de forma rápida e eficiente, haja visto o número de imagens atualmente manipuladas pelos bancos de dados disponíveis (JI et al., 2008). Portanto, torna-se importante o desenvolvimento de soluções computacionais capazes de otimizar a análise visual e computacional desses dados.

### 4.6.1 Visão linear dos núcleos celulares

O módulo em questão propõe uma nova forma de representação e visualização dos núcleos celulares contidos em imagens de embriões da *Drosophila* coletados das seções sagital e transverso. Trata-se da disposição dos núcleos celulares de uma forma linear, criando faixas horizontais. Para isso, foi utilizado o método inicialmente apresentado em (SOUZA; LOPES; TRAVENÇOLO, 2014), que é capaz de organizar os núcleos lado a lado, ao longo da direção horizontal, tal como se o embrião fosse “aberto”. Essa abordagem fornece uma maneira simples e eficiente de visualização e extração de características dos padrões de expressão gênica. Isto porque, na nova configuração, cada imagem é representada em uma linha, facilitando a identificação de padrões sobrepostos ou similares. A abordagem proposta também facilita a comparação visual de vários embriões com diferentes padrões de expressão nos estágios de desenvolvimento, sendo muito útil para a avaliação das interações entre genes. As etapas principais do módulo são descritas nos próximos parágrafos.

#### 4.6.1.1 Ordem nuclear

A ordem dos núcleos celulares na nova configuração linear deve ser a mesma estabelecida na imagem original do embrião. Assim, o primeiro passo é identificar a ordem em que cada componente aparece na imagem segmentada. Esta tarefa é executada por meio da rotulação de componentes conexos, de forma que cada núcleo na imagem segmentada receba um rótulo. A função rotula os componentes a partir das colunas da imagem. Qualquer objeto que apareça primeiro com base em uma coluna, da ordem de varredura de cima para baixo, recebe o rótulo 1 (um); e assim consecutivamente. Dessa forma, o algoritmo de rotulação não é capaz de exibir uma imagem em que a rotulação dos núcleos siga uma ordem sequencial de vizinhança entre eles.

#### 4.6.1.2 Vizinhança entre os núcleos

Para corrigir a questão anterior, primeiramente procura-se a vizinhança original de cada núcleo a fim de posteriormente atribuir diferentes rótulos. Para realizar essa tarefa, utilizou-se uma combinação das técnicas diagramas de Voronoi e triangulação de Delaunay (AMENTA; BERN; KAMVYSELIS, 1998). O centroide de cada núcleo foi passado como entrada para a triangulação de Delaunay que, a partir dessas informações, gera a relação de vizinhança entre os pontos. A combinação foi baseada na aplicação de um filtro às bordas geradas pela Triangulação de Delaunay. A filtragem permitiu isolar a superfície externa da combinação. A partir disso, tornou-se possível inferir a vizinhança de cada núcleo contido nos embriões de *Drosophila*.

#### 4.6.1.3 Rotulando os núcleos celulares

A saída do passo anterior é uma matriz em que cada linha representa o rótulo de um núcleo celular e seu respectivo par de núcleos vizinhos. Usando esse resultado, cada núcleo pode ser novamente rotulado, obedecendo agora uma ordem sequencial, de forma que núcleos vizinhos recebam rótulos sequenciais.

#### 4.6.1.4 Definindo o ponto central

Para padronizar o alinhamento e distribuição dos núcleos celulares, é necessário definir um ponto central (CP - *central point*). Isto é, o primeiro núcleo a partir do qual o reposicionamento dos componentes é realizado e o embrião é “aberto”. Para isso, foi encontrada a maior linha horizontal que pode ser desenhada entre dois pontos do contorno do embrião. Essa linha se ajusta ao longo do eixo AP dos embriões. E a partir disso, determinou-se como Ponto Central (*Central Point*) (CP) o núcleo mais próximo do seu início.

Essa abordagem é bem empregada no caso de imagens de embriões de *Drosophila* relativas à seção sagital (SOUZA; LOPES; TRAVENÇOLO, 2014). Porém, ela não apresenta um comportamento robusto em imagens obtidas da seção transversa. Isso porque as configurações geométricas dos embriões diferenciam-se conforme o plano de captura da imagem.

Portanto, neste trabalho é proposto um algoritmo para a definição do ponto central em imagens relativas à seção transversa dos embriões de *Drosophila*. Em tal caso, a estratégia baseia-se na detecção da simetria da expressão gênica (Fig. 18). A ideia consiste em encontrar o objeto (núcleo) da imagem que divide o embrião em duas partes de tal forma que essas partes possuam um perfil de expressão semelhante, caracterizando assim um ponto de alta simetria. Como é possível encontrar mais de um núcleo com esta propriedade, a busca é realizada considerando apenas os núcleos com maior intensidade de expressão gênica. Para calcular a simetria de reflexão para um núcleo  $i$ , os valores de expressão gênica dos núcleos são organizados sequencialmente ao longo de um vetor ( $CV_i$ ) começando no núcleo  $i$ . A partir de cada posição, o vetor  $CV_i$  é subtraído do vetor  $CV_{rev_i}$ . Este último é o próprio vetor  $CV_i$ , porém organizado em uma ordem reversa. A operação de subtração gera o vetor diferença ( $DV_i$ ). A simetria  $S_i$  de um núcleo é definida como a soma absoluta de  $DV_i$ . Quanto menor  $S_i$ , maior a simetria, sendo 0 o maior valor de simetria. As etapas para esta estratégia estão resumidas abaixo:

- ❑ Passo 1 - Encontrar os  $P$  núcleos com os maiores valores de intensidade da expressão gênica (aqui  $P = 15$ ).
- ❑ Passo 2 - Calcular a simetria de reflexão  $S_i$  para os núcleos encontrados no Passo 1.
- ❑ Passo 3 - Selecionar os  $D$  núcleos mais simétricos (aqui  $D = 5$ ).

- ❑ Passo 4 - Definir o ponto central (CP – *central point*) como o núcleo com maior nível de expressão gênica entre os núcleos selecionados no Passo 3.

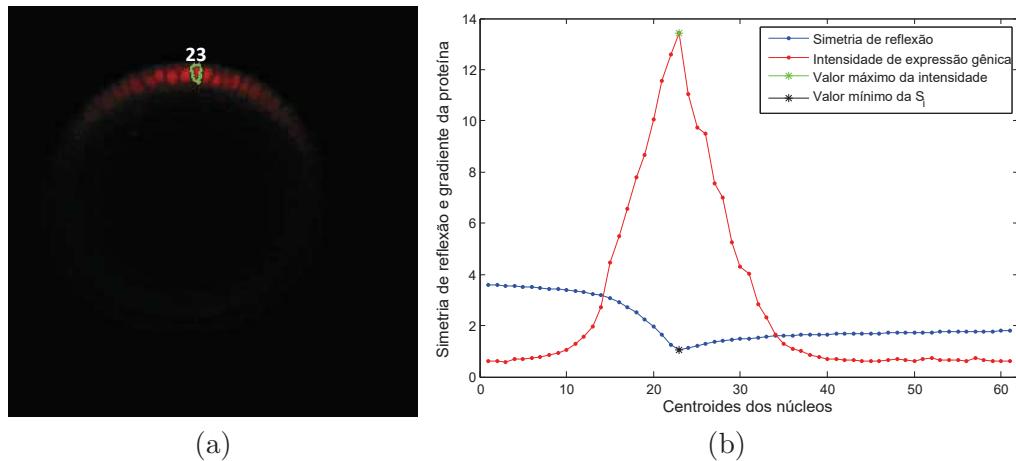


Figura 18 – Procedimento realizado para definir o ponto central (CP) com base na simetria de reflexão. (a) Embrião de *Drosophila* na seção transversa exibindo expressão gênica da proteína Dorsal. (b) Gráfico com as curvas simetria de reflexão (azul) e intensidade de expressão gênica da proteína (vermelha). Quanto menor o valor  $S_i$ , maior é a simetria entre as partes resultantes. Nesse exemplo encontrou-se o CP igual a 23, i. e., núcleo do embrião com maior valor de expressão gênica e localizado na região do embrião com menor valor de  $S_i$ , (aqui, 0.85).

#### 4.6.1.5 Repositionamento dos núcleos

Finalmente, os núcleos são dispostos ao longo de uma faixa horizontal. A distância entre eles deve ser a mesma que na imagem original. Para realizar essa tarefa alguns passos são realizados, tal como resumidos abaixo:

- ❑ Passo 1 - Identificar cada componente/núcleo ( $C_i$ ) do embrião e seus respectivos vizinhos, considerando  $0 < x \leq n$  em que  $n$  é o total de núcleos do embrião.
- ❑ Passo 2 - Determinar o ângulo inicial ( $\alpha$ ) que cada  $C_i$  apresenta em relação ao eixo-X (eixo horizontal em um sistema de coordenadas retangulares). Para encontrar o  $\alpha$ , foi utilizado a função *regionprops.orientation* do MATLAB. Ela especifica um escalar em graus que varia de  $-90^\circ$  a  $90^\circ$ , de tal modo que corresponde ao ângulo existente entre o eixo-X e o eixo maior da elipse que se encontra centrada em cada  $C_i$ .
- ❑ Passo 3 - Determinar o ângulo de rotação para cada  $C_i$  do embrião. A partir do  $\alpha$  (encontrado no passo anterior) determinou-se o ângulo ( $\phi$ ). Esse parâmetro é obtido a partir das Eqs. 4 e 5, e determina o valor de entrada para a função rotação.

$$\text{se } \alpha < 0, \phi = \phi + 90 \quad (4)$$

$$\text{se } \alpha > 0, \phi = \phi - 90 \quad (5)$$

- Passo 4 - Rotacionar cada componente do embrião para a direção vertical, a fim de posteriormente reposicioná-los na nova faixa linear. Em tal passo, utilizou-se a função de rotação *imrotate*. Ela gira cada componente no sentido anti-horário em torno de seu ponto central com base em três parâmetros de entrada: o componente  $C_i$  em análise, o ângulo de rotação ( $\phi$ ) e o método de interpolação (aqui foi utilizado o vizinho mais próximo).
- Passo 5 - Calcular o valor da distância existente entre cada componente  $C_i$  do embrião e seus respectivos vizinhos. Para isso foi usada a distância entre cada par de observações (com base no centroide dos respectivos componentes).
- Passo 6 - No último passo, cada componente  $C_i$  do embrião é reposicionado na nova matriz horizontal. A matriz é preenchida com base nos dados de entrada: rótulo, centroide, tamanho, coordenadas e valores dos pixels de cada componente  $C_i$  do embrião. O tamanho da nova matriz é definido em tempo de execução, conforme os dados dos componentes avaliados.

Essa forma de visualização é compacta e permite exibir dezenas de embriões ao mesmo tempo. Com isso, a análise visual do pesquisador é facilitada, e pode-se comparar mais rapidamente os padrões de expressão de diferentes embriões. Além disso, torna-se mais fácil observar a sobreposição de núcleos celulares usando essa técnica de abertura.

## 4.7 Módulo 5 - Representação dos dados do cito-plasma

As imagens de expressão gênica obtidas e estudadas nesta tese são relativas ao ciclo 14 do desenvolvimento embrionário da *Drosophila melanogaster* (Seção 2.2). Nessa fase, ocorre a celularização, isto é, formação de invaginações membranares ao redor de cada núcleo. Esse estágio embrionário, conhecido como blastoderme celular, é caracterizado pela maioria dos núcleos localizados na periferia do embrião. Morfologicamente, os núcleos são iguais e distribuídos isoladamente em uma monocamada, mas a nível molecular são diferentes. Quando o núcleo alcança a periferia do embrião, cada um fica imerso em uma respectiva região do citoplasma, conhecida como “ilha” do citoplasma. Cada núcleo é associado a um domínio de citoplasma estruturado (MAZUMDAR; MAZUMDAR, 2002; GILBERT, 2003).

Em imagens de embriões obtidas da superfície dos embriões, a ilha do citoplasma compreende os sulcos (vales) que se arranjam entre os núcleos celulares. Visualmente esse espaço é bem pequeno. O núcleo e suas ilhas citoplasmáticas associadas são chamados enérgides (GILBERT, 2003). Essa região não é comumente avaliada em estudos de Bioinformática. Por outro lado, no caso de imagens de embriões obtidas das seções sagital e transversa, a ilha do citoplasma é bem delimitada e geralmente discutida em estudos sobre a *Drosophila*. Nessas imagens, o citoplasma é dividido entre as partes (i) basal e (ii) apical. Nessa direção, neste módulo uma nova forma representação e visualização do citoplasma é apresentada.

A abordagem proposta para a representação e visualização do citoplasma (contido em imagens de embriões da *Drosophila*) segue a metodologia descrita na Seção 4.6. Isto porque, o objetivo principal de ambos os métodos coincidem, ou seja, dispor as regiões de interesse (neste caso, citoplasma apical e basal) em uma nova configuração linear. Sendo assim, cada passo descrito na Seção 4.6 é reutilizado aqui, com a diferença que, neste caso, os componentes de interesse são as regiões do citoplasma (basal ou apical) e não os núcleos celulares. Na Fig. 19 é ilustrado o resultado do esquema de abertura das faixas dos citoplasmas apical e basal em uma imagem relativa à seção transversa.

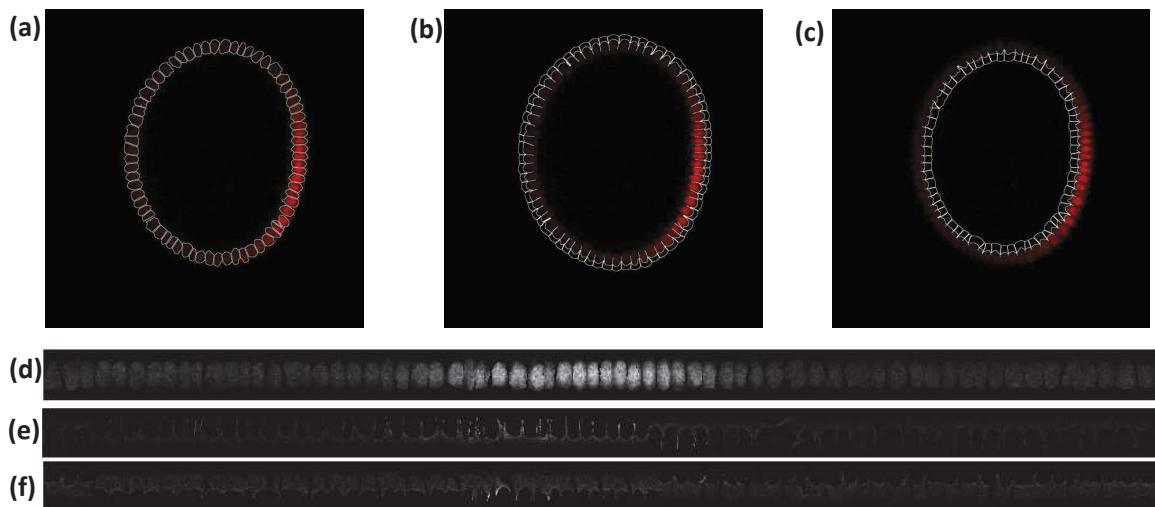


Figura 19 – Abertura das faixas nuclear e citoplasmática em imagem de embrião na seção transversa. (a) Sobreposição do contorno da máscara nuclear sobre a imagem com marcação de proteína. (b) e (c) Sobreposição do contorno dos citoplasmas apical e basal, respectivamente, sobre a imagem com marcação de proteína. Cada região definida pelo contorno branco identifica um componente para a faixa horizontal. (d, e, f) Exibem as respectivas regiões após aplicação do método de abertura do embrião.

## 4.8 Módulo 6 - Quantificação dos dados da expressão

Após obter a máscara de segmentação nuclear é possível avaliar a expressão gênica dos embriões de *Drosophila*. Para isso, incluem-se duas abordagens, uma para quantificação e outra para visualização da intensidade média (de proteína ou RNA) contida nas regiões intranuclear e citoplasmática dos embriões. O detalhamento das abordagens descritas segue nas próximas seções.

### 4.8.1 Quantificação da expressão nos núcleos

A quantificação da região nuclear é realizada para dois tipos principais de imagens de embriões, obtidas (a) da superfície e (b) das seções sagital (AP) e transversa (DV). No caso das imagens relativas à superfície dos embriões, o método inicia-se com a sobreposição da (i) máscara de segmentação nuclear com a (ii) imagem original com marcação do RNA ou proteína. A partir disso, torna-se possível quantificar a intensidade média dos sinais expressos na região intranuclear. Essa operação é realizada multiplicando, ponto a ponto, as imagens (i) e (ii) e, seguidamente, efetuando-se o cálculo da média dos pixels pertencentes a cada núcleo celular. Cada núcleo da máscara de segmentação é definido em termos de matrizes por um rótulo. As coordenadas ( $x, y$ ) do centroide de cada componente nuclear tornam-se marcadores para o mapa de quantificação, isto é, uma tabela contendo dados gerais sobre o processo de quantificação. Para cada núcleo lido obtém-se o rótulo, os valores de coordenadas do centroide e o valor médio da intensidade fluorescente (relativa aos níveis de expressão) para cada gene e cada classe de tempo marcada no embrião (JANSSENS et al., 2005). Na Fig. 20 esse processo é ilustrado.

### 4.8.2 Quantificação nas faixas Central e Mediatrix

Em vários trabalhos a quantificação dos dados de expressão refere-se a dados da região central de cada embrião. A faixa central é amplamente utilizada na literatura para extração de dados quantitativos relativos à superfície dos embriões (JANSSENS et al., 2005; PISAREV et al., 2009; OCHOA-ESPINOSA et al., 2009). Para isso, os valores médios de fluorescência considerados se limitaram à uma faixa horizontal que se estende ao longo do centro dos embriões 10% (AP). Esta região é normalmente descrita como faixa central e é amplamente utilizada para extração de dados quantitativos relativos à superfície dos embriões (JANSSENS et al., 2005; PISAREV et al., 2009). A partir dela, somente as médias de intensidades nucleares com posição do centroide dentro da faixa descritas (Fig. 21) são consideradas.

Neste trabalho é proposta a análise quantitativa dos dados obtidos das faixas central e mediatrix de imagens da superfície dos embriões. Observações preliminares indicam que dependendo da região de quantificação dos dados há diferença nas posições medidas para

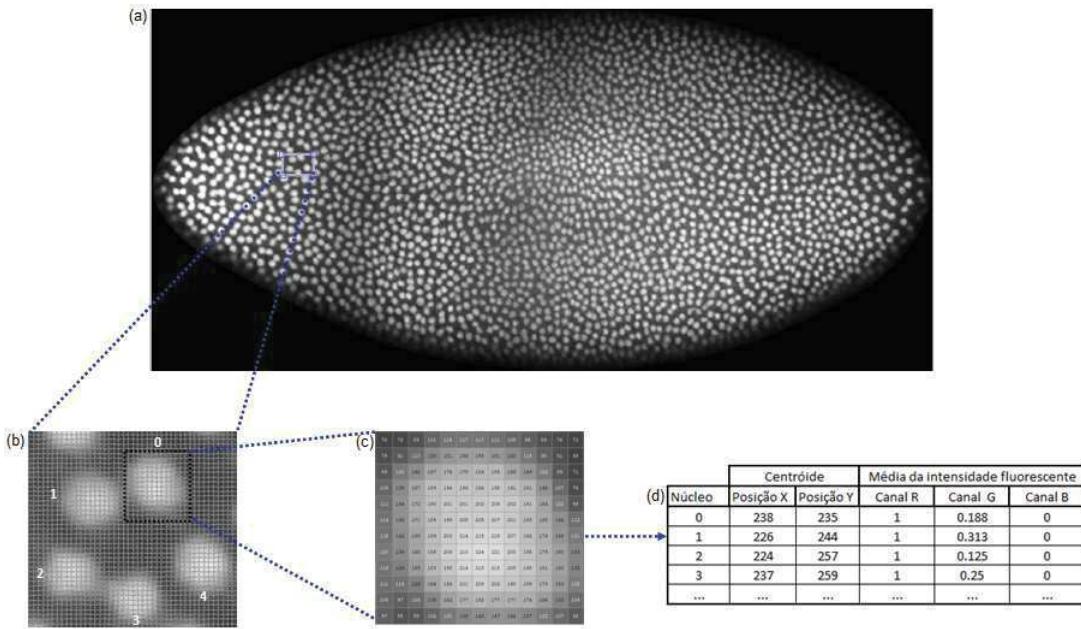


Figura 20 – Processo de quantificação da região nuclear em imagem de embrião do plano frontal. No esquema é ilustrado o cálculo da intensidade média de proteína por núcleo. Em (a) é exibido uma imagem de embrião de *Drosophila* no plano frontal. (b) Ampliação da região delimitada em (a). Os números 0...4 identificam os núcleos da região selecionada em (a). (c) Destaca o núcleo de número 0 para ilustração da quantificação. A resposta de saída (d) é uma tabela contendo a ID de cada núcleo, as coordenadas X e Y dos centroides respectivos, e os valores de intensidade expressos em cada canal (se houver mais de um). Disso, obtém-se a média da intensidade de fluorescência para cada núcleo segmentado.

os padrões. Diante disso, o problema de pesquisa consiste em verificar qual é a influência que a geometria dos embriões de *Drosophila* exerce sobre os resultados biológicos. Trata-se especificadamente da tarefa de comparar os sinais de expressão obtidos da faixa central em relação à faixa mediatrix. Essa abordagem permite resolver questões biológicas, tais como: o deslocamento de padrões de acordo com a idade dos embriões ocorre de fato? Ou somente é fruto da forma como é medido o sinal na faixa central? Em geral, mostrar que há uma influência da geometria dos embriões sobre resultados biológicos seria interessante para reforçar a importância da coleta de embriões na região sagital, em que esse efeito de geometria não existe ou pode ser minimizado.

Inicialmente definiu-se a faixa central como a maior linha horizontal que pode ser desenhada entre dois pontos da linha de contorno do embrião (linha amarela na Fig. 22). E a partir disso, determinou-se a localização da faixa Mediatrix. Essa faixa é inserida logo abaixo da faixa central, e igualmente usa dados de 10% da largura do embrião (eixo DV) e se estende ao longo do eixo AP (Fig.23).

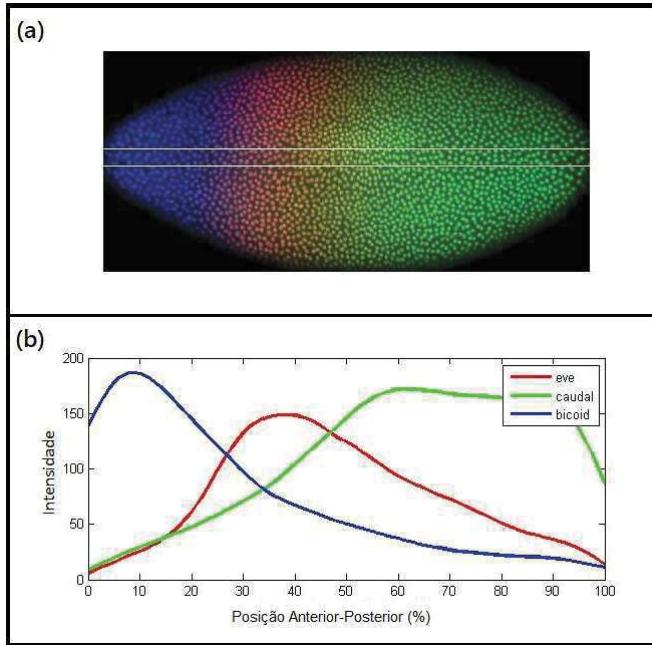


Figura 21 – Processo de quantificação em uma imagem arbitrária da base FlyEx. (a) Embrião de *Drosophila* da classe de tempo 8 e ciclo 14A com marcação dos genes *Bicoid* (azul), *Caudal* (verde) e *Even-skipped* (vermelho). As linhas brancas horizontais ilustram a posição da faixa central que delimita a região de interesse (10% do eixo DV estendendo-se ao longo do eixo AP). (B) Padrões de expressão unidimensionais dos três genes citados.

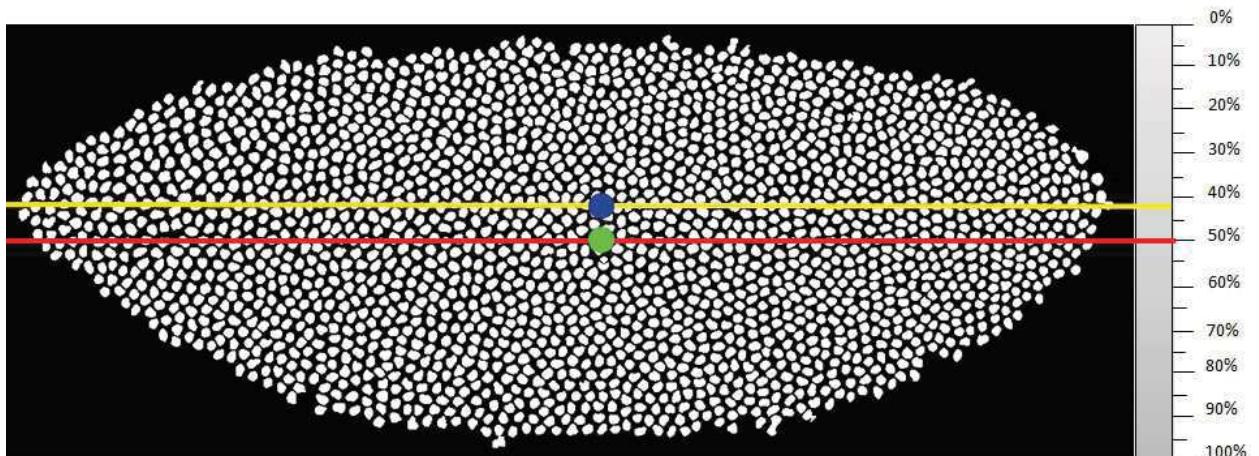


Figura 22 – Representação gráfica da faixa central. Em vermelho, a faixa central desenhada a partir do centroide (ponto verde). E em amarelo, a faixa central desenhada a partir da maior linha horizontal. O ponto azul define o centro do embrião.

#### 4.8.3 Quantificação das Seções Sagital e Transversa

O método de quantificação relativo às seções sagital e transversa dos embriões é semelhante ao plano frontal. A máscara de segmentação nuclear é superposta à imagem

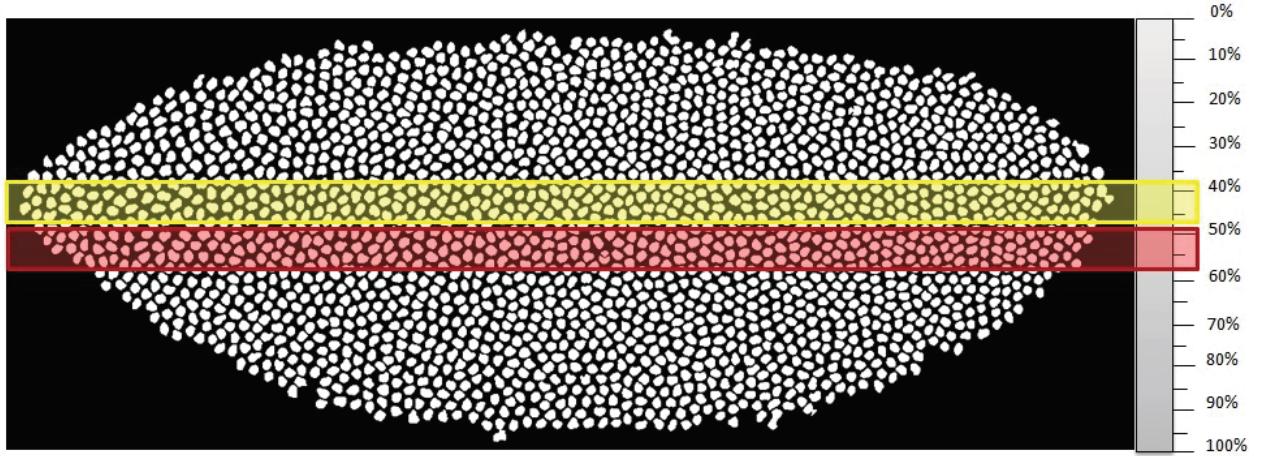


Figura 23 – Comparação gráfica das faixas central e mediatriz. Em amarelo, a faixa central desenhada a partir da maior linha horizontal. E em vermelho, a faixa mediatriz inserida abaixo da faixa central.

original com o sinal de expressão do gene. A diferença é que não existe a delimitação da faixa central do embrião. Assim, no caso dessas imagens a tabela contém os níveis médio de intensidade fluorescente de todos os núcleos celulares do embrião. Na Fig. 24 é ilustrado o esquema descrito. A resposta de saída é a representação matricial para cada núcleo, a partir da qual obtém-se a média da intensidade de fluorescência expressa por núcleo. A avaliação da expressão é então feita observando a intensidade ao longo do eixo AP para seções sagitais e referente ao eixo DV para as seções transversas, seguida uma ordem semelhante a visão linear dos embriões discutidas nas seções anteriores.

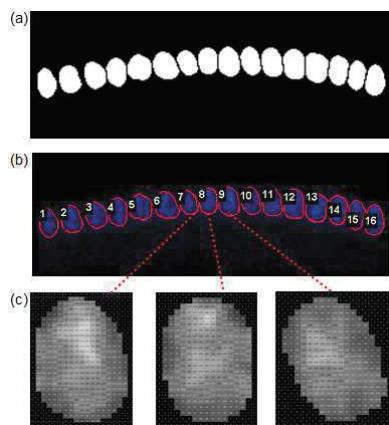


Figura 24 – Processo de quantificação da região nuclear em imagem de embrião na seção sagital. Em (a) é exibido um pequeno trecho da máscara de segmentação nuclear relativa a uma seção sagital. Em (b) a combinação entre (i) o contorno da máscara de segmentação nuclear exibida em (a) com o respectivo trecho da imagem do embrião com marcação DAPI (ii). Na imagem (c) é exemplificada a saída matricial de três núcleos.

#### 4.8.4 Quantificação do citoplasma

O método de quantificação da região citoplasmática é proposto para imagens capturadas da seção sagital ou transversa, relativas aos eixos AP e DV. A análise quantitativa do eixo DV é menos extensiva na literatura (KANODIA et al., 2011). No entanto, visando a proposição de uma abordagem mais completa também é proposto uma rotina para esse tipo de imagem.

De modo geral, o módulo aqui apresentado possui a mesma estratégia do módulo de quantificação da região nuclear, isto é, sobreposição de duas imagens, sendo (i) máscara de segmentação e (ii) marcação dos sinais de expressão, a fim de realizar a multiplicação dos pixels que se interseccionam na região de interesse. Neste caso, (i) é a máscara binária da região citoplasmática e (ii) é a imagem dos sinais de RNA ou proteína expressos no citoplasma. A principal diferença aqui é a localização citoplasma. Ele se localiza no entorno dos núcleos celulares e, para quantificá-lo inicialmente é necessário determinar o tamanho e a matriz de pixels para cada um de seus componentes.

Em suma, o método baseia-se em três etapas: (a) definir a máscara nuclear, (b) estender a região dos núcleos nas direções basal e apical e (c) quantificação. A primeira etapa (a) é obtida a partir do módulo de segmentação (definido na Seção 4.5.1). Após isso, aplica-se na máscara resultante a operação transformada distância com base na métrica Euclideana. O resultado é uma outra imagem em que o valor de cada pixel representa sua distância Euclidiana à borda do objeto. Como saída, também é obtido o mapa de distâncias na forma de uma matriz de índices (*idx*). Cada elemento de *idx* contém o índice do pixel diferente de zero mais próximo da imagem binária. Esse resultado estende a região dos núcleos celulares nas direções basal e apical com a propagação individual de cada rótulo. Isto é, cada componente da transformada distância (Fig. 25) é associado a um respectivo núcleo celular.

Esse resultado comprehende o mapa de distâncias completo, mas não foca a região do citoplasma. Assim, visando delimitar as regiões de interesse aplica-se um filtro de condição, o qual desconsidera os pixels localizados a uma distância superior a *d*. Na Fig. 26 são ilustrados as regiões delimitadas para diferentes valores de *d*. Neste trabalho, *d* é pré-estabelecido como o valor do tamanho médio dos núcleos celulares (*med*). Entretanto, o método é flexível e o usuário pode alterá-lo conforme as características das imagens analisadas.

Posteriormente, o objetivo é isolar as regiões apicais e basais associadas a cada núcleo. Para isso, foram desconsiderados os pixels localizados acima de uma distância *t* (definida previamente, empiricamente de acordo com a forma dos núcleos). A partir disso, torna-se possível exibir todos os pixels do embrião, exceto aqueles pertencentes aos núcleos, dividindo o embrião em duas regiões (Fig. 27(a)). Por fim, analisamos a área total de cada região para classificar como citoplasma apical ou basal (a área do apical é maior do que a basal). A região do citoplasma entre os núcleos e a membrana plasmática determina o



Figura 25 – Aplicação da transformada distância em uma máscara binária nuclear. A saída é representada por um mapa de distâncias na forma de uma matriz de índices ( $idx$ ). Cada elemento dessa matriz contém o índice do pixel diferente de zero mais próximo da imagem binária. Cada cor no mapa é associado à um respectivo núcleo da máscara.

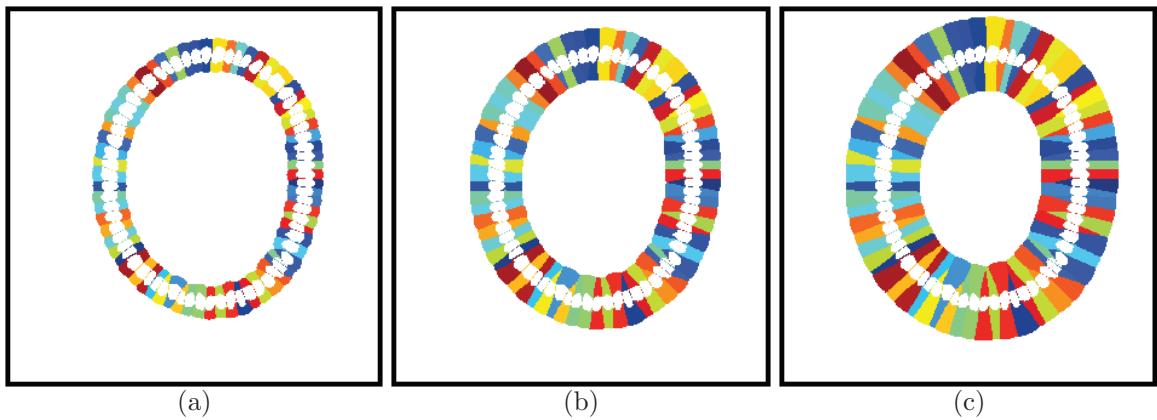


Figura 26 – Diferentes tamanhos da transformada distância utilizados na delimitação das regiões do citoplasma (do menor (a) para o maior (c)).

citoplasma apical (Fig. 27(b)) e, em contraste, a região interna do embrião próxima aos núcleos determina o citoplasma basal (Fig. 27(c)). Uma etapa final reduz a área de cobertura da região apical para a distância máxima igual a  $med/3$ . Esse ajuste, determinado empiricamente, foi necessário porque os núcleos tendem a estar mais próximos da região apical e, como consequência, a área da região apical é menor que a região basal.

Por fim, procede-se à quantificação dos sinais de interesse (etapa (c)). Nela, o objetivo é mensurar a intensidade média de expressão de RNA ou proteína contido nas regiões pertencentes à cada componente do citoplasma. Na Fig. 28 é ilustrado essa etapa final em uma amostra de embrião da base de dados. O contorno das regiões dos citoplasmas apical (Fig. 28(a)) e basal (Fig. 28(b)) são sobrepostos à imagem com marcação de proteína.

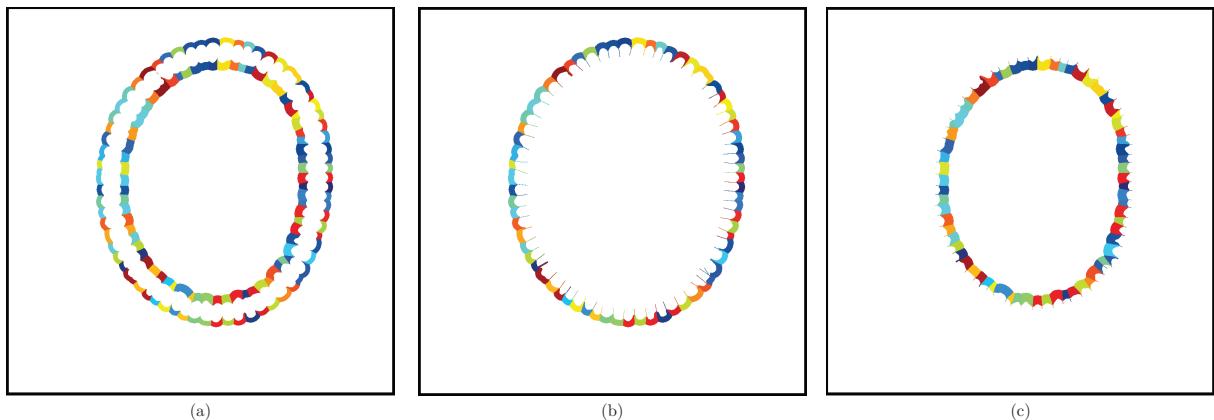


Figura 27 – Determinação das regiões do citoplasma completo (a), apical (b) e basal (c).

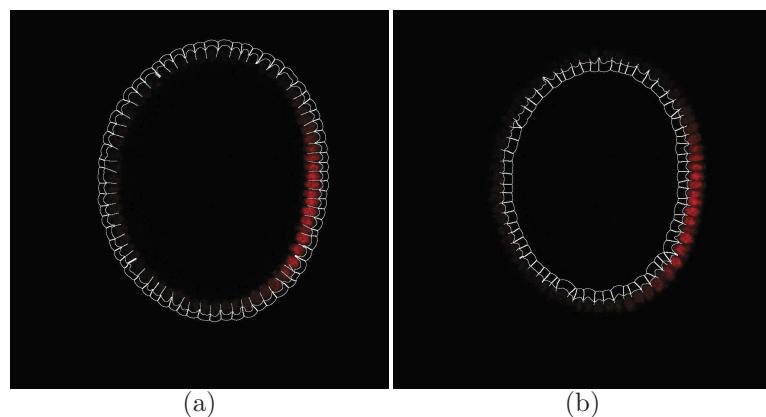


Figura 28 – Sobreposição do contorno dos citoplasmas apical (a) e basal (b) às respectivas imagens com marcação de proteína.

A partir disso, a intensidade média de proteína por região citoplasmática pode ser obtida multiplicando, ponto a ponto, a máscara do citoplasma pela imagem com os dados de expressão da proteína. Esse processo possibilita determinar o nível médio de expressão citoplasmática relativo à cada núcleo contido no embrião.

# CAPÍTULO 5

## Experimentos e Resultados

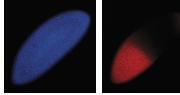
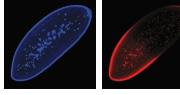
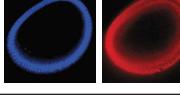
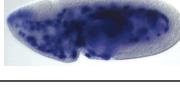
Nesse capítulo são descritos os experimentos realizados e os resultados alcançados pela aplicação dos seis módulos desenvolvidos. Para fins de comparação das propostas apresentadas, para alguns módulos também são apresentados resultados referentes a outros métodos presentes na literatura. Primeiramente serão descritas as bases de imagens usadas nos experimentos e em seguida os resultados quantitativos e qualitativos da aplicação do módulos propostos nessas bases.

### 5.1 Bancos de imagens

A metodologia foi aplicada à imagens confocais e não-confocais (digitais ópticas) de embriões da *Drosophila* relativos à três bancos de imagens distintos: (i) particular (composto por quatro conjuntos de imagens confocais), (ii) FlyEx (Sec. 3.2.3) e (iii) BDGP (Sec. 3.2.5). Na Tab. 3 é apresentado um resumo das bases usadas e a quantidade de imagens analisadas e quais dos módulos propostos foram usados. As seções seguintes apresentam maiores detalhes sobre as bases.

O banco de imagens particular foi construído ao longo do desenvolvimento deste trabalho. As imagens foram capturadas por um microscópio confocal Leica TCS SP5 no Inmetro (Instituto Nacional de Metrologia, Qualidade e Tecnologia). As células de interesse (núcleos) das amostras foram detectados com DAPI (4',6-diamidino-2-phenylindole), um marcador de DNA. Os experimentos de fluorescência foram realizados por uma equipe formada pelo Dr. Francisco José Pereira Lopes, do Campus Duque de Caxias da Universidade Federal do Rio de Janeiro, pela Dra. Helena Araújo e pela Dra. Maira A. Cardoso, do Instituto de Biofísica Carlos Chagas Filho, da Universidade Federal do Rio de Janeiro. Parte das imagens foram capturadas na Universidade de Princeton, em que foi utilizado um aparelho de matriz microfluídica para orientar os embriões. A base completa formou-se a partir de quatro etapas de captura das imagens.

Tabela 3 – Experimentos e bases de imagens tratadas neste trabalho. Imagens próprias, exceto das bases Flyex (PISAREV et al., 2009) e DBGP (TOMANCAK et al., 2007)

Base	Tipo de Imagem	Qtd	Isolamento	Padronização	Segmentação	Rep. Núcleo	Rep. Citopl.	Quantif.
SURFBASE 	Superfície do embrião Marcação dos núcleos: DAPI Padrão de expressão: Hunchback	77	✓	✓	embrião núcleos			✓
SAGIBASE 	Seção sagital Marcação dos núcleos: DAPI Padrão de expressão: Hunchback	77	✓	✓	embrião núcleos	✓		✓
TRANSBASE 	Seção transversa Marcação dos núcleos: DAPI Padrão de expressão: Dorsal	45			núcleos	✓	✓	✓
FlyEx 	Superfície do embrião Marcação dos núcleos: a partir de padrões de expressão de vários genes	250	✓	✓	embrião			
BDGP 	Superfície do embrião Sem identificação dos núcleos Microscopia óptica	61	✓	✓	embrião			

### 5.1.1 SURFBASE e SAGIBASE

Essas bases são compostas por 154 imagens. Desse total, metade das imagens referem-se à superfície do embrião e a outra metade à seção sagital. A caráter de simplificação, esses subconjuntos são descritos neste trabalho, respectivamente, como SURFBASE e SAGIBASE. O conjunto de imagens dessas bases foram coletados em três momentos distintos, nos quais foram utilizadas três lâminas. Na Tab. 5 é detalhado o número de imagens considerada para cada lâmina.

Tabela 5 – Número de imagens considerada para cada lâmina nas bases SAGIBASE e SURFBASE.

	SAGIBASE	SURFBASE
<i>Lam 1</i>	34	34
<i>Lam 2</i>	30	30
<i>Lam3</i>	13	13
Total	77	77

As Figs. 29(a) e 29(b) exibem uma amostra de cada base. Todos os embriões analisados referem-se ao ciclo mitótico 14 (coletados de 130 a 170 minutos após a fecundação) e tiveram os núcleos detectados com DAPI. As imagens foram geradas usando a mesma configuração de um microscópio confocal Leica TCS SP5, obtendo como saída o formato TIFF, dimensão 1024 x 1024 pixels e representadas numa escala de 256 níveis de cinza.

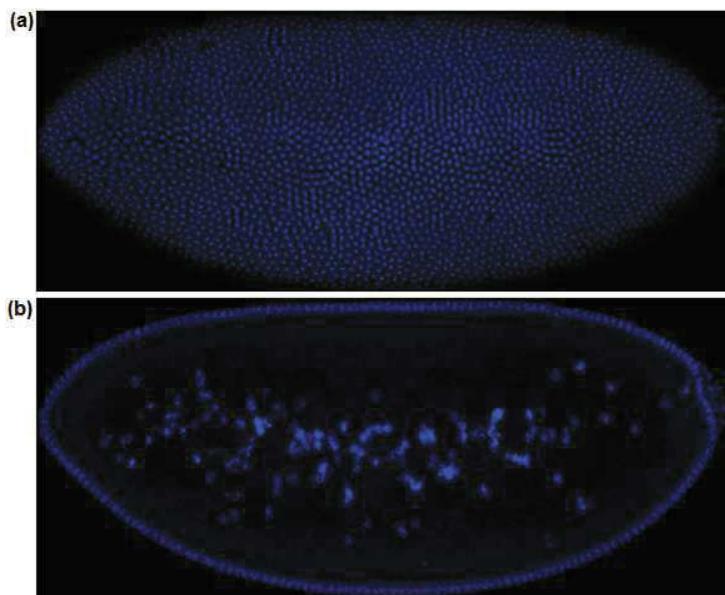


Figura 29 – Exemplos de imagens de embriões tratadas pelo primeiro banco de imagens formado durante o desenvolvimento deste trabalho. Em (a) superfície do embrião (amostra da base SURFBASE) e em (b) seção sagital (amostra da base SAGIBASE).

### 5.1.2 TRANSBASE

Uma outra base particular refere-se à imagens de embriões capturados da seção transversa. Nesse caso, o interesse foi investigar o gradiente de Dorsal e genes-alvo de Dorsal a partir de embriões no 14º estágio do ciclo de desenvolvimento. Foram coletadas 45 imagens confocais no formato Formato de arquivo de imagem marcado (*Tagged Image File Format*) (TIFF) e tamanho  $512 \times 512$  pixels. Dessas, 23 imagens referiam-se a embriões do tipo selvagem e 22 a embriões do tipo mutante. Três diferentes genótipos foram preparados e analisados, sendo *cactA2-Df<sup>cact</sup>*, *cactA2-dl<sup>6</sup>* e *dl<sup>6+</sup>*. Todas as imagens foram geradas usando a mesma configuração confocal e dois canais. No primeiro canal, as células de interesse (núcleos) das amostras foram detectadas a partir do marcador de DNA - DAPI (4',6-diamidino-2-phenylindole). No segundo canal, o nível do gradiente de Dorsal foi coletado. Para cada genótipo mutante, um conjunto de imagens do tipo selvagem foi coletado no mesmo lote, sendo considerado como controle do genótipo. Na Tabela 6 é detalhado o número de imagens considerada para cada genótipo.

Genótipo	Selvagem (controle)	Mutante
<i>cactA2-Df<sup>cact</sup></i>	8	4
<i>cactA2-dl<sup>6</sup></i>	12	10
<i>dl<sup>6+</sup></i>	3	8
Total	23	22

Tabela 6 – Número de imagens considerada para cada genótipo.

A base em questão é nomeada neste trabalho como TRANSBASE. É importante dizer que várias dessas imagens exibiram baixo contraste e grande variabilidade no nível de intensidade dos pixels, dificultando até mesmo a identificação visual das regiões nucleares. Essas imagens também apresentaram uma maior quantidade de ruído nas região circundante aos núcleos. Em imagens confocais o ruído é considerado o sinal fluorescente não originado da região de interesse. Tais características prejudicam os resultados de etapas posteriores, como a segmentação. Na Fig. 30 são exemplificadas algumas amostras de imagens dessa base.

### 5.1.3 FlyEx

A base FlyEx (PISAREV et al., 2009) contém imagens de padrões de expressão gênica relativas à embriões da *Drosophila* coletados durante 80 a 170 minutos após a fecundação. Todas as imagens referem-se ao plano frontal e nível de resolução subcelular do embrião. A proteína alvo e os núcleos foram corados com moléculas fluorescentes, como descrito por (KOSMAN; REINITZ; SHARP, 1999). Para a realização do experimento selecionou-se um conjunto contendo 250 imagens arbitrárias, tais que cobrissem o período de clivagem de 10 – 14 e todas as classes temporais do ciclo 14. A Fig. 31 exibe seis em-

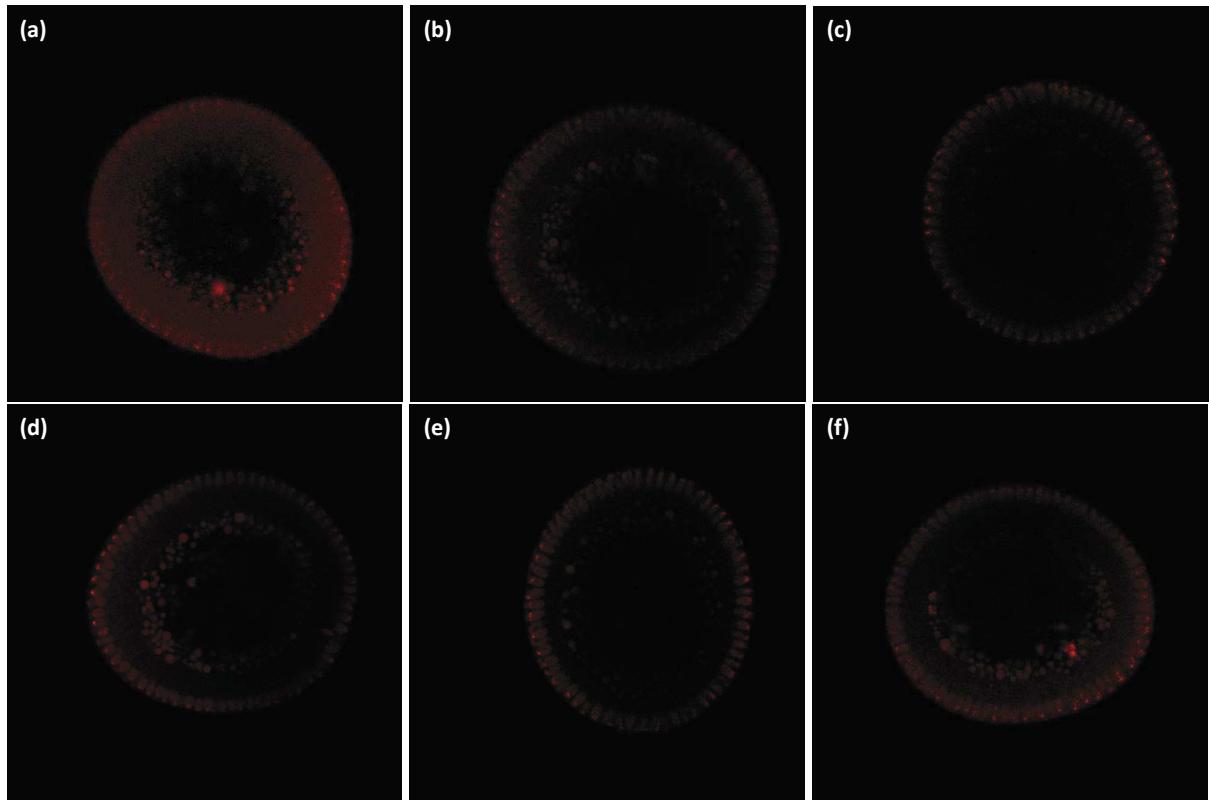


Figura 30 – Exemplos de imagens contidas na base TRANSBASE. (a)–(c) Imagens com baixo contraste e pouca definição das regiões nucleares. (d)–(f) Imagens com conjunto de pixels com altos valores de fluorescência dispersos pelo embrião.

briões de diferentes estágios do desenvolvimento. Originalmente, as imagens dessa base são disponibilizadas na direção horizontal. Entretanto, a fim de avaliar o desempenho do módulo Padronização, as imagens selecionadas para experimentos foram arbitrariamente rotacionadas em diferentes graus.

#### 5.1.4 BDGP

A base BDGP (TOMANCAK et al., 2007) contém imagens digitais coletadas da superfície dos embriões de *Drosophila*, as quais exibem a distribuição espacial de RNA de oito mil genes no estágio pré-gastrula, detectadas por hibridização fluorescente *in situ* (FISH). As imagens dessa base não são padronizadas quanto a orientação, direção e quantidade de objetos, e ainda, podem apresentar múltiplos ou fragmentos de embriões. Na Fig. 32 são exibidas seis imagens de embriões que reforçam a variação das características anteriormente descritas. Para os experimentos foram selecionadas 61 imagens arbitrárias dos estágios de desenvolvimento diversos. Como as imagens são coloridas, na etapa de pré-processamento foi realizada a conversão de RGB para níveis de cinza (usando o componente I do espaço de cores HSI) e em seguida foi aplicada a filtragem por mediana (máscara  $3 \times 3$ ). Essa sequência possibilita melhorar a qualidade das imagens para fases

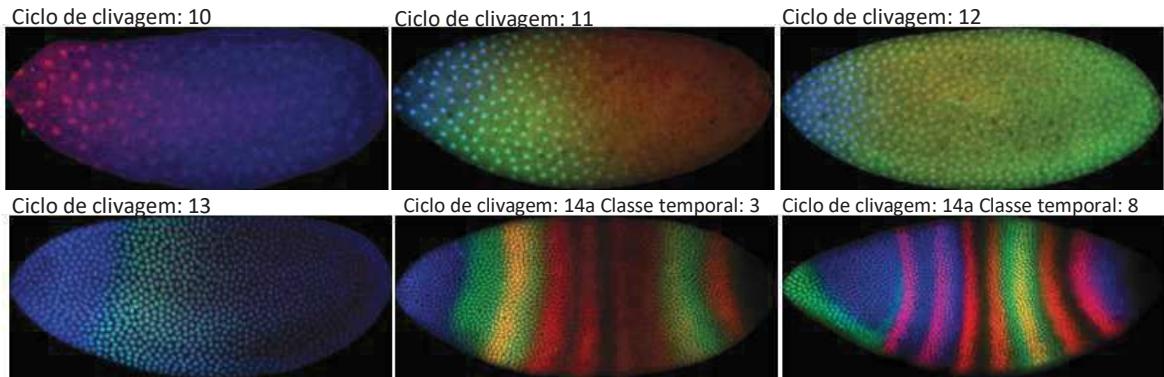


Figura 31 – Exemplos de imagens de embriões de *Drosophila* coletadas da base FlyEx (PISAREV et al., 2009). Todas as imagens analisadas referem-se à embriões capturados do ciclo mitótico 10 – 14, nível de resolução subcelular e plano frontal.

posteriores, tal como ilustrado na Fig. 33.

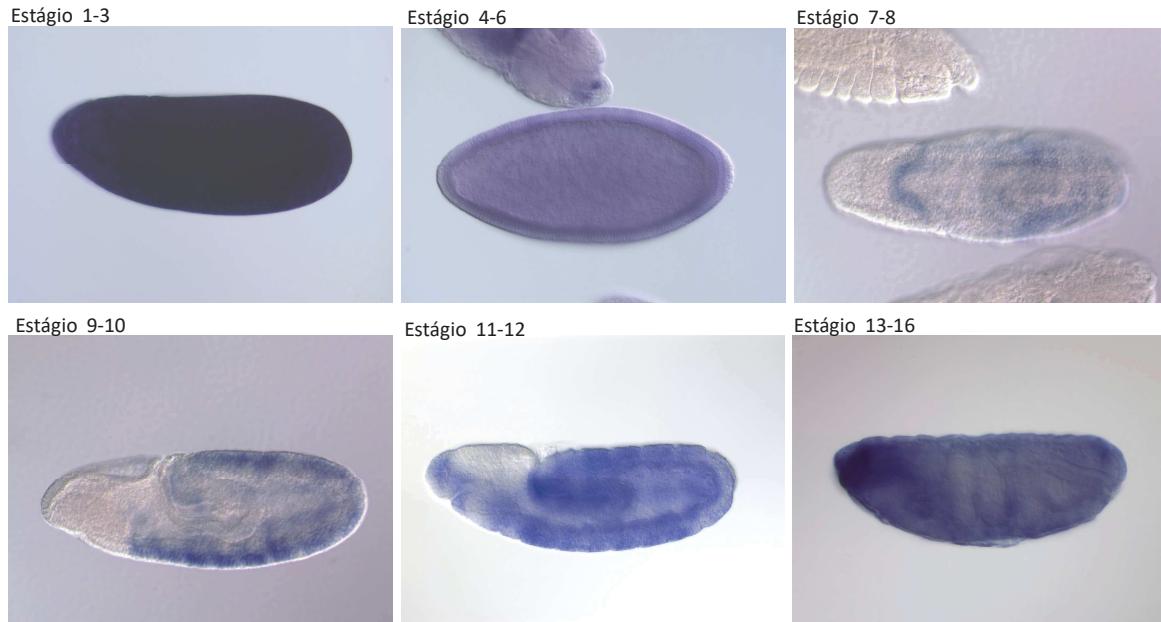


Figura 32 – Exemplos de imagens de embriões de *Drosophila melanogaster* obtidos da base BDGP (TOMANCAK et al., 2007). As imagens são agrupadas em seis grupos 1 – 3, 4 – 6, 7 – 8, 9 – 10, 11 – 12 ou 14 – 16.

A seleção de ambas as bases busca enfatizar a robustez da metodologia proposta, uma vez que permite avaliar os resultados em imagens de embriões de diferentes configurações (como por exemplo, nível de resolução, plano e seção).

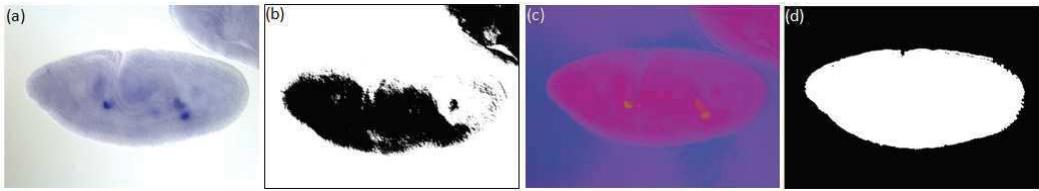


Figura 33 – Pré-processamento em embriões da base BDGP. (a) Imagem do embrião no espaço de cores RGB. (b) Resultado da binarização da imagem exibida em (a). (c) Conversão RGB para níveis de cinza (usando o componente I do espaço de cores HSI). (d) Resultado da binarização da imagem (c).

## 5.2 Resultados

Nesta seção são apresentados os resultados alcançados pela aplicação dos seis módulos desenvolvidos. Os resultados referentes ao Módulo 3 foram dividido em duas seções, visto a etapa intermediária de pré-processamento possuir resultados relevantes de serem mostrados.

### 5.2.1 Módulo 1 – Isolamento do embrião

Nesta seção são descritos os resultados da aplicação do módulo Isolamento do embrião. Esse módulo foi aplicado sobre as bases SAGIBASE, SURFBASE e BDGP. Não houve aplicação sobre a base FlyEx e TRANSBASE, uma vez que essas bases disponibilizam imagens com o embrião principal já isolado.

O módulo de isolamento aplicado às bases SURFBASE e SAGIBASE obtiveram um desempenho de, respectivamente, 98,70% e 93,50%. Esses valores referem-se ao número de máscaras de segmentação do embrião principal corretamente obtidas. Na Fig. 34 são ilustrados alguns resultados qualitativos obtidos da base SAGIBASE, ao passo que, na Fig. 35 são destacadas as saídas positivas da base SURFBASE. Vale observar que, de 154 imagens avaliadas, 50 representavam casos de imagens com mais de um embrião. A abordagem proposta mostrou-se capaz de isolar o embrião principal em todos os casos testados. A Fig. 36 ilustra alguns resultados da aplicação desse módulo em imagens da base SURFBASE.

Na base SURFBASE, o módulo não foi capaz de gerar a máscara do embrião para somente uma imagem do conjunto, e no caso da SAGIBASE, quatro imagens processadas não tiveram um resultado de saída esperado. Na Fig. 37 algumas dessas imagens são destacadas. É possível observar que as imagens que apresentaram erros são marcadas por baixo contraste e qualidade baixa na identificação dos núcleos.

Os resultados da aplicação desse módulo sobre a base BDGP também foram considerados satisfatórios. A metodologia adotada mostrou-se eficiente para a tarefa de isolamento do embrião principal. Do total de imagens contidas na base (61), 11 imagens não foram corretamente binarizadas (essas imagens não forma consideradas nas etapas seguintes da

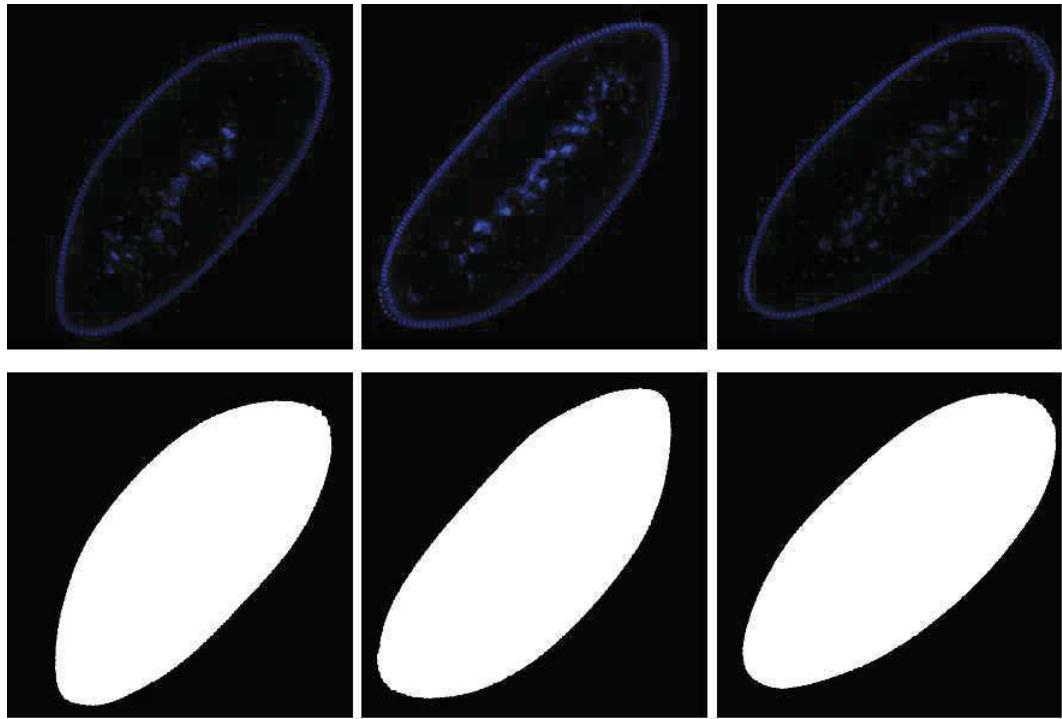


Figura 34 – Exemplos de resultados qualitativos obtidos na geração da máscara binária do embrião principal relativos a base SAGIBASE

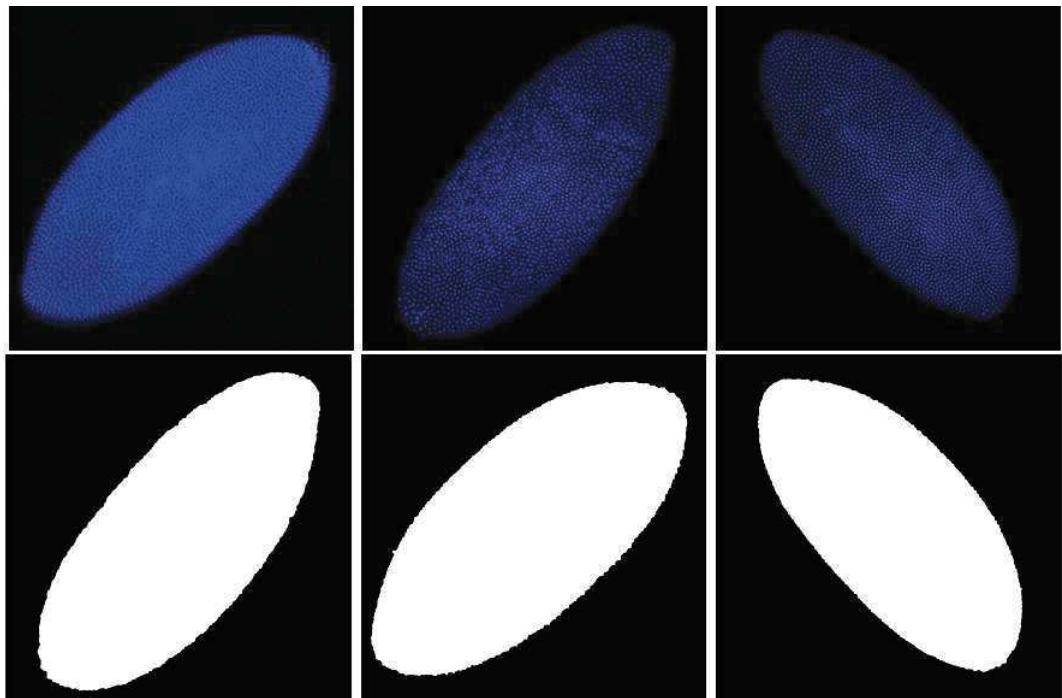


Figura 35 – Exemplos de resultados qualitativos obtidos na geração da máscara binária do embrião principal relativos a base SURFBASE.

metodologia). Vale ressaltar que poucas imagens dessa base foram escolhidas na análise. O resultado obtido não necessariamente representa que o método proposto teria o mesmo desempenho em toda a base. Na Fig. 38 alguns resultados qualitativos são exibidos.

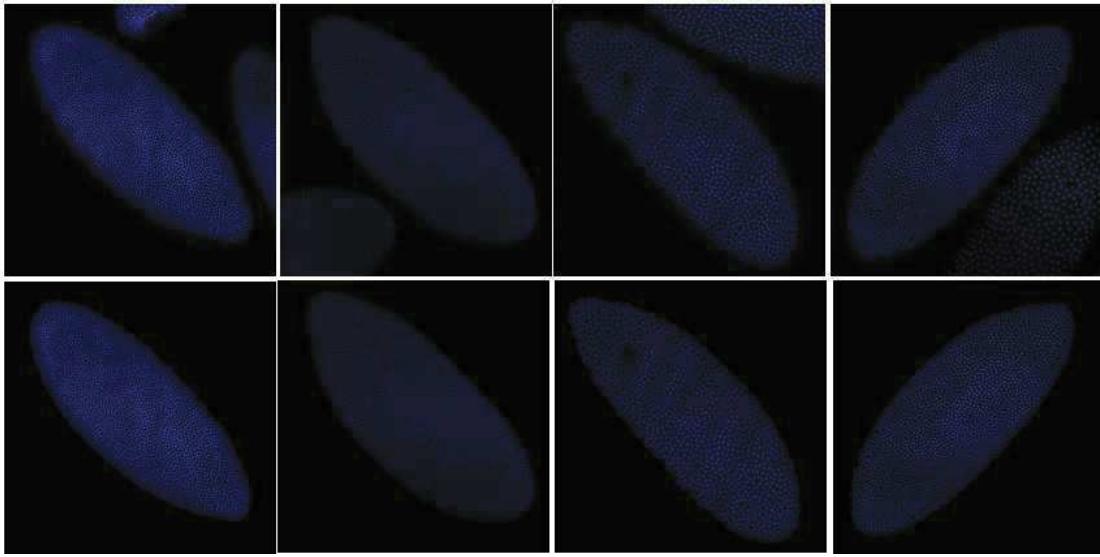


Figura 36 – Isolamento do embrião principal - imagens contidas na base SURFBASE. Na primeira linha, imagens originais da base. Na segunda linha, o resultado do módulo isolamento do embrião principal. É possível perceber que o embrião principal das imagens de entrada foram identificados e isolados, ao passo que fragmentos ou embriões secundários foram removidos.

### 5.2.2 Módulo 2 – Padronização do embrião

Os experimentos com o módulo de Padronização foram conduzidos nas bases FlyEx, BDGP, SURFBASE e SAGIBASE. Em cada base de imagens, os resultados obtidos foram comparados com o desempenho de duas outras técnicas. As técnicas escolhidas para a etapa de avaliação foram propostas por (KONIKOFF et al., 2012) e (GARGESHA et al., 2005). Elas tratam, respectivamente, de uma (i) abordagem semi-supervisionada na qual a correção da orientação requer intervenção humana e (ii) uma estratégia não-supervisionada para padronização da orientação dos embriões.

Simplificadamente, (KONIKOFF et al., 2012) propõem uma abordagem que utiliza um ângulo de rotação para girar a imagem do embrião, de modo que seu eixo principal se alinhe ao eixo horizontal da imagem. O algoritmo de rotação usa interpolação de vizinho mais próximo para definir em 0 (zero) os pixels da imagem de saída que situam-se fora da imagem rotacionada. Essa abordagem padroniza a direção dos embriões para a posição horizontal, mas a correção da orientação, quando necessária, é realizada a partir de supervisão humana. Ao passo que, (GARGESHA et al., 2005) padronizam a direção dos embriões com base em uma estratégia que usa a comparação entre a taxa de pixels contido nas extremidades anterior e posterior do embrião.

O módulo Padronização alcançou 98,6% de sucesso quando aplicado às imagens da base SAGIBASE e 100% quando aplicado na base SURFBASE. No caso dessas bases, do total de imagens nelas contidas (Tab. 5), um pequeno número (seis) não foi considerado. Isso porque trataram-se de casos em que ocorreram erros na saída do módulo anterior,

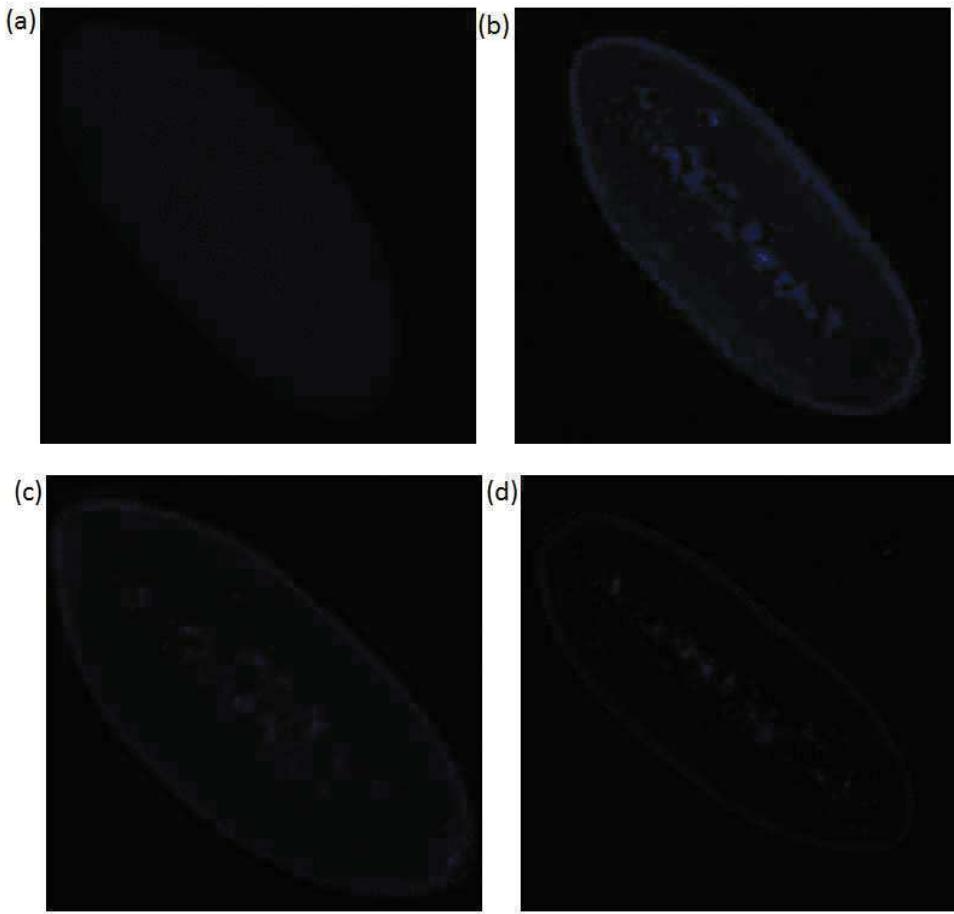


Figura 37 – Exemplos de desafios encontrados nas imagens de embriões de *Drosophila*. (a, c, d) ilustram casos de imagens com baixo contraste. É possível perceber ainda problemas advindos do processo de aquisição das imagens, como má definição do formato nuclear (b). Em tal caso, as regiões dos núcleos celulares apresentam borramentos.

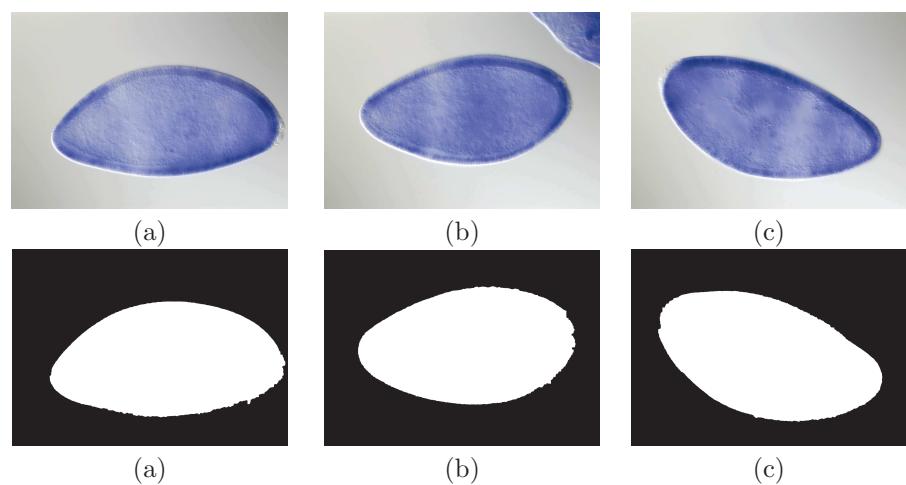


Figura 38 – Exemplos de resultados qualitativos obtidos na geração da máscara binária do embrião principal relativos a base BDGP.

especificamente, na geração das máscaras binárias do embrião principal. Considerando que as máscaras binárias são objetos fundamentais para a execução do módulo Padronização, esse conjunto não foi contabilizado para efeitos de avaliação. Portanto, o total utilizado para experimentos do módulo Padronização foi 148 imagens, das quais 76 eram relativas ao plano frontal e 72 à seção sagital. As bases FlyEx e BDGP foram totalmente utilizadas. Na Tabela 7 são exibidos os resultados quantitativos alcançados, bem como os resultados obtidos pelas propostas (i) e (ii).

Tabela 7 – Taxa de sucesso do módulo Padronização quando aplicado às bases SAGIBASE e SURFBASE.

Método	SAGIBASE	SURFBASE
Método proposto	98,6%	100%
Gargesha et al. (GARGESHA et al., 2005)	48,6%	51,3%
Konikoff et al. (KONIKOFF et al., 2012)	40,3%	48,7%

Na base FlyEx, o módulo apresentou uma taxa de acerto de 92,0%, comparado aos 66,4% alcançados pela abordagem de (GARGESHA et al., 2005) e 60,0% pela abordagem de (KONIKOFF et al., 2012) (Tabela 8). A aplicação do mesmo na base BDGP também mostrou um desempenho superior, como pode ser conferido na Tabela 9.

Tabela 8 – Taxa de sucesso do módulo Padronização quando aplicado à base FlyEx.

Método	FlyEx Base
Método proposto	92,0%
Gargesha et al. (GARGESHA et al., 2005)	66,4%
Konikoff et al. (KONIKOFF et al., 2012)	60,0%

Diferentes morfologias exibidas por embriões de diferentes bancos de imagens explicam a baixa taxa de sucesso das estratégias de (KONIKOFF et al., 2012) e (GARGESHA et al., 2005). Por outro lado, o módulo proposto mostrou-se robusto para contornar esse desafio da variação e complexidade dos dados.

Tabela 9 – Taxa de sucesso do módulo Padronização quando aplicado à base BDGP.

Método	BDGP Base
Método proposto	90,0%
Gargesha et al. (GARGESHA et al., 2005)	60,0%
Konikoff et al. (KONIKOFF et al., 2012)	40,0%

Em todas as bases analisadas, os resultados experimentais provaram que o módulo Padronização é confiável e altamente eficaz, podendo ser facilmente aplicado em uma ampla

gama de imagens de embriões da *Drosophila*. Resultados qualitativos de sua aplicação podem ser conferidos nas Figs. 39, 40, 41 e 42.

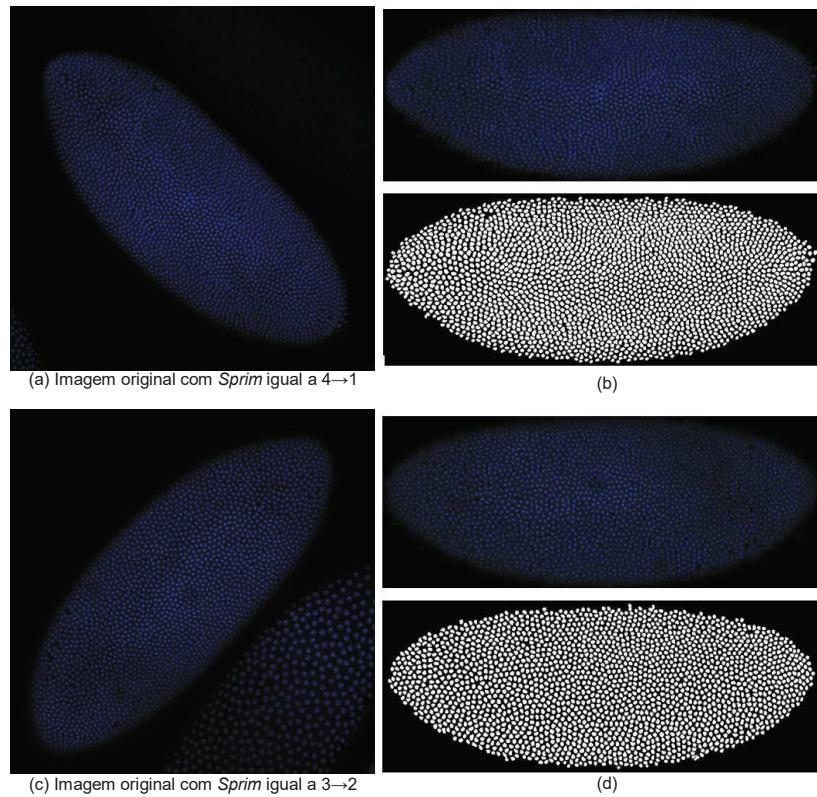


Figura 39 – Resultado qualitativo do módulo Padronização quando aplicado à base de imagens SURFbase. (a) e (c) Imagens originais da base com diferentes orientações. (c) Uma parte de embrião secundário é automaticamente removida pelo módulo Isolamento do EP. (b) e (d) Subfiguras superiores exibem os resultados do módulo Padronização quando aplicado às imagens (a) e (c); Subfiguras inferiores ilustram os resultados do módulo quando aplicado às respectivas máscaras de segmentação nuclear.

A metodologia mostrou-se capaz de padronizar as propriedades direção e orientação de embriões da *Drosophila* em diferentes configurações. Confiável e robusta, ela pode ser aplicada a variados tipos de imagens, incluindo confocais e não-confocais (como, ópticas *in situ*). Os resultados alcançaram uma taxa de sucesso superior a 90% em comparação com os máximos de 67% e 60% obtido pelas técnicas correlatas avaliadas. Diferenças na morfologia exibidas pelos embriões de diferentes bancos de dados explicam a baixa taxa de sucesso das estratégias de (GARGESHA et al., 2005) e (KONIKOFF et al., 2012). Isso porque, apesar de proporem os métodos mais próximos do presente módulo, elas não foram desenvolvidas para trabalhar com a heterogeneidade apresentada pelas bases de dados aqui abordadas. Além disso, o método de (KONIKOFF et al., 2012) é semi-supervisionado, e os resultados apresentados limitam-se a saída das etapas automáticas.

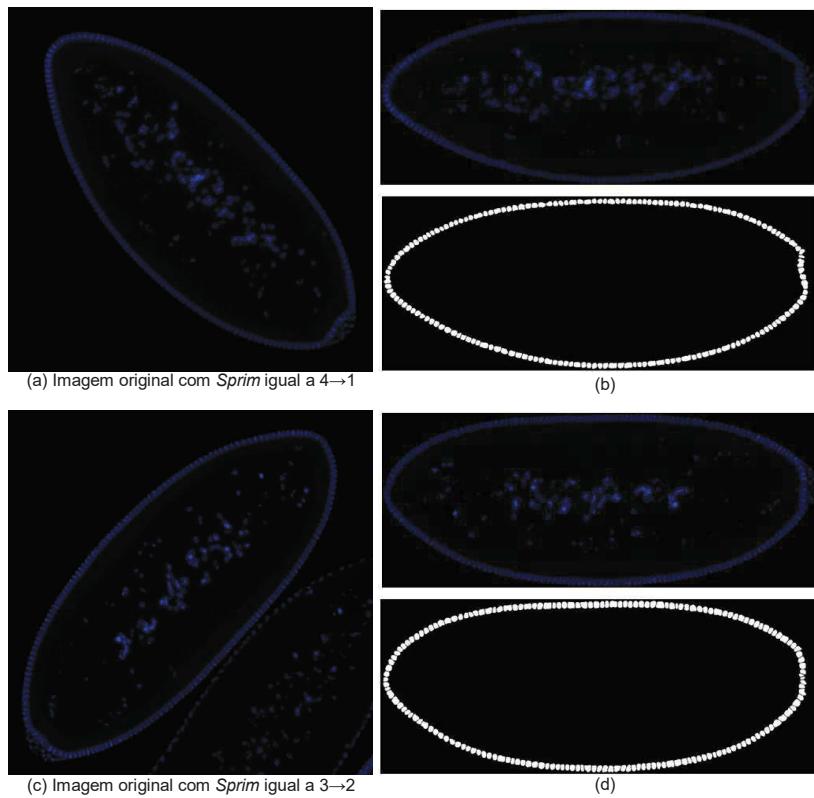


Figura 40 – Resultado qualitativo do módulo Padronização quando aplicado à base de imagens SAGIBASE. (a) e (c) Imagens originais da base com diferentes orientações. (c) Uma parte de embrião secundário é automaticamente removida pelo método. (b) e (d) Subfiguras superiores exibem os resultados do módulo quando aplicado às imagens (a) e (c); Subfiguras inferiores ilustram os resultados do módulo quando aplicado às respectivas máscaras de segmentação nuclear.

### 5.2.3 Módulo 3 – Segmentação dos núcleos: pré-processamento

Os resultados da aplicação da parte de pré-processamento do módulo de segmentação podem ser avaliados qualitativamente. As técnicas utilizadas, tal como descritas na Seção 5.2.4, são aplicadas visando diminuir o ruído das imagens, e com isso, melhorar o resultado de operações posteriores, como a segmentação.

Na Fig. 43 é destacado alguns resultados da aplicação deste módulo sobre imagens da superfície dos embriões. Para realizar essa avaliação, ao invés de exibir a imagem após o pré-processamento, será mostrado o resultado de uma binarização simples (utilizando a técnica Otsu) antes e depois da aplicação do pré-processamento. É nítido a obtenção de melhores resultados após o pré-processamento das imagens. Antes disso, a saída apresenta vários núcleos aglomerados em regiões da imagem com baixo contraste, ou ainda, núcleos ausentes ou super-segmentados em áreas do embrião de iluminação irregular.

A Fig. 44 destaca os efeitos do pré-processamento em imagens de embriões da seção

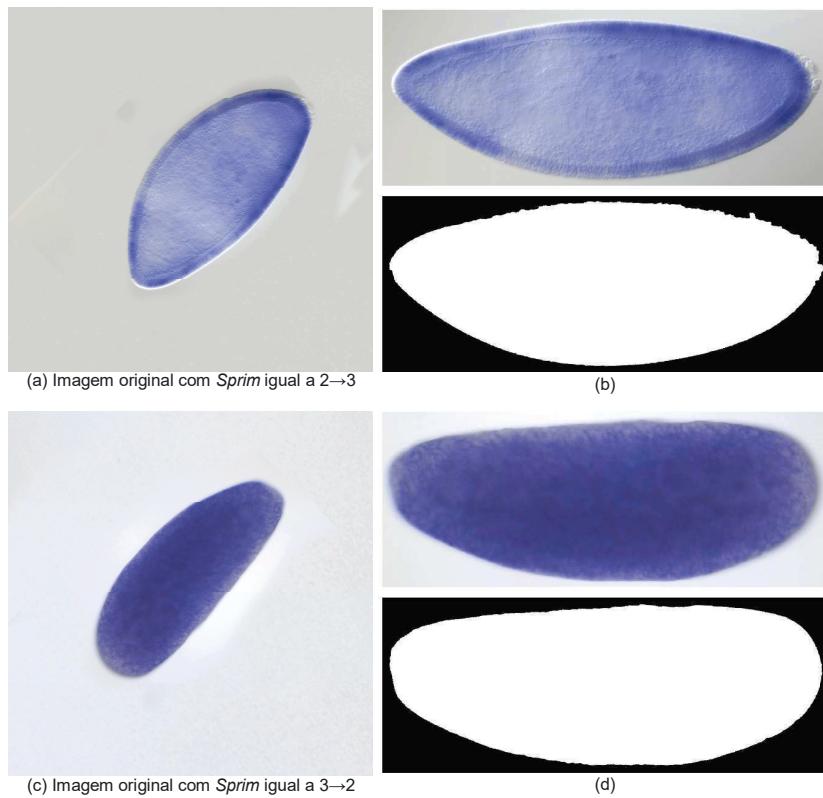


Figura 41 – Resultado qualitativo do módulo Padronização quando aplicado à base de imagens BDBP . (a) e (c) Imagens originais da base com diferentes orientações. (b) e (d) Subfiguras superior exibem os resultados do módulo quando aplicado às imagens (a) e (c); subfiguras inferior ilustram os resultados do módulo quando aplicado às respectivas máscaras de segmentação nuclear.

sagital. Os resultados obtidos não são os esperados, uma vez que a técnica de segmentação aplicada é uma binarização simples. Entretanto, ainda no caso dessa abordagem, o pré-processamento é capaz de melhorar o resultado final. A saída apresenta núcleos com formatos melhor conservados e um embrião mais limpo.

Com relação ao nível de resolução celular e plano frontal objetivou-se a obtenção da máscara binária do embrião. Esse tipo de imagem foi abordado na base BDGP. Conforme mencionado anteriormente, as imagens dessa base são bem variadas no que diz respeito às características de iluminação das imagens, quantidade e formato dos embriões. Sendo assim, o Pré-processamento adotado nesse contexto mostrou-se bastante útil para contornar esse desafio. Em alguns casos analisados da base (especificamente 11), o módulo em questão não foi capaz de melhorar/ou corrigir as características das imagens para etapas posteriores (alguns exemplos são mostrados na Fig. 45).

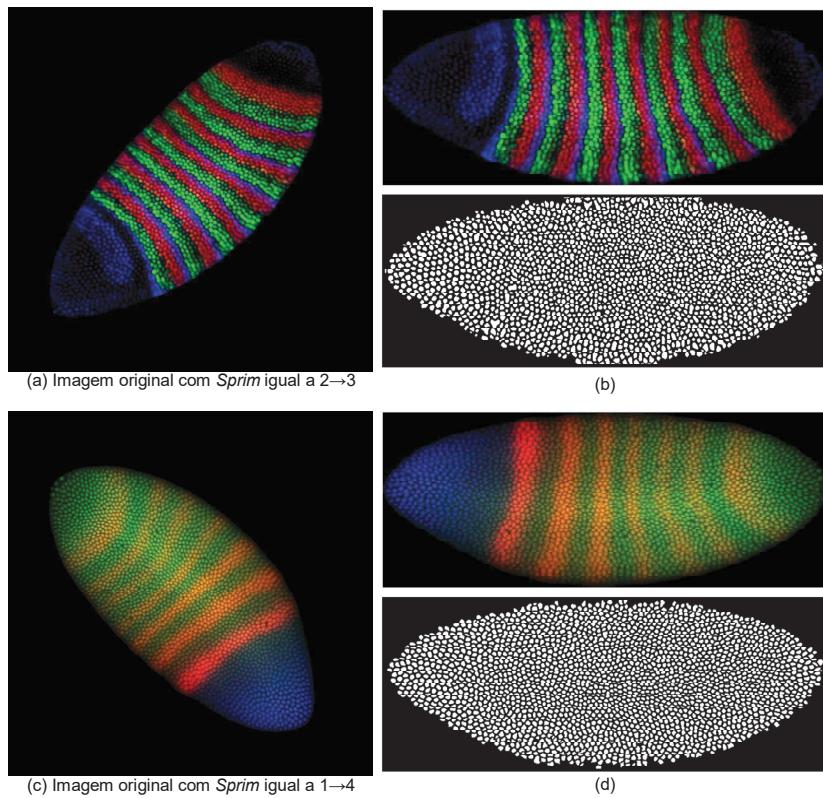


Figura 42 – Resultado qualitativo do módulo Padronização quando aplicado à base de imagens FlyEx. (a) e (c) Imagens originais da base com diferentes orientações. (b) e (d) Subfiguras superiores exibem os resultados do módulo quando aplicado às imagens (a) e (c); Subfiguras inferiores ilustram os resultados do módulo quando aplicado às respectivas máscaras de segmentação nuclear.

#### 5.2.4 Módulo 3 – Segmentação dos núcleos

Nesta seção são apresentados os resultados do módulo Segmentação após a etapa de pré-processamento. Esse módulo foi avaliado utilizando imagens das seções sagital e transversa, e também em imagens da superfície do embrião.

O método foi aplicado nas bases SAGIBASE (imagens relativas à seção sagital) e TRANSBASE (imagens relativas à seção transversa). Os experimentos foram avaliados apenas qualitativamente. Em tal caso, foi realizada uma sobreposição do contorno da máscara de segmentação nuclear sobre a imagem original com marcação DAPI. Essa estratégia possibilita observar detalhes importantes nos resultados, tais como a preservação das formas e regularidade dos núcleos segmentados. Além disso, torna-se possível identificar possíveis erros da segmentação, tais como núcleos ausentes, sobresegmentados ou supersegmentados. A Fig. 47 destaca alguns resultados alcançados a partir da base SAGIBASE, ao passo que, a Fig. 46 ilustra resultados da base TRANSBASE. Vale observar que o mesmo método, em um trabalho anterior, foi aplicado sobre a base de imagens sagitais em que obteve 94,4% de núcleos corretamente segmentados. Em (SOUZA et al.,

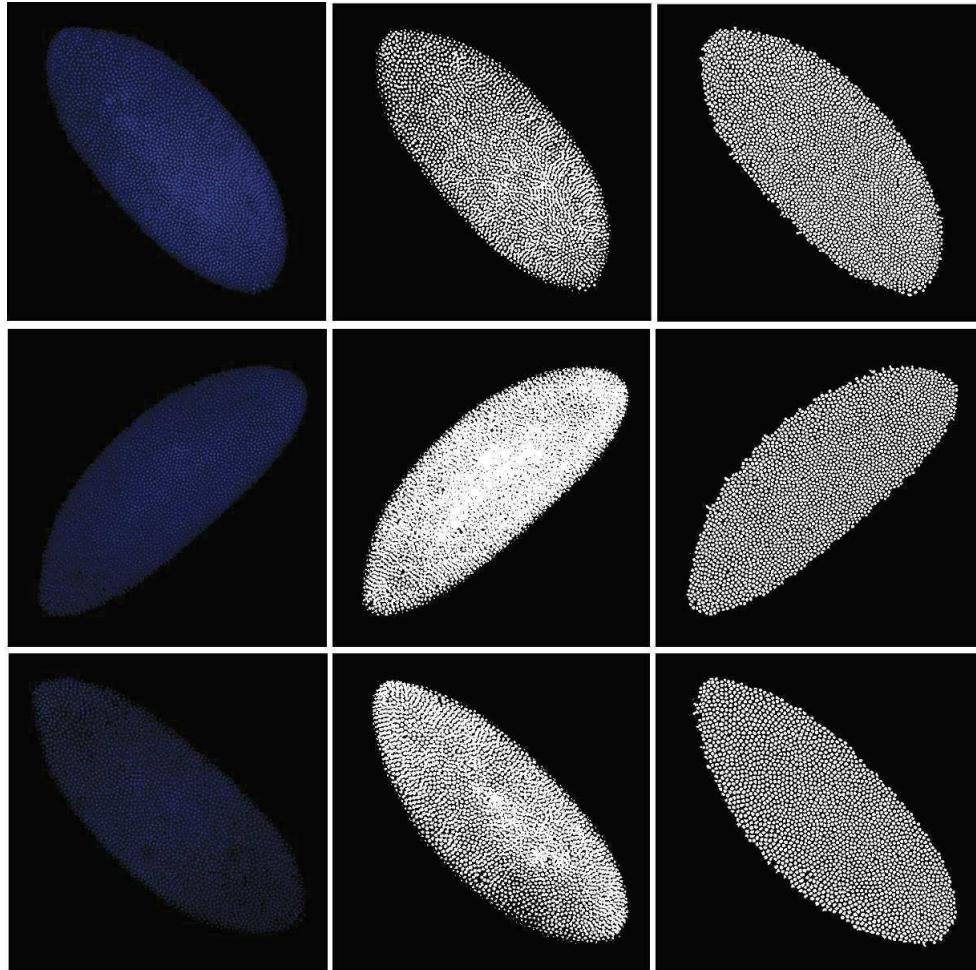


Figura 43 – Pré-processamento aplicado à imagens da base SURFBASE. A primeira coluna exibe três imagens arbitrárias com marcação DAPI. Na segunda coluna são ilustrados os resultados da segmentação sobre as imagens DAPI antes do pré-processamento. E por fim, a terceira coluna exibe o resultado da segmentação após aplicação do módulo em questão.

2013) é possível encontrar uma discussão ampliada a respeito dos experimentos realizados e resultados alcançados para essa estratégia de segmentação.

A partir das bases anteriormente descritas é possível afirmar que os resultados obtidos foram satisfatórios. Erros de núcleos sobresegmentados ou supersegmentados foram encontrados, tal como exemplificado na Fig. 48. Características específicas desse tipo de imagem, tais como baixo contraste e iluminação irregular justificam em partes esse resultado negativo. Entretanto, ainda assim, o desempenho do método de segmentação proposto é superior quando comparado à utilização isolada de técnicas tradicionais da literatura (SOUZA et al., 2013). A aplicação eficaz em diferentes bases comprova a robustez e flexibilidade da abordagem descrita.

A segmentação das imagens da superfície do embrião foram realizados sobre a base SURFBASE. Os resultados alcançados foram bastante positivos. Em ampla maioria dos casos, o método proposto foi capaz de identificar corretamente os núcleos celulares da

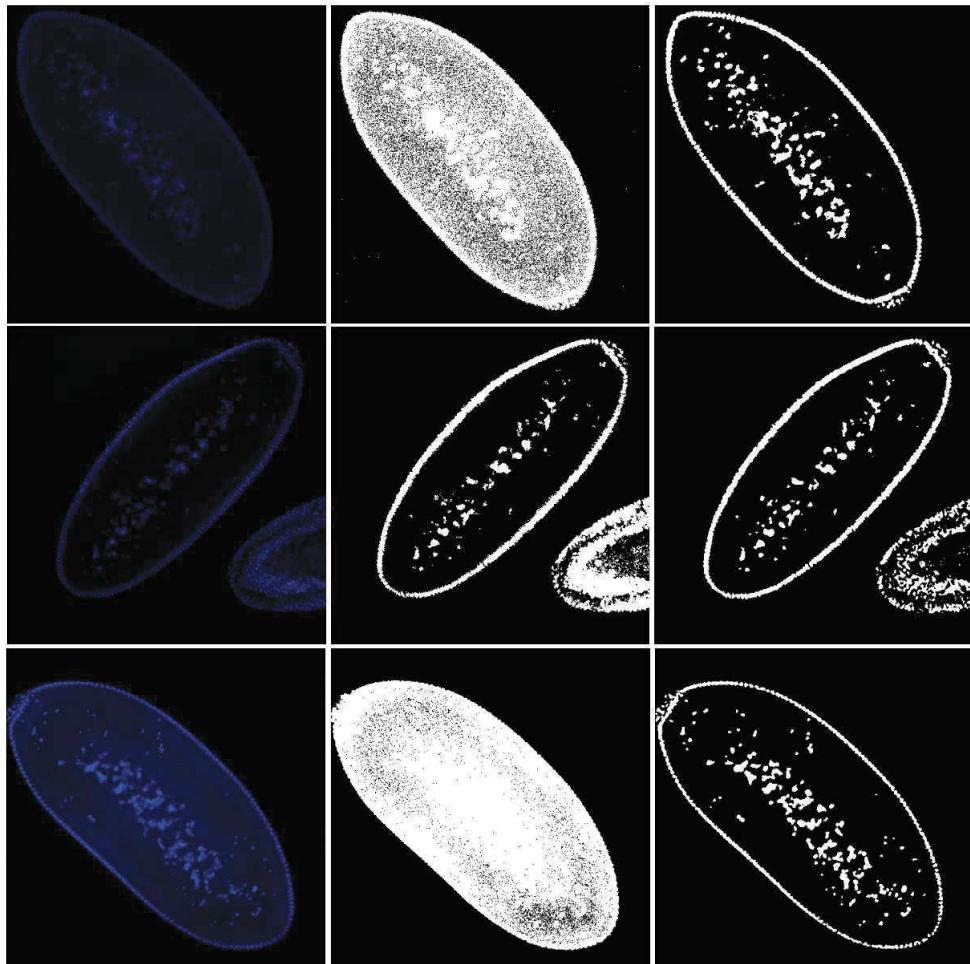


Figura 44 – Pré-processamento aplicado à imagens da base SAGIBASE. A primeira coluna exibe três imagens arbitrárias com marcação DAPI. Na segunda coluna são ilustrados os resultados da segmentação sagital sobre as imagens DAPI antes do pré-processamento. E por fim, a terceira coluna exibe o resultado da segmentação após aplicação do módulo em questão.

superfície dos embriões. Os resultados alcançados foram superiores quando comparado à aplicação isolada de técnicas existentes. Na Fig. 49 são ilustrados os resultados alcançados com as imagens da base SURFbase.

De modo geral, os resultados alcançados comprovam a eficácia do módulo Segmentação. A diversidade e complexidade dos dados tratados demonstra sua capacidade de contornar alguns dos principais problemas desse cenário. Na Fig. 50 alguns dos desafios encontrados nas bases são exibidos.

### 5.2.5 Módulo 4 – Representação dos dados nucleares

Nesta seção são exibidos os resultados qualitativos alcançados com o módulo Representação dos dados nucleares (Seção 4.6) a partir da base SAGIBASE. Esse módulo apresenta a estratégia de visualização e organização dos dados de expressão gênica conti-

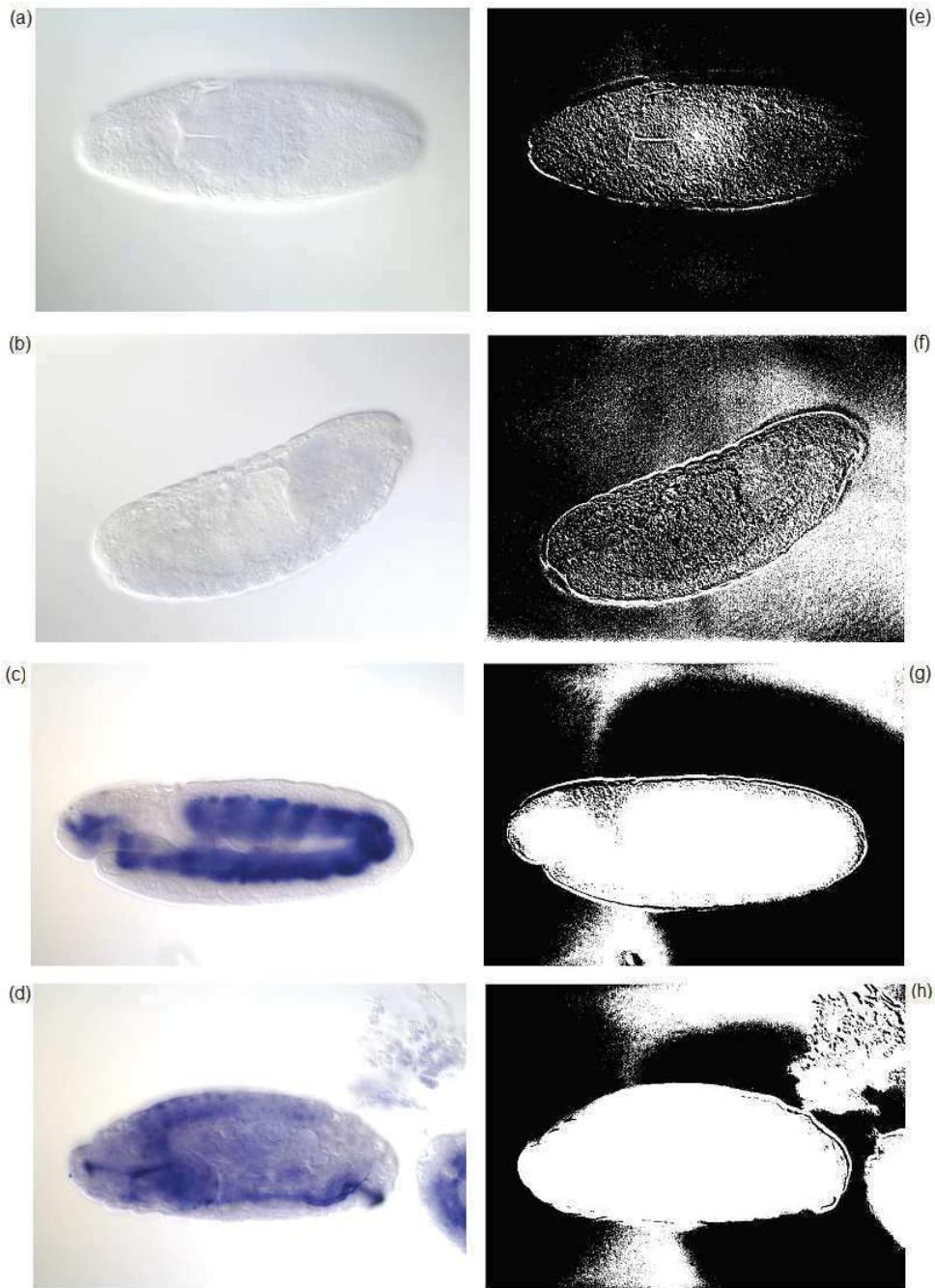


Figura 45 – Exemplos de imagens de embriões da base BDGP com variação das características de brilho, contraste e iluminação. Na primeira coluna, as imagens (a–b) destacam casos em que o embrião principal apresenta-se muito difuso com o plano de fundo. Nas figuras (c–d) são ilustrados exemplos de imagens com iluminação irregular, as quais geram pixels (ou aglomerados de pixels) indesejados na imagem de saída. A segunda coluna, por sua vez, ilustra os resultados ruins da etapa de binarização, obtidos, mesmo após a aplicação do método de pré-processamento nas imagens originais (a–d).

dos em imagens de embriões da seção sagital. É importante lembrar que essa abordagem de representação é inovadora.

Na Fig. 51 são apresentadas algumas saídas do módulo. Os resultados são apenas

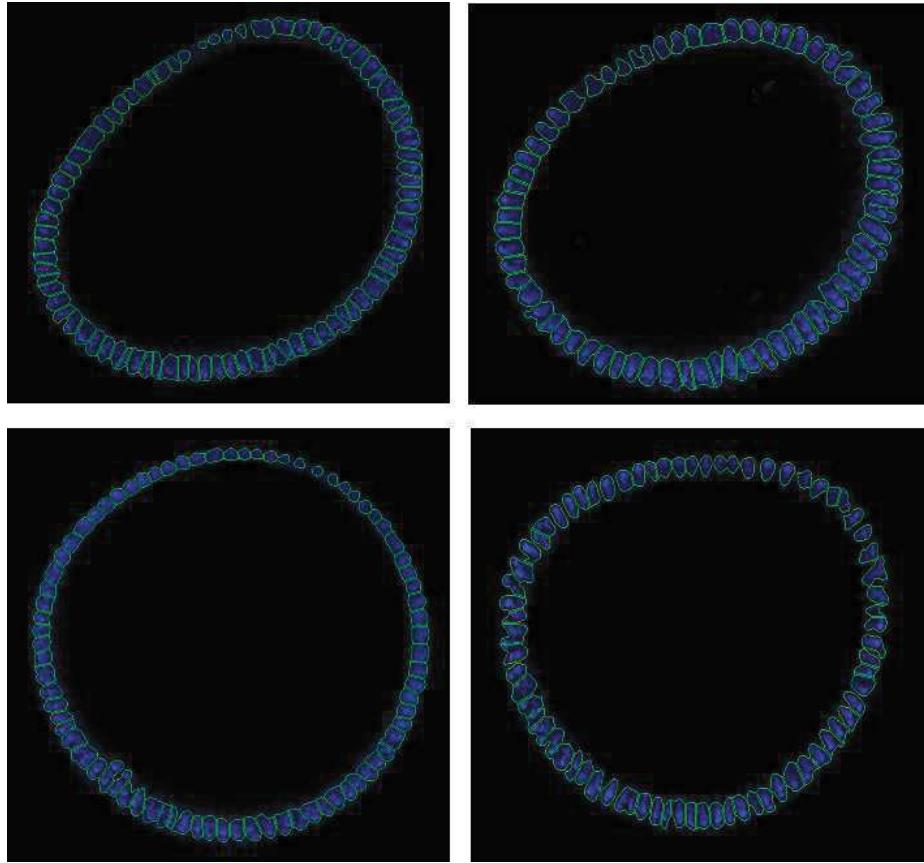


Figura 46 – Resultado qualitativo do método de segmentação nuclear sagital obtidos a partir de imagens da base TRANSBASE.

qualitativos. Para cada objeto inserido na faixa horizontal três parâmetros são associados: (a) lista de *pixels* do ROI (núcleo ou citoplasma), (b) centroide e (c) a média de expressão gênica da região avaliada.

O módulo descrito também é capaz de exibir faixas horizontais para cada lado (dorsal e ventral) dos embriões de *Drosophila* separadamente. Nas Figs. 52–54 são ilustradas as saídas respectivas à Fig. 51.

Por fim, a Fig. 54 exibe o resultado da linearização para alguns embriões da base TRANSBASE.

O módulo em questão mostrou-se prático para a comparação dos padrões de expressão gênica de diferentes embriões, e ainda, útil para a identificação de núcleos agregados.

### 5.2.6 Módulo 5 – Representação dos dados do citoplasma

Nesta seção são exibidos resultados qualitativos alcançados com o módulo cinco (Seção 4.7) a partir da base TRANSBASE. Esse módulo apresenta a estratégia de visualização e organização dos dados de expressão gênica contidos em imagens de embriões da seção transversa. As ROIs aqui são os componentes citoplasma apical e basal.

Os resultados demonstraram que a estratégia proposta cumpre o objetivo inicial pro-

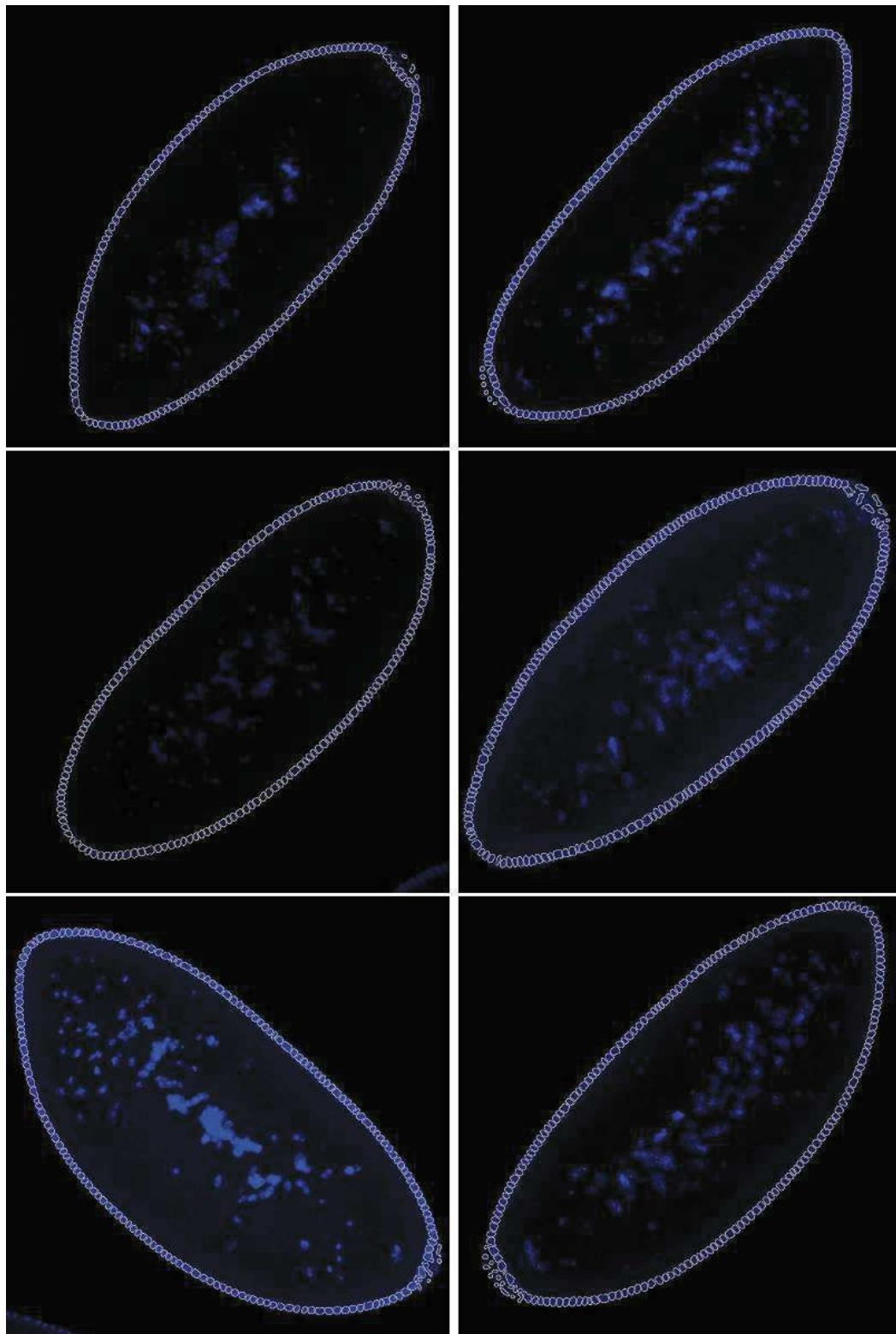


Figura 47 – Resultado qualitativo do método de segmentação nuclear sagital obtidos a partir de imagens da base SAGIBASE.

posto. Isto é, para cada embrião, é possível dispor todos os seus componentes citoplasmáticos (apical ou basal) em faixas horizontais e visualizá-los lado a lado. No entanto, a abordagem não foi considerada totalmente eficiente. Isso porque, parte dos componentes

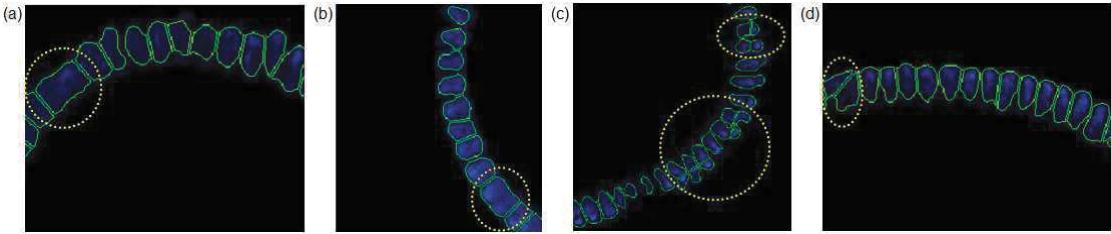


Figura 48 – Erros na segmentação nuclear de imagens sagitais de embriões capturados na seção transversa. Os círculos tracejados na cor amarelo enfatizam os problemas encontrados neste tipo de imagem, como núcleos agregados (a–b) ou supersegmentados (c).

não foram dispostos nas faixas com a orientação adequada. Essa situação é decorrente da estratégia adotada no algoritmo para realizar a rotação das regiões de interesse. A abordagem ajusta a direção vertical nos componentes, mas não efetua uma etapa automática para controle da orientação. Esse processo é detalhado na Seção 4.6.1.5. Consequentemente, uma parcela dos componentes citoplasmáticos não são dispostos na orientação esperada.

Nas Figs. 55 e 56 são exibidos os resultados qualitativos do módulo para um conjunto de imagens de embriões do tipo Selvagem (previamente descritas na Tab. 6). O problema descrito anteriormente pode ser conferido em alguns trechos da saída exibida.

Apesar do desafio descrito, os resultados alcançados comprovam a robustez da abordagem frente à heterogeneidade comum dos tipos de dados tratados nesse cenário, como diferentes tamanhos e formatos das regiões de interesse. Por consequente, a metodologia opera bem em qualquer etapa do ciclo de desenvolvimento embrionário. Além disso, a abordagem pode ser promissora para investigações biológicas, como ordenação dos embriões por idade.

### 5.2.7 Módulo 6 – Quantificação nuclear e citoplasmática

O conhecimento dos padrões espaciais e temporais da expressão gênica é crucial para a compreensão dos processos de desenvolvimento no embrião. No entanto, em grandes bases de dados a inspeção visual desse tipo de imagens é um trabalho lento e repetitivo. Nesta seção são apresentados resultados do módulo Quantificação, em que é realizado a análise automatizada da expressão gênica contida em imagens de embriões da *Drosophila*. Ela trata, paralelamente, da quantificação da região dos núcleos e da região do citoplasma, considerando apenas imagens de embriões no nível de resolução subcelular.

Geralmente, três imagens por gene são geradas para as análises, a saber: uma com marcação DAPI (que identifica os núcleos), uma com marcação dos sinais de proteína codificados e outra dos sinais do RNA (Fig. 57). Nas bases usadas neste trabalho temos somente imagens com intensidades de proteínas. Em todos os casos, o módulo gera dados

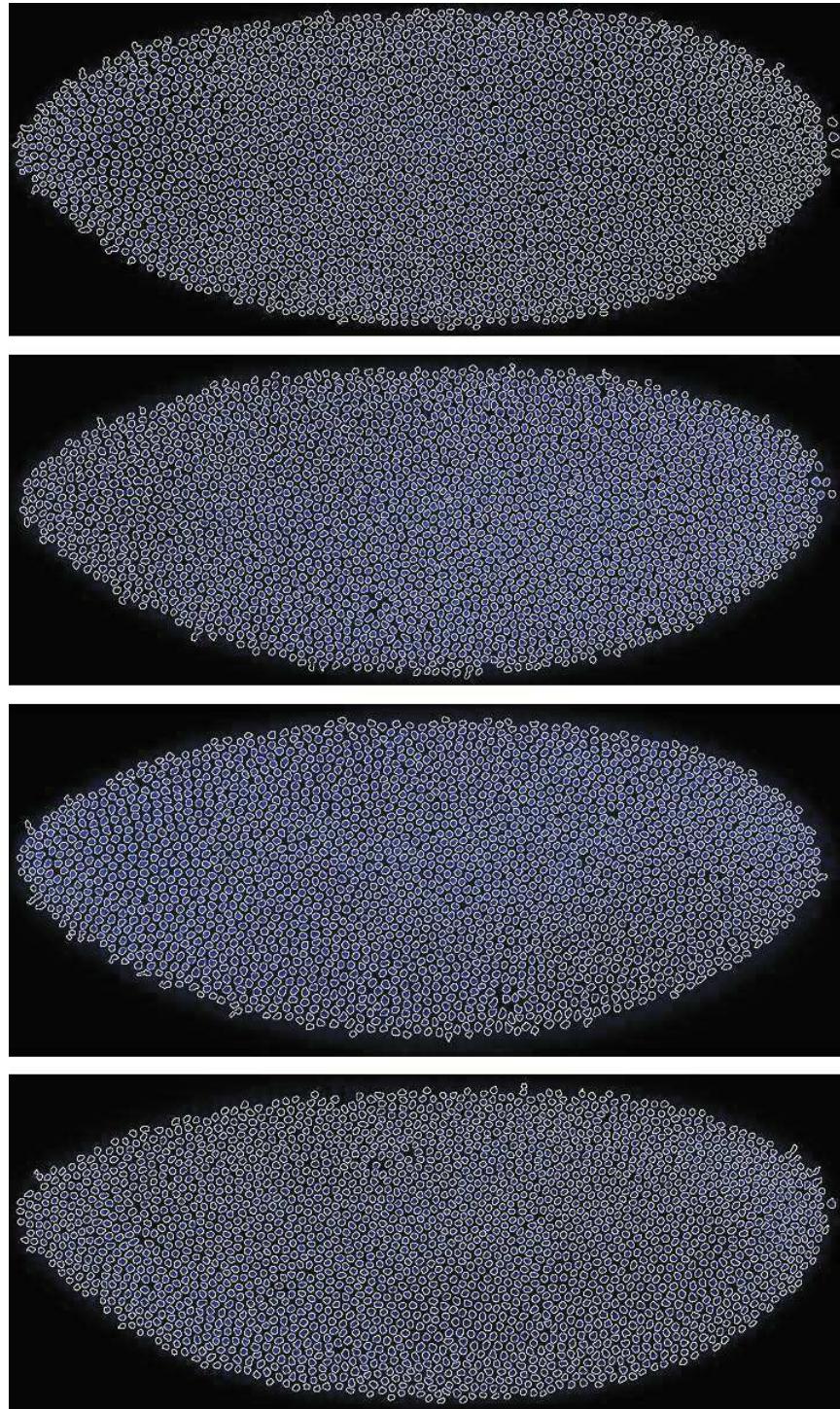


Figura 49 – Resultado qualitativo do método de segmentação nuclear proposto para imagens de embriões da *Drosophila* obtidos do plano frontal e nível de resolução subcelular.

quantitativos a respeito dos sinais de expressão gênica contidos nas regiões analisadas, conforme descrito na Sec. 4.8.

Fig. 58 exibe um resultado da sobreposição das máscaras do citoplasma (apical e basal) sobre uma imagem com expressão de proteína obtida da base TRANSBASE.

Em geral, os resultados da Quantificação possibilitam caracterizar os diferentes pa-

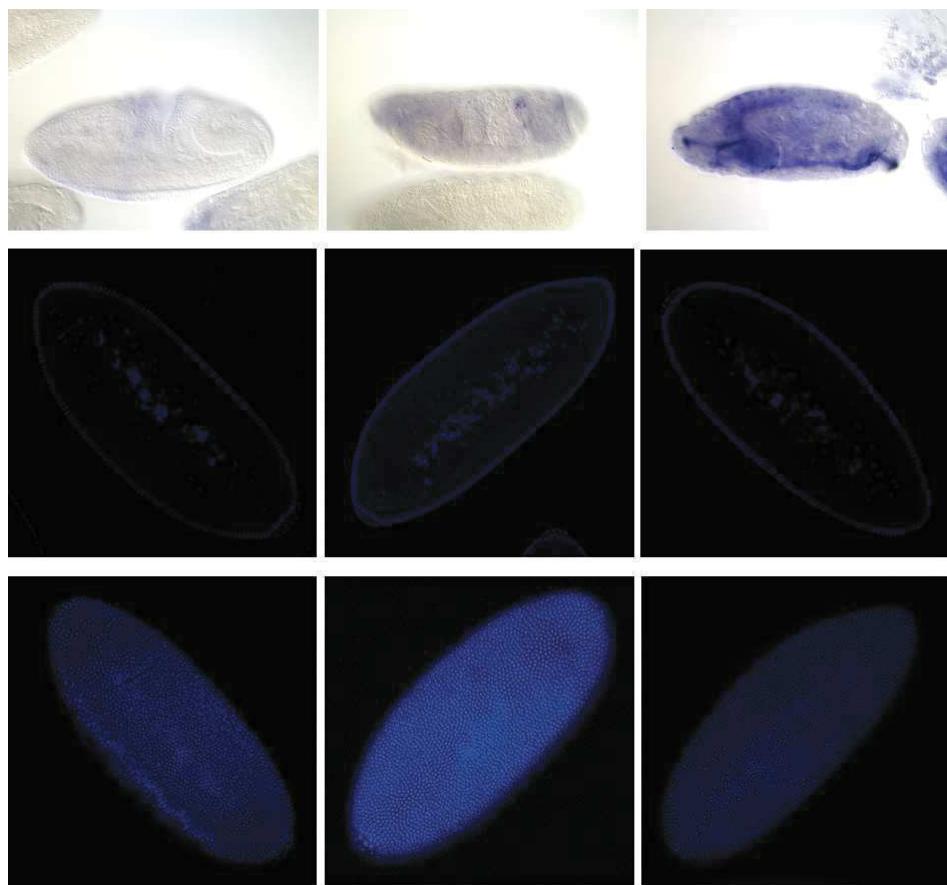


Figura 50 – Exemplos de desafios encontrados nas imagens abordadas pelas bases tratadas neste trabalho. Na primeira são ilustrados casos de iluminação irregular, na segunda linha é exibido problemas de baixo contraste e na última linha imagens com qualidade pobre das amostras coletadas. Neste caso, a identificação dos núcleos celulares se torna difícil até mesmo aos olhos humanos. A maioria destas questões (a–f) podem ser tratadas pela metodologia proposta.

drões de expressão gênica das proteínas e transcritos codificados, respectivamente, nos núcleos e no citoplasma.

### 5.3 Resultados biológicos

O objetivo desta seção é apresentar resultados biológicos relevantes que foram obtidos a partir da utilização da metodologia proposta. Nas seções seguintes são exibidos alguns resultados de perfis de expressão gênica obtidos a partir das bases SURFBASE, SAGIBASE e TRANSBASE. Os resultados dos experimentos estão organizados conforme o tipo de imagem analisada.

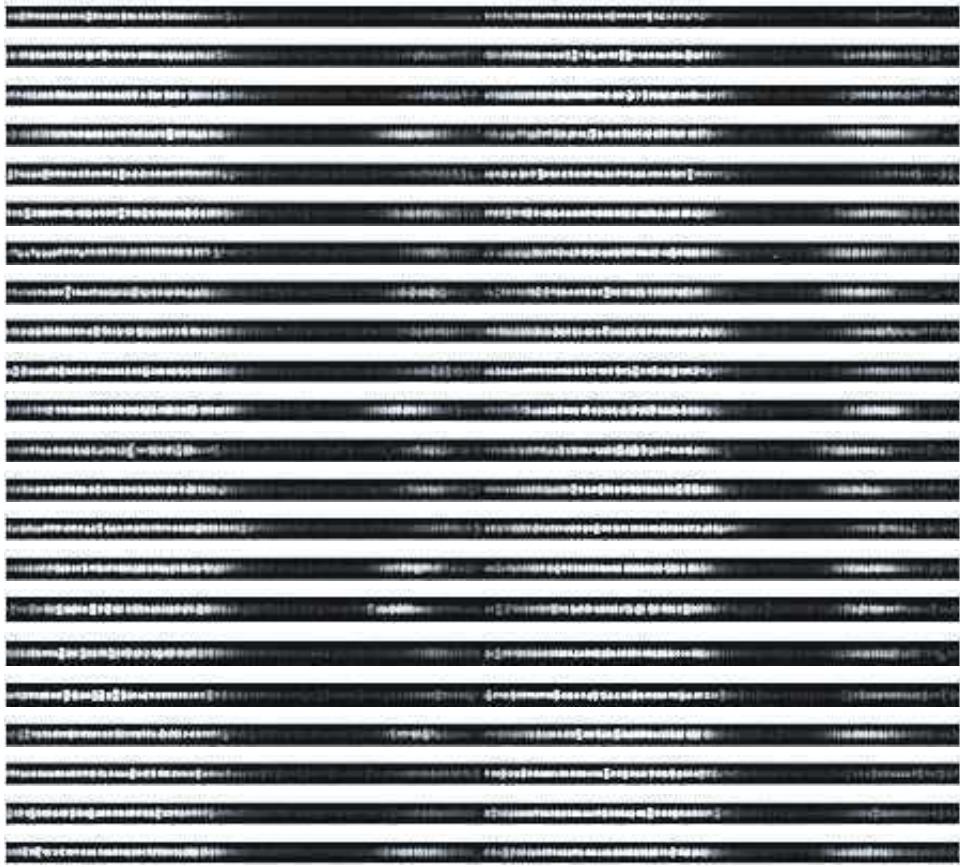


Figura 51 – Resultado do módulo de representação dos dados nucleares quando aplicado sobre algumas imagens da base SAGIBASE. Cada linha representa uma embrião completo linearizado. Essas amostras são do grupo LAM1.

### 5.3.1 Imagens da superfície dos embriões

As imagens relativas à superfície dos embriões foram organizadas na base SURFBASE (Sec. 5.1.1). Os experimentos conduzidos nessa base, objetivaram sobretudo, realizar uma análise quantitativa dos dados obtidos a partir das faixas central e mediatriz. Na presente seção são apresentados os resultados a respeito dos primeiros experimentos biológicos realizados. Nenhuma afirmação ainda pode ser feita, mas é possível observar e procurar por informações relevantes nos perfis de expressão gênica obtidos das faixas central e mediatriz. A Fig. 59 exemplifica quatro resultados obtidos. De modo geral, elas incluem uma comparação dos sinais de expressão gênica das regiões em questão.

Na Fig. 60 é exibida a sobreposição das curvas médias de expressão gênica (central *vs* mediatriz) obtidas de 73 embriões contidos na base SURFBASE. Os experimentos foram realizados separadamente para o conjunto de imagens de cada lâmina, tendo sido selecionando as seguintes quantidades de imagens: Lam1 (33), Lam2 (28) e Lam3(12).

Uma análise geral, indicou que os locais do embrião de maior variabilidade localizam-se na região anterior e no final da região posterior, que corresponde ao padrão já conhecido da proteína *Hunchback*. Essa condição é igual para ambas as faixas, central e mediatriz. Além



Figura 52 – Linearizações geradas pelo módulo de representação dos dados nucleares referentes ao lado dorsal dos embriões ilustrados na Fig. 51. Cada linha representa meio embrião aberto. Essas amostras são do grupo LAM1.

disso, foi observado que a curva média da região mediatrix apresentou um menor gradiente de expressão nos três conjuntos analisados (Lam 1, Lam 2 e Lam 3), demonstrando que há diferença entre realizar medidas entre essas faixas e que isso deve ser levado em consideração nas análises de padrões de expressão.

### 5.3.2 Seção Sagital

As imagens relativas à seção sagital foram organizadas na base SAGIBASE (Sec. 5.1.1). Os experimentos conduzidos nessa base objetivaram avaliar os perfis de expressão gênica da região nuclear dos embriões de *Drosophila*. Para extrair os dados quantitativos, cada máscara nuclear foi sobreposta à respectiva imagem com o sinal de expressão desejado. Valores médios das regiões de interesse (núcleos) então são obtidos ao longo do eixo AP dos embriões. O perfil de expressão traduz o nível médio de fluorescência por núcleo.



Figura 53 – Linearizações geradas pelo módulo de representação dos dados nucleares referentes ao lado ventral dos embriões ilustrados na Fig. 51. Cada linha representa meio embrião aberto. Essas amostras são do grupo LAM1.

A SAGIBASE não possui imagens relativas à expressão no citoplasma, sendo assim gerou-se o perfil de expressão das regiões nucleares com sinais da proteína *Hunchback*. Os perfis mostrados referem-se ao lado dorsal e ventral do embrião, no que diz respeito às intensidades fluorescentes médias das regiões de interesse.

Na Fig. 61 são ilustrados resultados dos experimentos obtidos com essa base. No eixo vertical do gráfico é representada a escala de intensidade de fluorescência e no eixo horizontal é descrito a posição das regiões analisadas no eixo AP do embrião de *Drosophila*.

Na Fig. 62 são mostradas as curvas médias de expressão gênica (dorsal *vs* ventral) obtidas de 67 embriões contidos na base SAGIBASE. Os experimentos foram realizados separadamente para os conjuntos de imagens de cada lâmina, tendo sido selecionando as seguintes quantidades de imagens: Lam1 (28), Lam2 (26) e Lam3(12). É possível notar pelos gráficos uma diferença significativa entre os padrões do lado dorsal e ventral. Isso



Figura 54 – Linearizações geradas pelo módulo de representação dos dados nucleares referentes à base TRANSBASE.

é em parte explicado pela diferença no tamanho e na curvatura entre esses lados. Assim como ocorre na avaliação da expressão entre as faixas central e mediatriz, as diferenças encontradas nessas curvas devem ser consideradas em experimentos que avaliam os padrões de expressão gênica nesses embriões.



Figura 55 – Linearizações geradas pelo módulo de representação dos dados do citoplasma referentes à região basal de embriões da base TRANSBASE. Cada linha representa a linaerização do citoplasma basal completo. Essas amostras são do grupo selvagem.

### 5.3.3 Seção transversa

As imagens relativas à seção transversa foram organizadas na base TRANSBASE (Sec. 5.1.2). Para os experimentos foram selecionadas 45 imagens, a partir das quais objetivou-se gerar os perfis de expressão da proteína Dorsal relativo às regiões nucleares e citoplasmáticas (apical e basal) – ao longo do eixo DV – dos embriões de *Drosophila*.

Considerando as características geométricas desse tipo de imagem (embrião com for-



Figura 56 – Linearizações geradas pelo módulo de representação dos dados do citoplasma referentes à região apical de embriões da base TRANSBASE. Cada linha representa a abertura do citoplasma apical completo. Essas amostras são do grupo selvagem.

mato circular), a configuração do ponto central (CP) para plotagem dos dados baseou-se na detecção da reflexão de simetria, isto é, em encontrar o objeto (núcleo) da imagem que divide o embrião em duas partes, de forma que maximize a simetria do nível de expressão entre as duas partes (detalhadamente descrito na Seção 4.6.1.4). A intensidade de Dorsal foi calculada como o valor médio dentro de cada núcleo e região citoplasmática (apical e

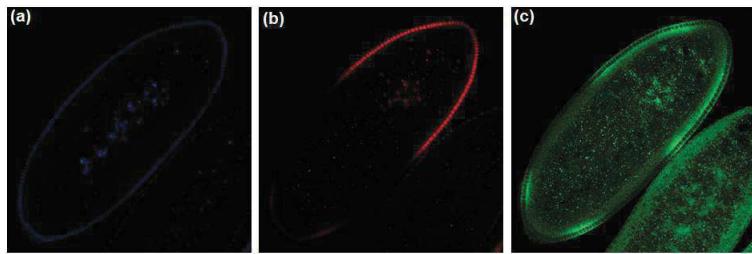


Figura 57 – Imagens do processo de quantificação gênica em embriões de *Drosophila*. (a) Marcação DAPI em um embrião capturado da seção sagital. A partir dessa imagem obtém-se a máscara de segmentação nuclear. (b) Padrão de expressão da proteína *Hunchback*. (c) Distribuição de RNA de *hunchback*. As imagens (b) e (c) são utilizadas, respectivamente, para quantificação das regiões nuclear e do citoplasma.

basal) associada aos seus respectivos núcleos.

Para obter padrões de expressão claros, aplicou-se a análise de espectro singular (SSA), uma técnica não paramétrica com um filtro adaptativo. O SSA é um método destinado a realizar a decomposição de uma sequência de medições em uma soma de componentes interpretáveis. Seu uso para extração de sinal a partir dos dados espaciais de expressão gênica unidimensional permitiu remover o ruído decorrente da variabilidade na ordem nuclear e na distribuição citoplasmática dos produtos gênicos (ALEXANDROV; GOLYANDINA; SPIROV, 2008). Durante os testes foi utilizado a rotina em MATLAB de (GOLYANDINA; NEKRUTKIN; ZHIGLJAVSKY, 2001). Então, a partir dos dados de expressão gênica normalizados, curvas SSA foram geradas para cada embrião. O mesmo comprimento de janela ( $L = 40$ ) foi usado para todas as imagens e duas tendências de diferentes resoluções para cada região – nuclear ( $T = 5$ ) e citoplasmática ( $T = 3$ ). Foram realizados testes empíricos para estimar os parâmetros descritos. A partir disso, foi avaliado os perfis de expressão nos núcleos, citoplasma apical e basal, para cada genótipo, embriões selvagens e mutantes. Com o objetivo de padronizar os perfis para permitir as comparações entre selvagens e mutantes, as curvas SSA foram alteradas por meio de três operações: (i) normalização da intensidade do perfil de expressão; (ii) reamostragem da curva para um comprimento comum a fim de padronizar o comprimento dos embriões; e (iii) deslocamento das curvas para centralizar os picos no gráfico e aparar as extremidades em que as curvas são nulas. Na Fig. 63 é ilustrado o resultado desse processo em uma imagem da base TRANSBASE.

Em tal passo, o uso do CP é fundamental para definir um ponto de referência comum na plotagem dos dados entre diferentes embriões. CP foi posicionado no centro do gráfico, portanto, o valor máximo da expressão está localizado nesta região. Esse processo foi desenvolvido para todas as imagens de cada genótipo. CP também pode ser usado para traçar meia curva, que é uma forma tradicional de mostrar o nível de expressão em embriões de *Drosophila* (LOPES et al., 2008; COPPEY et al., 2008).

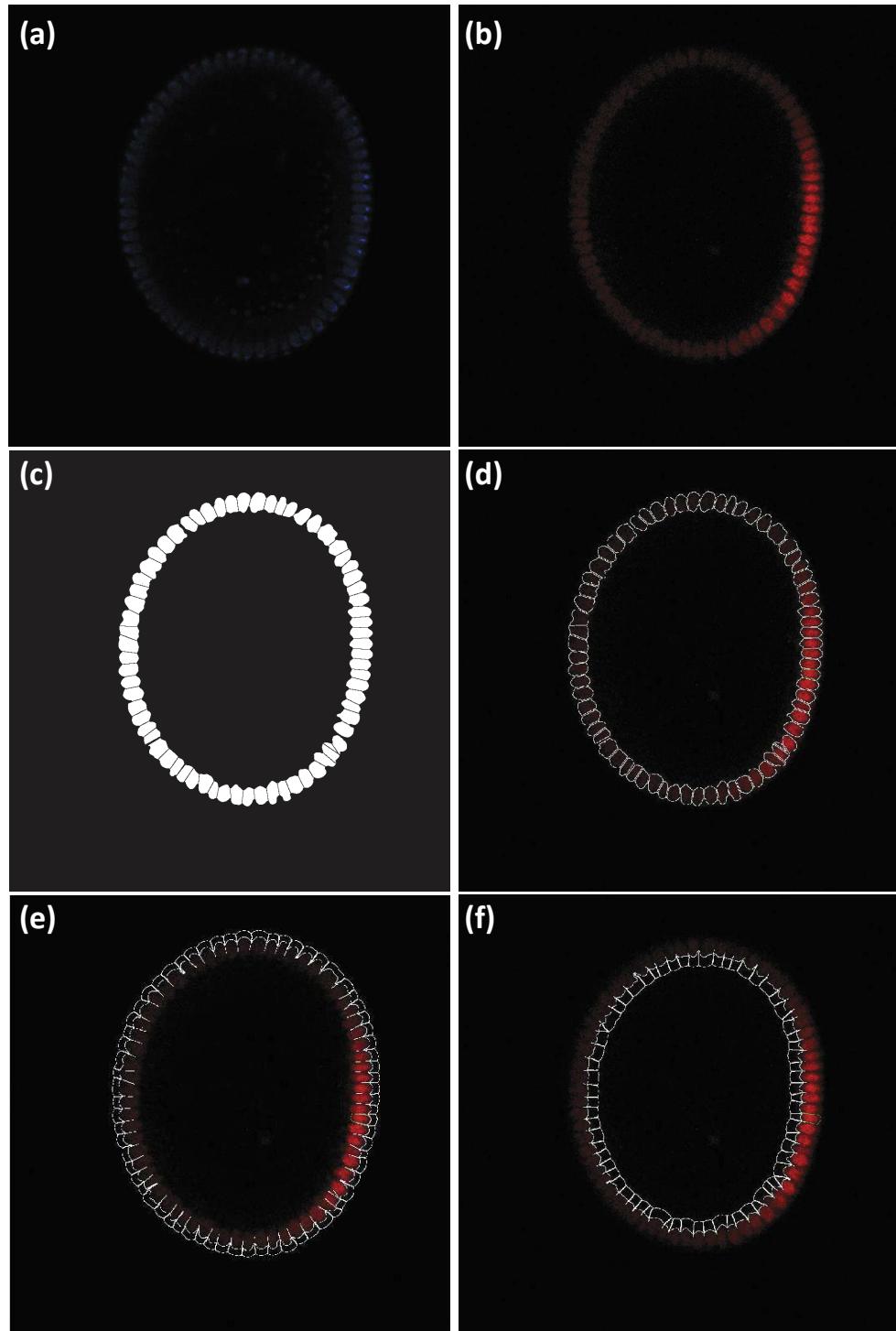


Figura 58 – Sobreposição do contorno das regiões de interesse da seção transversa. (a) Imagem do embrião com marcação DAPI. (b) Imagem do embrião com marcação da proteína. (c) Máscara de segmentação obtida a partir da imagem DAPI. (d)-(e)-(f) Sobreposição das regiões nuclear, apical e basal, respectivamente, sobre a imagem DAPI. Cada contorno branco gera um valor de intensidade média.

As Figuras 64 e 65 exibem o perfil de expressão para os embriões do tipo selvagem e mutante, respectivamente. Nessas imagens, o perfil de expressão foi calculado como

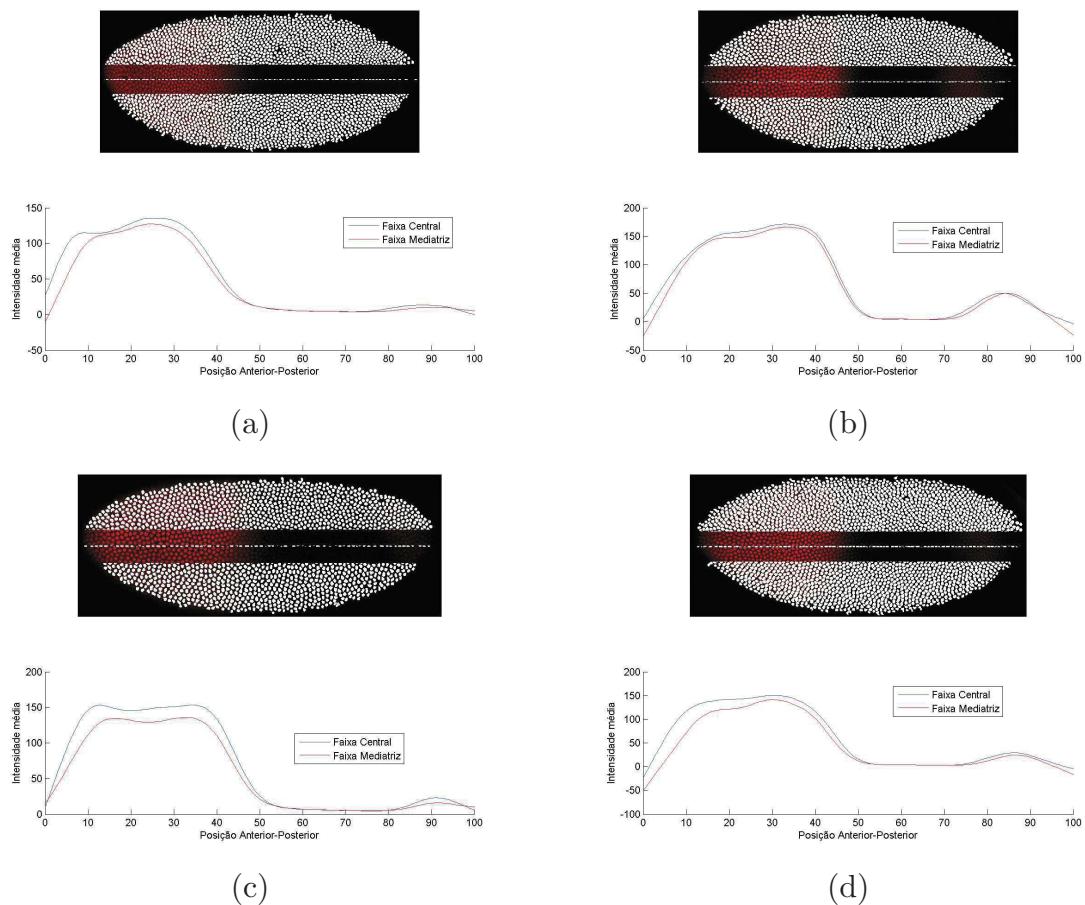


Figura 59 – Perfis de expressão gênica da proteína *Hunchback* relativos às faixas Central e Mediatrix. Exemplos dos experimentos realizados na base SURFBASE.

a média (e desvio padrão) de todos os embriões para cada genótipo (*cact*[A2]/Df(*cact*), *dl*[6]/*cact*[A2] e *dl*[6]/+) e para cada ROI (núcleos, citoplasma apical e citoplasma basal).

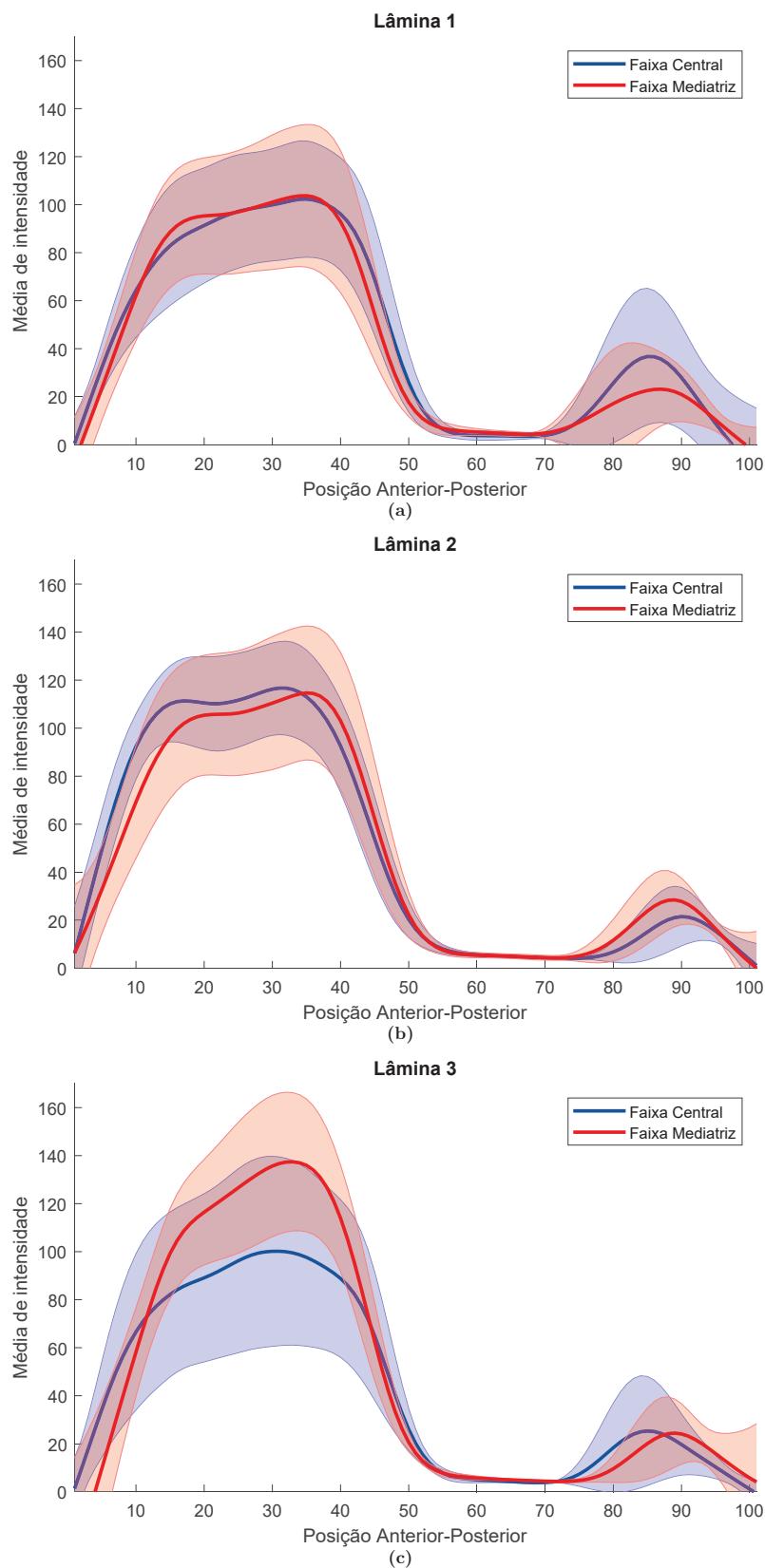


Figura 60 – Comparação gráfica dos sinais de expressão gênica contidos nas faixas central e mediatriz. Sobreposição das curvas médias e de erro relativa a cada região.

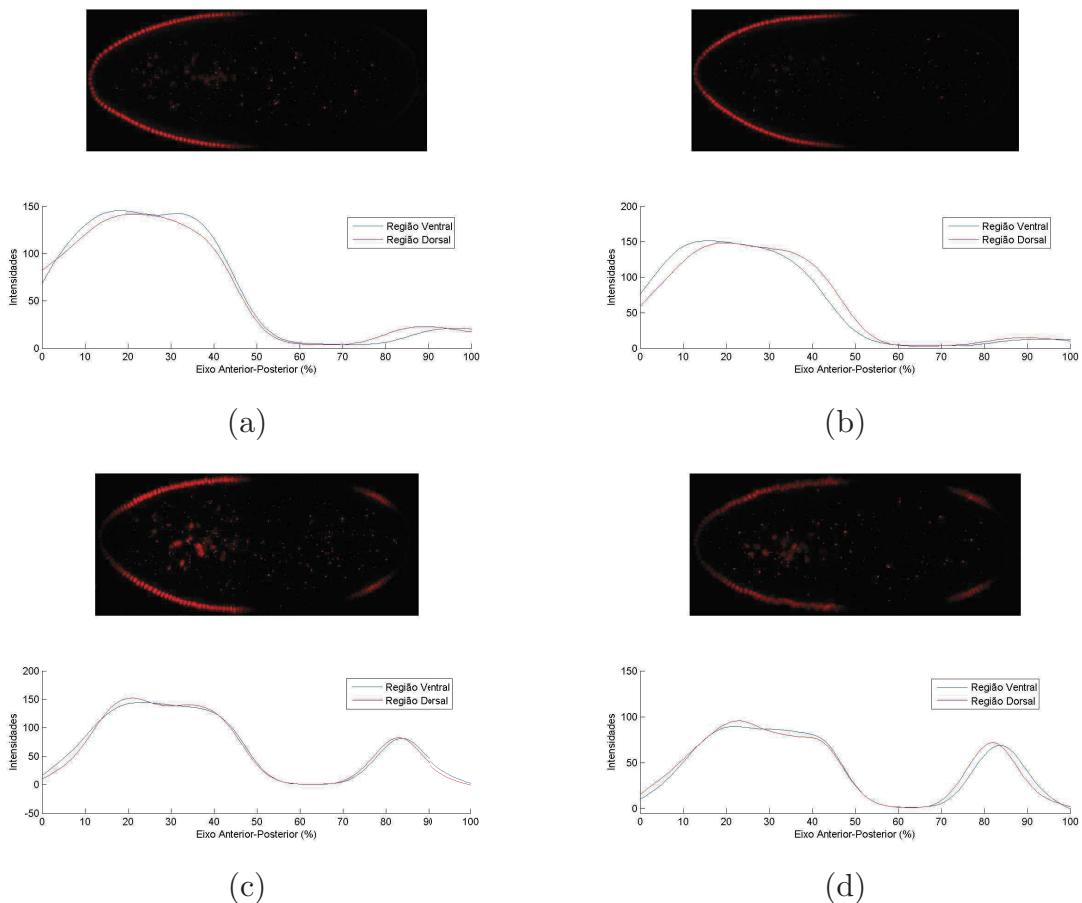


Figura 61 – Perfis de expressão gênica da proteína *Hunchback* expressos na região nuclear. As curvas relativas aos lados dorsal e ventral são sobrepostos. Exemplos dos experimentos realizados na base SAGIBASE.

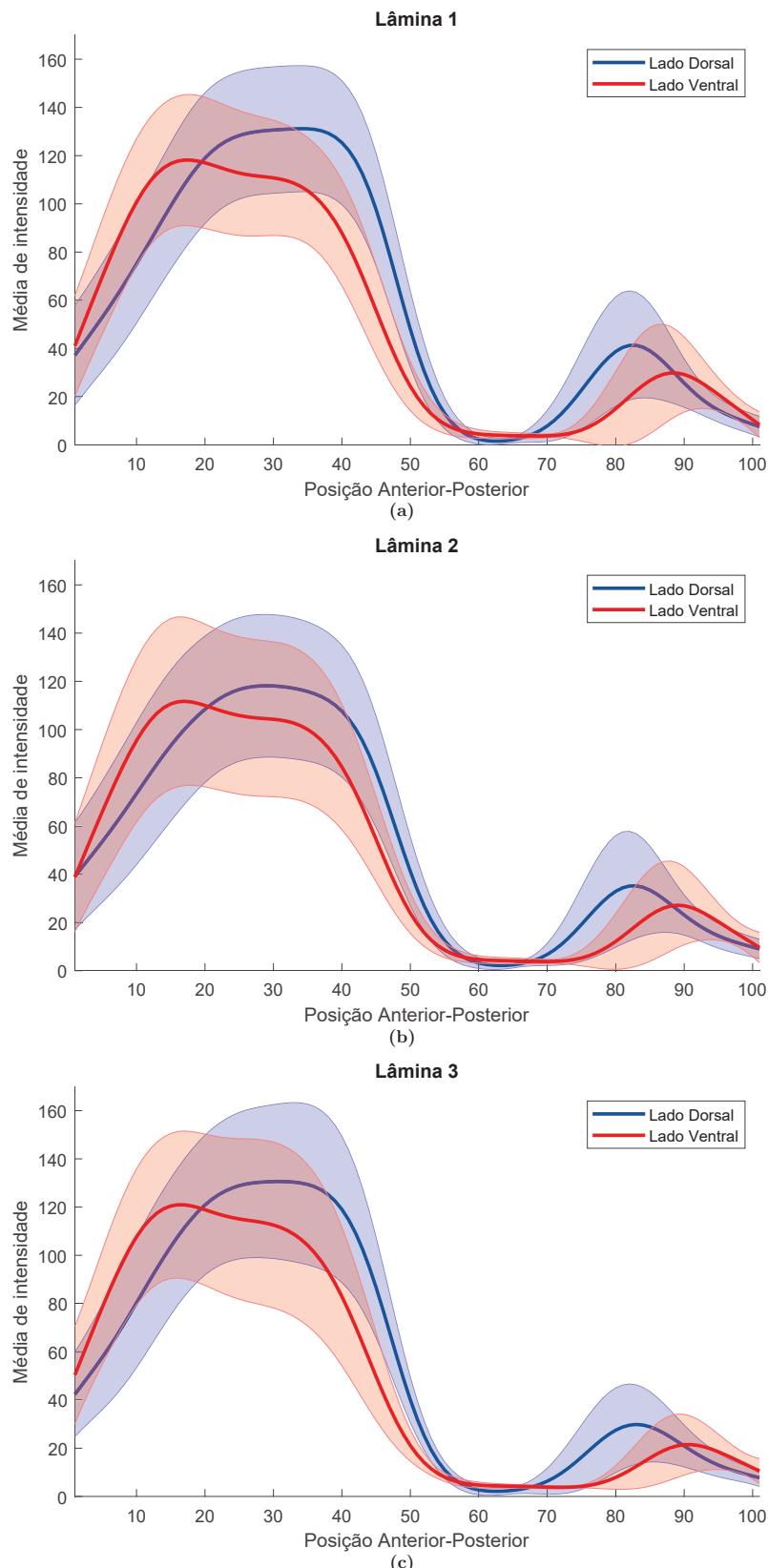


Figura 62 – Comparação gráfica dos sinais de expressão gênica contidos regiões dorsal e ventral. Sobreposição das curvas médias e de erro relativos a cada região.

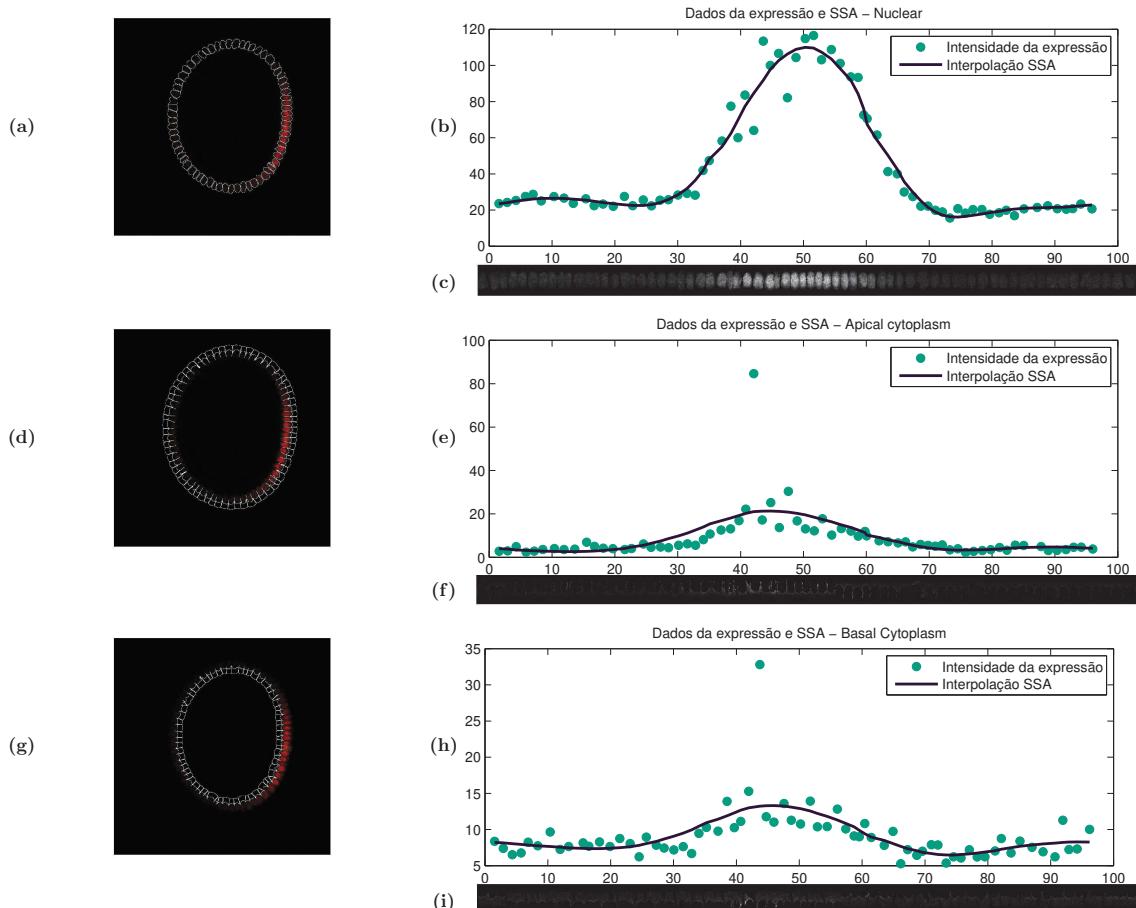


Figura 63 – Resultado individual do perfil de expressão gênica em uma imagem de *Drosophila* na seção transversa. Sobreposição do contorno das máscaras nuclear (a), apical (d) e basal (g) na imagem com sinal de expressão da proteína Dor-sal. Faixa linear horizontal e gráfico do perfil de expressão das regiões nuclear (b-c), apical (e-f) e basal (h-i) ao longo do eixo dorsoventral. Os pontos em (b, e, h) mostram o valor da expressão média de cada ROI (cor verde) e a curva interpolada definida pelo SSA (linha roxa).

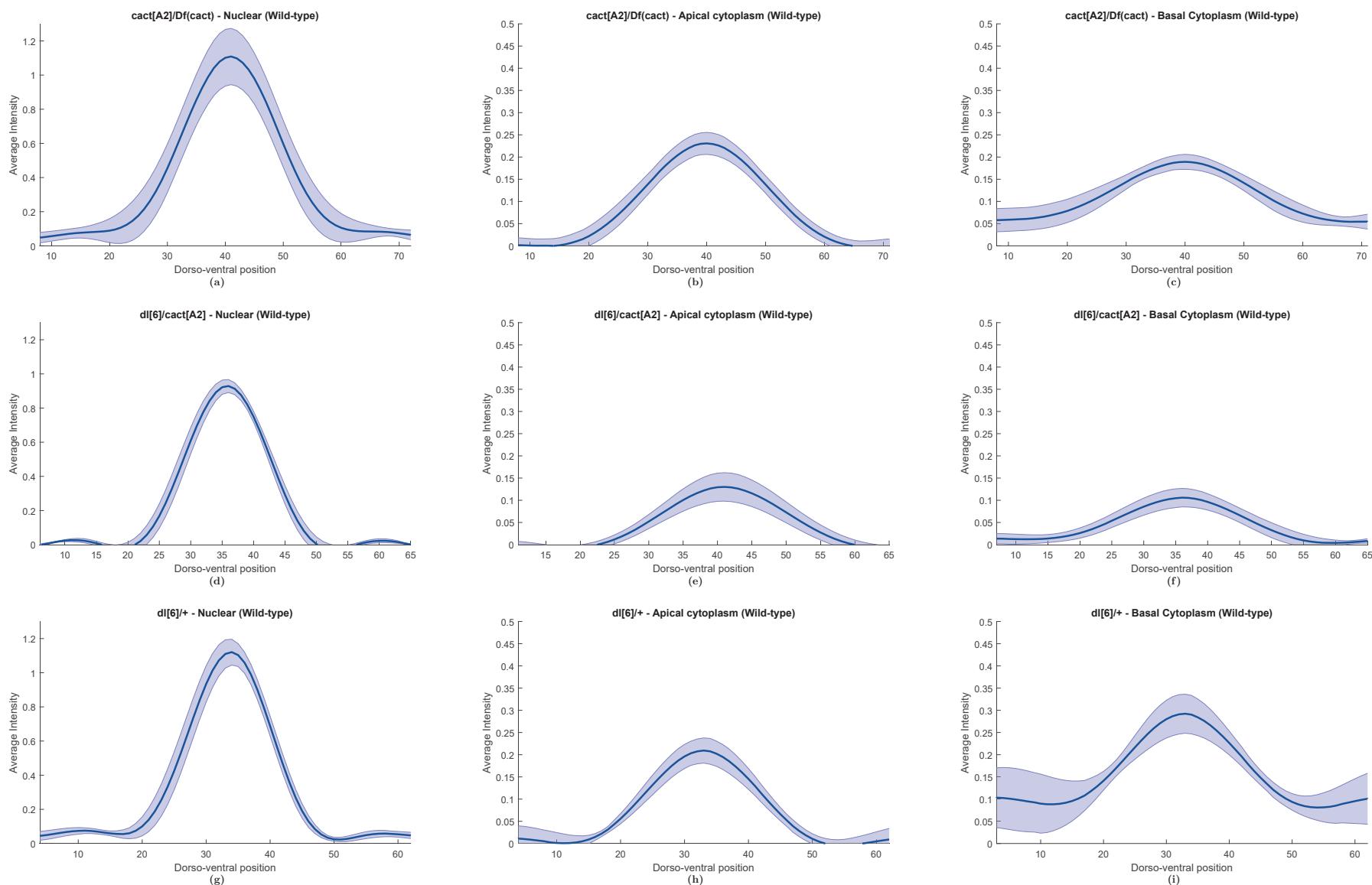


Figura 64 – Perfil da intensidade média das regiões nuclear (primeira coluna), apical (segunda coluna) e basal (terceira coluna) para todos os embriões selvagens, categorizados por seus genótipos. *cact[A2]/Df(cact)* (a–c), *dl[6]/cact[A2]* (d–f) e *dl[6]/+* (g–i).

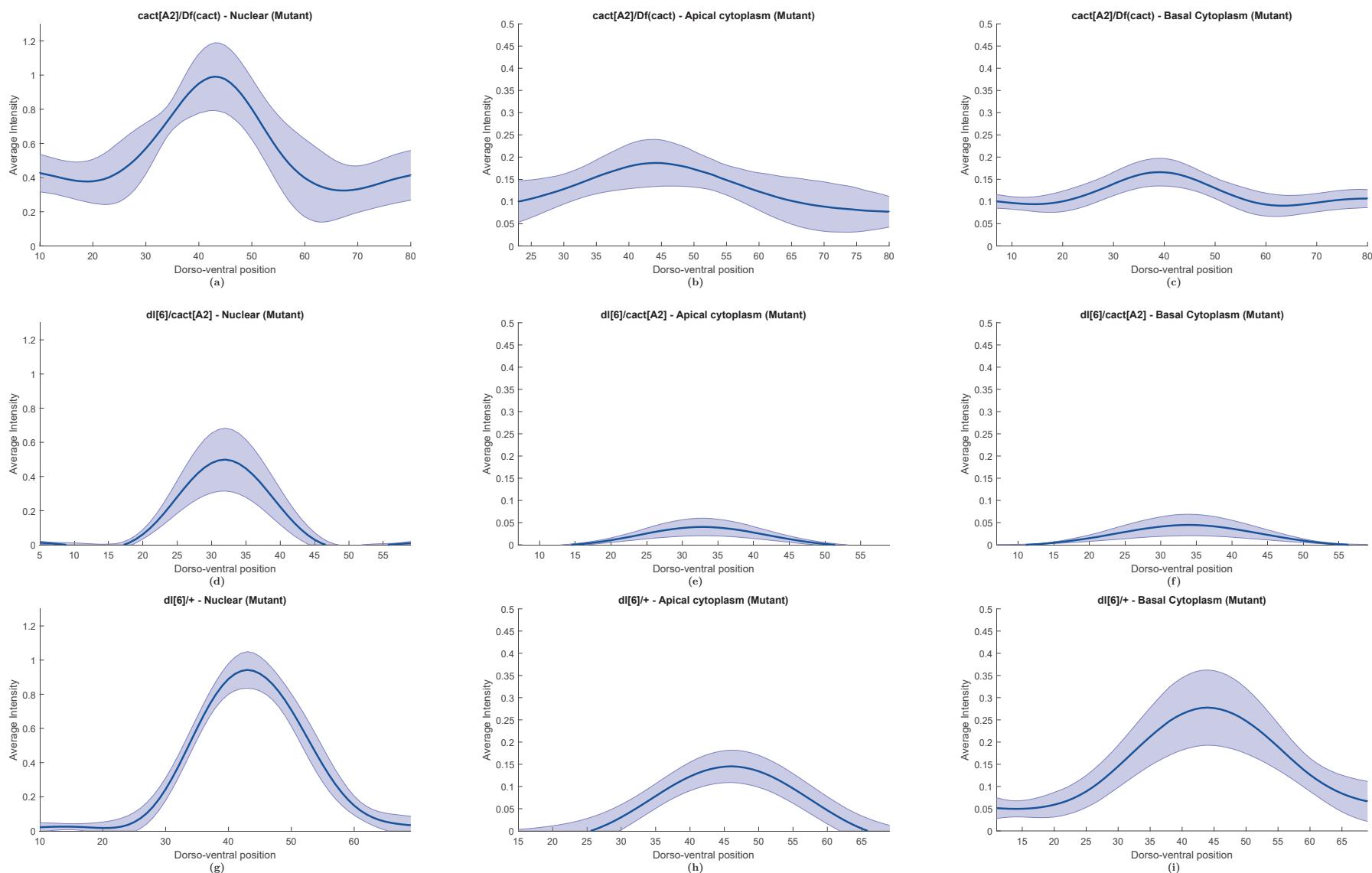


Figura 65 – Perfil da intensidade média das regiões nuclear (primeira coluna), apical (segunda coluna) e basal (terceira coluna) para todos os embriões mutantes, categorizados por seus genótipos. *cact[A2]/Df(cact)* (a–c), *dl[6]/cact[A2]* (d–f) e *dl[6]/+* (g–i).

A partir desses gráficos, observa-se que o padrão de expressão da proteína Dorsal para embriões do tipo selvagem mostra um padrão semelhante a uma curva em forma de sino e com um pequeno desvio padrão. Por outro lado, como esperado, embriões mutantes apresentam variações significativas entre cada genótipo e desvios maiores. Em ambos os conjuntos de dados, encontramos um gradiente de Dorsal no citoplasma (apical e basal) que mimetiza o nuclear. Até o momento, acredita-se que essa é a primeira descrição simultânea dos padrões nuclear e citoplasmático (basal e apical) na literatura, retratando resultados biológicos ainda não observados com o mesmo nível de detalhamento deste trabalho. Essa descrição contribui decisivamente para o entendimento dos mecanismos de translocação nuclear de Dorsal, uma vez que o fluxo contínuo entre o núcleo e o citoplasma é um aspecto fundamental nesse processo. Posteriormente, o gradiente de Dorsal foi comparado entre imagens de embriões selvagens e mutantes, para cada genótipo e região de interesse. Na Fig. 66 é exibido o resultado obtido.

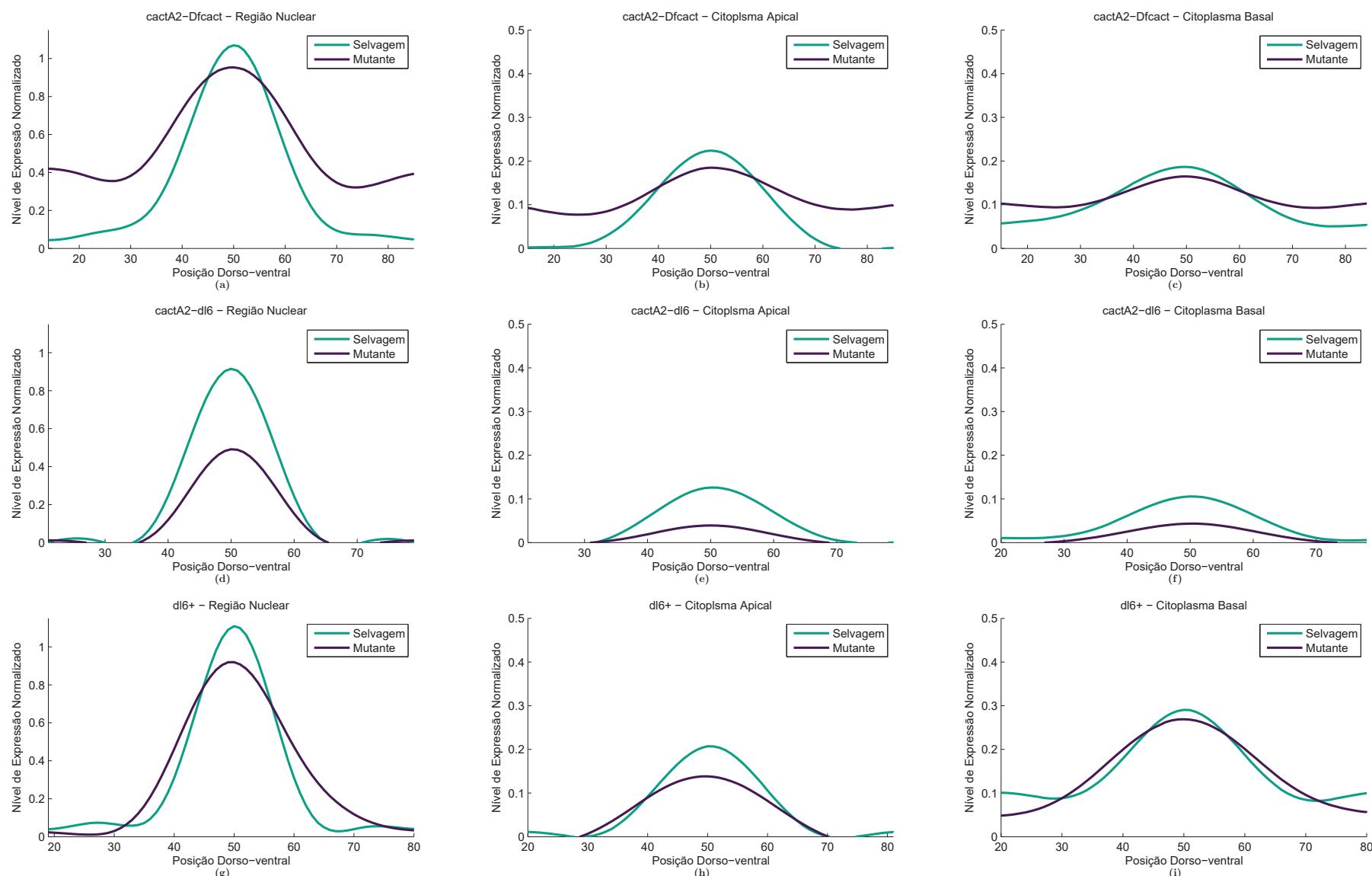


Figura 66 – Comparação entre os perfis de intensidade média das regiões nucleares (primeira coluna), apical (segunda coluna) e basal (terceira coluna) para todas as imagens de embriões mutantes e selvagens, categorizadas pelos genótipos: *cact[A2]/Df(cact)* (a–c), *dl[6]/cact[A2]* (d–f) e *dl[6]/+* (g–i).

Resultados biológicos relevantes foram encontrados nessas comparações. O perfil de expressão na região nuclear para *dl[6]/+* evidenciou que embriões mutantes têm menor expressão da proteína Dorsal na região central da curva (perto de CP), mas ligeiramente mais alto do que o tipo selvagem nos outros núcleos do embrião. A baixa expressão também é encontrada no mutante *dl[6]/cact[A2]*, no entanto, neste caso, o perfil para os outros núcleos também é inferior ao embrião de tipo selvagem. Já o perfil de expressão na região nuclear para *cact[A2]/Df(cact)* evidenciou uma relação inversa do gradiente de Dorsal. Embriões selvagens demonstraram maior concentração de Dorsal na região central da curva (perto de CP) em comparação com embriões mutantes. No entanto, exibiram uma concentração de Dorsal menor na região dorsal do embrião em comparação com embriões mutantes.

Em relação à região citoplasmática, para todos os genótipos, o mesmo padrão encontrado nas regiões nucleares também foi encontrado no citoplasma. Encontramos os níveis de expressão mais elevados de Dorsal no citoplasma basal de *dl[6]/+* e no citoplasma apical de *cact[A2]/Df(cact)*, em ambos os tipos de embriões.

Ao comparar os citoplasmas basal e apical, uma expressão ligeiramente superior foi encontrada no citoplasma basal para embriões de tipo selvagem e mutantes (Fig. 66).

# CAPÍTULO 6

## Conclusão

Nesta tese é proposta uma metodologia computacional para o processamento e análise de imagens de embriões da *Drosophila melanogaster*. A proposta compõe-se de seis módulos, que em conjunto, oferecem uma solução genérica para o tratamento da variabilidade e complexidade dos dados biológicos desse tipo de imagem. O presente trabalho sugere uma abordagem ampla, capaz de contornar os principais obstáculos/desafios normalmente encontrados nesse campo de estudo da Bioinformática.

A metodologia proposta permite realizar tarefas que vão desde a identificação do tipo de embrião até a visualização gráfica dos padrões de expressão gênica. Além de representar uma solução inovadora, os módulos propostos também se destacam individualmente. A metodologia possui flexibilidade de uso e facilidade de adaptação a outros campos de pesquisa. Em geral, a proposta apresentada busca tratar a heterogeneidade e complexidade dos dados normalmente encontrados nesse contexto. Os resultados comprovam que ela é capaz de tratar imagens de diferentes bases de dados e com configurações diversas.

### 6.1 Limitações

Uma clara limitação deste trabalho é o tamanho das bases de dados tratadas e a não existência de bases de imagens já segmentadas (*ground truth*) para uma avaliação quantitativa de parte da metodologia, e, por essa razão, várias análises foram feitas de forma qualitativa. Embora a quantidade de imagens tratadas aqui, inicialmente tenha sido suficiente para os experimentos, uma quantidade maior de amostras consequentemente melhora a amostragem de dados. É preciso considerar que, de um conjunto inicial, nem todas as imagens são adequadas para processamento, haja visto os ruídos de aquisição característicos desse tipo de imagem.

Até o momento, não se conhece um banco de dados público ou imagens *ground truth* referente às imagens das seções transversas e sagital aqui abordadas. Consequentemente, algumas tarefas (como validação das metodologias) são condicionadas à inspeção humana tornando-se limitadas e mais lentas.

Outra limitação do trabalho é o ajuste manual de alguns parâmetros. Embora para a maioria dos casos isso não seja um problema sério, visto que o domínio das imagens analisadas não costuma variar significativamente nas diferentes análises, minimizar o número de parâmetros é uma tarefa traz melhoria aos algoritmos.

## 6.2 Trabalhos Futuros

O estudo da expressão gênica e da morfologia de organismos complexos se moverá cada vez mais para estratégias de análise de imagens baseada em computação, dado a complexidade, a heterogeneidade e a quantidade dos tipos de dados em questão. Nesse sentido, elencou-se os seguintes trabalhos futuros:

- ❑ Adicionar um novo módulo ao conjunto da metodologia proposta referente à padronização da visualização anatômica dos embriões. A visão refere-se ao posicionamento do embrião durante o processo de aquisição da imagem. Ela pode ter dois estados principais, lateral e não lateral (sendo, dorsal ou ventral). A visão mais utilizada em experimentos biológicos é a lateral. Entretanto, no processo de captura das imagens é bastante comum obter amostras com diferentes visões em um mesmo experimento (PENG et al., 2006).
- ❑ Adicionar um novo módulo ao conjunto da metodologia proposta referente ao estudo da organização dos núcleos celulares utilizando a medida de hexagonalidade (TRAVENÇOLO, 2007).
- ❑ Aumentar o espectro de imagens de embriões analisadas.
- ❑ Aprofundar a análise das diferenças entre os padrões de expressão nas faixas central e mediatriz, bem como a diferenciação do lado dorsal e ventral em imagens sagitais.
- ❑ Revisar o algoritmo de segmentação, tarefa importante para superar os desafios remanescentes na máscara de segmentação, como objetos fragmentados ou fundidos.
- ❑ Aprofundar a aplicação dos métodos desenvolvidos em estudos biológicos diversos, e com isso, favorecer novos *insights* no campo da *Drosophila melanogaster*.
- ❑ Aprofundar o estudo de métodos de interpolação como o SSA e aplicá-los em novas análises.
- ❑ Criar um repositório com as rotinas desenvolvidas e disponibilizar para acesso público.

### 6.3 Contribuições bibliográficas

Foram produzidos dois artigos científicos. O primeiro, descreve método *Embrystandard*, para padronização automática dos embriões. Esse trabalho foi publicado no *Journal of Visual Communication and Image Representation*. O segundo artigo descreve o processo completo de obtenção dos padrões de expressão no plano transverso e discute as implicações biológicas dos resultados obtidos. Esse trabalho está sob revisão no periódico *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. Um terceiro trabalho está na fase de escrita e engloba todo a metodologia apresentada nesta tese. A seguir é apresentado a listagem detalhada desses trabalhos:

- SOUSA, D. J. de; CARDOSO, M. A.; BISCH, P. M.; LOPES, F. J. P.; TRAVENÇOLO, B. A. N. (2020). Automated standardization of images of *Drosophila* embryos. *Journal of Visual Communication and Image Representation* (SOUSA et al., 2020).
- SOUSA, D. J. de; CARDOSO, M. A.; ARAUJO, H. M. M.; LOPES, F. J. P.; TRAVENÇOLO, B. A. N. (2020). Computational Image Analysis from the Transverse Plane of *Drosophila* embryos. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* (submetido).
- Descrição completa da metodologia: *An automatic pipeline for extraction of quantitative gene expression from confocal images of Drosophila embryos* (em preparação).

---

## Referências

- AIZENBERG, I.; MYASNIKOVA, E.; SAMSONOVA, M.; REINITZ, J. Temporal classification of Drosophila segmentation gene expression patterns by the multi-valued neural recognition method. **Mathematical Biosciences**, v. 176, n. 1, p. 145–159, 2002. ISSN 0025-5564. Disponível em: <[https://doi.org/10.1016/S0025-5564\(01\)00104-3](https://doi.org/10.1016/S0025-5564(01)00104-3)>.
- ALBERTS, B.; JOHNSON, A.; LEWIS, J.; RAFF, M.; ROBERTS, K.; WALTER, P. **Molecular Biology of the Cell**. 4. ed. [S.l.]: New York: Garland Science, 2002. ISBN 10: 0-8153-3218-1 10: 0-8153-4072-9.
- ALEXANDROV, T.; GOLYANDINA, N.; SPIROV, A. Singular Spectrum Analysis of Gene Expression Profiles of Early Drosophila embryo: Exponential-in-Distance Patterns. **Research Letters in Signal Processing**, v. 2008, 07 2008. Disponível em: <<https://doi.org/10.1155/2008/825758>>.
- ALLENA, R.; AUBRY, D. A novel technique to parametrize shell-like deformations inside biological membranes. **Computational Mechanics**, v. 47, n. 4, p. 409–423, Apr 2011. ISSN 1432-0924. Disponível em: <<https://doi.org/10.1007/s00466-010-0551-8>>.
- AMENTA, N.; BERN, M.; KAMVYSELIS, M. Crust: A new voronoi-based surface reconstruction algorithm. In: **Proceedings of the 25th annual conference on Computer graphics and interactive techniques**. New York, NY, USA: ACM, 1998. (SIGGRAPH '98), p. 415–421. ISBN 0-89791-999-8. Disponível em: <<http://doi.acm.org/10.1145/280814.280947>>.
- ARBEITMAN, M. N.; FURLONG, E. E. M.; IMAM, F.; JOHNSON, E.; NULL, B. H.; BAKER, B. S.; KRASNOW, M. A.; SCOTT, M. P.; DAVIS, R. W.; WHITE, K. P. Gene Expression during the life cycle of Drosophila melanogaster. **Science (New York, N.Y.)**, United States, v. 297, p. 2270–5, Sep 2002. Disponível em: <<https://www.doi.org/10.1126/science.1072152>>.
- ARIAS, A. M. Drosophila melanogaster and the Development of Biology in the 20th century. **Methods in molecular biology (Clifton, N.J.)**, United States, v. 420, p. 1–25, 2008. Disponível em: <[https://doi.org/10.1007/978-1-59745-583-1\\_1](https://doi.org/10.1007/978-1-59745-583-1_1)>.
- BAGATINI, P. B.; SAUR, L.; RODRIGUES, M. F.; BERNARDINO, G. C.; PAIM, M. F.; COELHO, G. P.; SILVA, D. V. d.; OLIVEIRA, R. M. de; SCHIRMER, H.; SOUTO, A. A.; VIANNA, M. R. M. R.; XAVIER, L. L. The role of calcium channel blockers and resveratrol in the prevention of paraquat-induced parkinsonism

- in drosophila melanogaster: a locomotor analysis. **Invertebrate Neuroscience**, Springer-Verlag, v. 11, n. 1, p. 43–51, 2011. ISSN 1354-2516. Disponível em: <<http://dx.doi.org/10.1007/s10158-011-0116-3>>.
- BAKER, M. Cellular imaging: Taking a long, hard look. **Nature**, England, v. 466, p. 1137–40, Aug 2010. Disponível em: <<https://doi.org/10.1038/4661137a>>.
- BESSINGER, Z.; XING, G.; LI, Q. Localization of Drosophila embryos using connected components in scale space. In: **19th IEEE International Conference on Image Processing**. [s.n.], 2012. p. 497–500. ISSN 1522-4880. Disponível em: <<https://doi.org/10.1109/ICIP.2012.6466905>>.
- BOROVEC, J.; KYBIC, J.; NAVA, R. Detection and Localization of Drosophila Egg Chambers in Microscopy Images. In: WANG, Q.; SHI, Y.; SUK, H.-I.; SUZUKI, K. (Ed.). **Machine Learning in Medical Imaging**. Cham: Springer International Publishing, 2017. p. 19–26. ISBN 978-3-319-67389-9. Disponível em: <[https://doi.org/10.1007/978-3-319-67389-9\\_3](https://doi.org/10.1007/978-3-319-67389-9_3)>.
- CANNY, J. A. A Computational Approach to Edge Detection. **IEEE Pattern Analysis and Machine Intelligence**, v. 8, n. 6, p. 679–698, 1986. Disponível em: <<https://www.doi.org/10.1109/TPAMI.1986.4767851>>.
- CHEN, K.; CROWTHER, D. C. Functional genomics in Drosophila models of human disease. **Brief Funct Genomics**, v. 11(5), p. 405–15, 2012. Disponível em: <<https://doi.org/10.1093/bfgp/els038>>.
- CHONG, J.; AMOURDA, C.; SAUNDERS, T. E. Temporal development of Drosophila embryos is highly robust across a wide temperature range. **Journal of The Royal Society Interface**, v. 15, n. 144, p. 20180304, 2018. Disponível em: <<https://doi.org/10.1098/rsif.2018.0304>>.
- CICIN-SAIN, D.; PULIDO, A. H.; CROMBACH, A.; WOTTON, K. R.; JIMENEZ-GURI, E.; TALY, J.-F.; ROMA, G.; JAEGER, J. SuperFly: a comparative database for quantified spatio-temporal gene expression patterns in early dipteran embryos. **Nucleic Acids Research**, England, v. 43, p. D751–5, Jan 2015. Disponível em: <<https://doi.org/10.1093/nar/gku1142>>.
- COPPEY, M.; BOETTIGER, A. N.; BEREZHKOVSII, A. M.; SHVARTSMAN, S. Y. Nuclear trapping shapes the terminal gradient in the Drosophila embryo. **Current Biology**, v. 18, p. 915–9, Jun 2008. Disponível em: <<https://doi.org/10.1016/j.cub.2008.05.034>>.
- CROIX, C. M. S.; SHAND, S. H.; WATKINS, S. C. Confocal microscopy: comparisons, applications, and problems. **BioTechniques**, v. 39, n. 6S, p. S2–S5, 2005. PMID: 20158500. Disponível em: <<https://doi.org/10.2144/000112089>>.
- CROMBACH, A.; WOTTON, K. R.; CICIN-SAIN, D.; ASHYRALIYEV, M.; JAEGER, J. Efficient Reverse-Engineering of a Developmental Gene Regulatory Network. **PLOS Computational Biology**, Public Library of Science, v. 8, n. 7, p. 1–21, 07 2012. Disponível em: <<https://doi.org/10.1371/journal.pcbi.1002589>>.
- FILHO, O. M.; NETO, H. V. **Processamento Digital de Imagens**. [S.l.]: Brasport, 1999. ISBN 9788574520094.

- FOWLKES, C.; HENDRIKS, C. L. L.; KERANEN, S.; WEBER, G.; RUBEL, O.; HUANG, M.-Y.; CHATOOR, S.; DEPACE, A.; SIMIRENKO, L.; HENRIQUEZ, C.; BEATON, A.; WEISZMANN, R.; CELNIKER, S.; HAMANN, B.; KNOWLES, D.; BIGGIN, M.; EISEN, M.; MALIK, J. A Quantitative Spatiotemporal Atlas of Gene Expression in the Drosophila Blastoderm. *Cell*, v. 133, n. 2, p. 364–374, 2008. Disponível em: <<https://doi.org/10.1016/j.cell.2008.01.053>>.
- FRISE, E.; HAMMONDS, A. S.; CELNIKER, S. E. Systematic image-driven analysis of the spatial Drosophila embryonic expression landscape. *Molecular Systems Biology*, v. 6, p. 345, 2009. Disponível em: <<https://doi.org/10.1038/msb.2009.102>>.
- GAO, H.; YAN, L.; CAI, W.; HUANG, H. Anatomical Annotations for Drosophila Gene Expression Patterns via Multi-Dimensional Visual Descriptors Integration: Multi-Dimensional Feature Learning. In: **Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: ACM, 2015. (KDD '15), p. 339–348. ISBN 978-1-4503-3664-2. Disponível em: <<https://doi.org/10.1145/2783258.2783384>>.
- GARCIA, D. A fast all-in-one method for automated post-processing of PIV data. **Experiments in Fluids**, Springer Science and Business Media LLC, v. 50, n. 5, p. 1247–59, Oct. 2010. Disponível em: <<https://doi.org/10.1007/s00348-010-0985-y>>.
- \_\_\_\_\_. Robust smoothing of gridded data in one and higher dimensions with missing values. **Computational Statistics & Data Analysis**, Elsevier BV, v. 54, n. 4, p. 1167–1178, apr 2010. Disponível em: <<https://doi.org/10.1016/j.csda.2009.09.020>>.
- GARCIA, M.; NAHMAD, M.; REEVES, G. T.; STATHOPOULOS, A. Size-dependent regulation of dorsal–ventral patterning in the early Drosophila embryo. **Developmental Biology**, v. 381, n. 1, p. 286 – 299, 2013. ISSN 0012-1606. Disponível em: <<https://doi.org/10.1371/journal.pcbi.1002589>>.
- GARCIA, S. M. L.; FERNÁNDEZ, C. G. **Embriologia**. 3. ed. Porto Alegre: Grupo A - Artmed, 2012. ISBN 9788536326207.
- GARGESHA, M.; YANG, J.; EMDEN, B. V.; PANCHANATHAN, S.; KUMAR, S. Automatic annotation techniques for gene expression images of the fruit fly embryo. In: . [s.n.], 2005. v. 8. Disponível em: <<https://doi.org/10.1117/12.631575>>.
- GILBERT, S. **Developmental Biology**. 7. ed. [S.l.]: Sinauer Associates, 2003.
- GOLYANDINA, N.; NEKRUTKIN, V.; ZHIGLJAVSKY, A. Analysis of Time Series Structure: SSA and Related Techniques. **Monographs on Statistics and Applied Probability**, v. 90, 01 2001. Disponível em: <<https://doi.org/10.1201/9781420035841>>.
- GOLYANDINA, N. E.; HOLLOWAY, D. M.; LOPES, F. J.; SPIROV, A. V.; SPIROVD; SPIROVA, E. N.; USEVICHA, K. D. Measuring gene expression noise in early Drosophila embryos: nucleus-to-nucleus variability. **International Conference on Computational Science**, p. 373–382, 2012. Disponível em: <<https://doi.org/10.1016/j.procs.2012.04.040>>.
- GONZALEZ, R.; WOODS, R. **Processamento de Imagens Digitais**. 3. ed. [S.l.]: Prentice Hall, 2010. ISBN 978-8521202646.

- GRIMM, O.; COPPEY, M.; WIESCHAUS, E. Modelling the Bicoid gradient. **Development**, v. 137, n. 0950-1991, p. 2253–2264, 2010. Disponível em: <<https://doi.org/10.1242/dev.032409>>.
- GRUMBLING, G.; STRELETS, V. Flybase: anatomical data, images and queries. **Nucleic Acids Research**, v. 34, p. D484–8, Jan 2006. Disponível em: <<https://doi.org/10.1093/nar/gkj068>>.
- HAMMONDS, A. S.; BRISTOW, C. A.; FISHER, W. W.; WEISZMANN, R.; WU, S.; HARTENSTEIN, V.; KELLIS, M.; YU, B.; FRISE, E.; CELNIKER, S. E. Spatial expression of transcription factors in *Drosophila* embryonic organ development. **Genome Biology**, England, v. 14, p. R140, Dec 2013. Disponível em: <<https://doi.org/10.1186/gb-2013-14-12-r140>>.
- HAN, D. Comparison of Commonly Used Image Interpolation Methods. In: . Atlantis Press, 2013. p. 1556–1559. ISBN 978-90-78677-61-1. ISSN 1951-6851. Disponível em: <<https://doi.org/10.2991/iccsee.2013.391>>.
- HE, F.; WEN, Y.; CHEUNG, D.; DENG, J.; LU, L. J.; JIAO, R.; MA, J. Distance measurements via the morphogen gradient of bicoid in *Drosophila* embryos. **BMC Developmental Biology**, v. 10, n. 1, p. 80, Aug 2010. ISSN 1471-213X. Disponível em: <<https://doi.org/10.1186/1471-213X-10-80>>.
- HEFFEL, A.; PROHASKA, S. J.; STADLER, P. F.; KAUER, G.; KUSKA, J. P. Automatic Classification of Embryonic Fruit Fly Gene Expression Patterns. In: **Bildverarbeitung fur die Medizin**. Springer, Berlin, Heidelberg, 2009. p. 415–419. Informatik aktuell. Disponível em: <<https://doi.org/10.1007/978-3-540-93860-6\84>>.
- HELPAND, S. L.; ROGINA, B. Genetics of Aging in the Fruit Fly, *Drosophila melanogaster*. **Annual Review of Genetics**, United States, v. 37, p. 329–48, 2003. Disponível em: <<https://doi.org/10.1146/annurev.genet.37.040103.095211>>.
- HENDRIKS, C. L. L.; KERANEN, S. V.; 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. Three-dimensional morphology and gene expression in the *Drosophila* blastoderm at cellular resolution I: data acquisition pipeline. **Genome Biology**, Springer Science and Business Media LLC, v. 7, n. 12, p. R123, 2006. ISSN 1474-760X. Disponível em: <<https://doi.org/10.1186/gb-2006-7-12-r123>>.
- HILLIS, D. M.; SADAUA, D. E.; HILL, R. W.; PRICE, M. V. **Principles of Life**. [S.l.]: Sinauer Associates W.H. Freeman, 2012. ISBN 978-1464109478.
- HIRTH, F. *Drosophila melanogaster* in the Study of Human Neurodegeneration. **CNS Neurol Disord Drug Targets**, v. 9, n. 4, p. 504–523, 2010. Disponível em: <<https://doi.org/10.2174/187152710791556104>>.
- HOPPERT, M.; WREDE, C. Immunolocalization. In: \_\_\_\_\_. **Encyclopedia of Geobiology**. Dordrecht: Springer Netherlands, 2011. p. 482–486. ISBN 978-1-4020-9212-1. Disponível em: <[https://doi.org/10.1007/978-1-4020-9212-1\\_116](https://doi.org/10.1007/978-1-4020-9212-1_116)>.

- HOSKINS, R. A.; CARLSON, J. W.; WAN, K. H.; PARK, S.; MENDEZ, I.; GALLE, S. E.; BOOTH, B. W.; PFEIFFER, B. D.; GEORGE, R. A.; SVIRSKAS, R.; KRZYWINSKI, M.; SCHEIN, J.; ACCARDO, M. C.; DAMIA, E.; MESSINA, G.; MENDEZ-LAGO, M.; PABLOS, B. de; DEMAKOVA, O. V.; ANDREYEVA, E. N.; BOLDYREVA, L. V.; MARRA, M.; CARVALHO, A. B.; DIMITRI, P.; VILLASANTE, A.; ZHIMULEV, I. F.; RUBIN, G. M.; KARPEN, G. H.; CELNIKER, S. E. The Release 6 reference sequence of the *Drosophila melanogaster* genome. **Genome Res.**, v. 25, n. 3, p. 445–458, Mar 2015. Disponível em: <<https://doi.org/10.1101/gr.185579.114>>.
- JAEGER, J.; MANU; REINITZ, J. *Drosophila* blastoderm patterning. **Current Opinion in Genetics & Development**, v. 22, n. 6, p. 533 – 541, 2012. ISSN 0959-437X. Genetics of system biology. Disponível em: <<https://doi.org/10.1016/j.gde.2012.10.005>>.
- JANSSENS, H.; CROMBACH, A.; WOTTON, K. R.; CICIN-SAIN, D.; SURKOVA, S.; LIM, C. L.; SAMSONOVA, M.; AKAM, M.; JAEGER, J. Lack of tailless leads to an increase in expression variability in *Drosophila* embryos. **Developmental Biology**, v. 377, p. 305–17, May 2013. Disponível em: <<https://doi.org/10.1016/j.ydbio.2013.01.010>>.
- JANSSENS, H.; KOSMAN, D.; VANARIO-ALONSO, C.; JAEGER, J.; SAMSONOVA, M.; REINITZ, J. A high-throughput method for quantifying gene expression data from early *Drosophila* embryos. **Development Genes and Evolution**, Springer, v. 215, n. 7, p. 374–381, 2005. Disponível em: <<https://doi.org/10.1007/s00427-005-0484-y>>.
- JANSSENS, H.; SIGGENS, K.; CICIN-SAIN, D.; JIMÉNEZ-GURI, E.; MUSY, M.; AKAM, M.; JAEGER, J. A quantitative atlas of Even-skipped and Hunchback expression in *Clogmia albipunctata* (Diptera: Psychodidae) blastoderm embryos. **EvoDevo**, v. 5, n. 1, p. 1, Jan 2014. ISSN 2041-9139. Disponível em: <<https://doi.org/10.1186/2041-9139-5-1>>.
- JENNINGS, B. H. *Drosophila* – a versatile model in biology & medicine. **Materials Today**, Elsevier BV, v. 14, n. 5, p. 190–195, may 2011. Disponível em: <[https://www.doi.org/10.1016/s1369-7021\(11\)70113-4](https://www.doi.org/10.1016/s1369-7021(11)70113-4)>.
- JI, S.; SUN, L.; JIN, R.; KUMAR, S.; YE, J. Automated annotation of drosophila gene expression patterns using a controlled vocabulary. **Bioinformatics (Oxford, England)**, England, v. 24, p. 1881–8, Sep 2008.
- JIA, D.; XU, Q.; XIE, Q.; MIO, W.; DENG, W.-M. Automatic stage identification of *Drosophila* egg chamber based on DAPI images. **Scientific Reports**, v. 6, 01 2016. Disponível em: <<https://doi.org/10.1038/srep18850>>.
- JOHNSTON, D. S.; NUSSLEIN-VOLHARD, C. The origin of pattern and polarity in the *Drosophila* embryo. **Cell**, United States, v. 68, p. 201–19, Jan 1992. Disponível em: <[https://doi.org/10.1016/0092-8674\(92\)90466-P](https://doi.org/10.1016/0092-8674(92)90466-P)>.
- KANODIA, J. S.; KIM, Y.; TOMER, R.; KHAN, Z.; CHUNG, K.; STOREY, J. D.; LU, H.; KELLER, P. J.; SHVARTSMAN, S. Y. A computational statistics approach for estimating the spatial range of morphogen gradients. **Development**, The Company of Biologists Ltd, v. 138, n. 22, p. 4867–4874, 2011. ISSN 0950-1991. Disponível em: <<https://doi.org/10.1242/dev.071571>>.

- KARAISKOS, N.; WAHLE, P.; ALLES, J.; BOLTENGAGEN, A.; AYOUB, S.; KIPAR, C.; KOCKS, C.; RAJEWSKY, N.; ZINZEN, R. P. The Drosophila embryo at single-cell transcriptome resolution. **Science**, American Association for the Advancement of Science (AAAS), v. 358, n. 6360, p. 194–199, aug 2017. ISSN 0036-8075. Disponível em: <<https://doi.org/10.1126/science.aan3235>>.
- KAZMAR, T.; KVON, E. Z.; STARK, A.; LAMPERT, C. H. Drosophila Embryo Stage Annotation Using Label Propagation. In: **2013 IEEE International Conference on Computer Vision**. [s.n.], 2013. p. 1089–1096. ISSN 1550-5499. Disponível em: <<https://doi.org/10.1109/ICCV.2013.139>>.
- KERANEN, S. V.; FOWLKES, C. C.; HENDRIKS, C. L. L.; SUDAR, D.; KNOWLES, D. W.; MALIK, J.; BIGGIN, M. D. Three-dimensional morphology and gene expression in the Drosophilablastoderm at cellular resolution II: dynamics. **Genome Biology**, Springer Science and Business Media LLC, v. 7, n. 12, p. R124, 2006. R124. Disponível em: <<https://doi.org/10.1186/gb-2006-7-12-r124>>.
- KNOWLES, D. W. Three-dimensional morphology and gene expression mapping for the Drosophila blastoderm. **Cold Spring Harbor Perspectives in Biology Protoc.**, v. 2, p. 150–161, 2012. Disponível em: <<https://doi.org/10.1101/pdb.top067843>>.
- KONIKOFF, C. E.; KARR, T.; MCCUTCHAN, M.; NEWFELD, S. J.; KUMAR, S. Comparison of embryonic expression within multigene families using the FlyExpress discovery platform reveals more spatial than temporal divergence. **Developmental dynamics: an official publication of the American Association of Anatomists**, v. 241, p. 150–60, 01 2012. Disponível em: <<https://doi.org/10.1002/dvdy.22749>>.
- KOSMAN, D.; REINITZ, J.; SHARP, D. H. Automated assay of gene expression at cellular resolution. In: **Proceedings of the 1998 Pacific Symposium on Biocomputing**. [S.l.: s.n.], 1999. p. 6–17.
- KUMAR, S.; JAYARAMAN, K.; PANCHANATHAN, S.; GURUNATHAN, R.; MARTI-SUBIRANA, A.; NEWFELD, S. BEST: A novel computational approach for comparing gene expression patterns from early stages of Drosophila melanogaster development. **Genetics**, v. 162, n. 4, p. 2037–2047, 2002.
- KUMAR, S.; KONIKOFF, C.; EMDEN, B. V.; BUSICK, C.; DAVIS, K. T.; JI, S.; WU, L.-W.; RAMOS, H.; BRODY, T.; PANCHANATHAN, S.; YE, J.; KARR, T. L.; GEROLD, K.; MCCUTCHAN, M.; NEWFELD, S. J. FlyExpress: visual mining of spatiotemporal patterns for genes and publications in Drosophila embryogenesis. **Bioinformatics (Oxford, England)**, England, v. 27, p. 3319–20, Dec 2011. Disponível em: <<https://doi.org/10.1093/bioinformatics/btr567>>.
- KUMAR, S.; KONIKOFF, C.; SANDERFORD, M.; LIU, L.; NEWFELD, S.; YE, J.; KULATHINAL, R. J. Flyexpress 7: An Integrated Discovery Platform To Study Coexpressed Genes Using in Situ Hybridization Images in Drosophila. **G3 (Bethesda, Md.)**, United States, v. 7, p. 2791–2797, Aug 2017. Disponível em: <<https://doi.org/10.1534/g3.117.040345>>.
- LECUYER, E.; YOSHIDA, H.; PARTHASARATHY, N.; ALM, C.; BABAK, T.; CEROVINA, T.; HUGHES, T. R.; TOMANCAK, P.; KRAUSE, H. M. Global analysis of mrna localization reveals a prominent role in organizing cellular architecture and

- function. **Cell**, v. 131, n. 1, p. 174 – 187, 2007. ISSN 0092-8674. Disponível em: <<https://doi.org/10.1016/j.cell.2007.08.003>>.
- LI, Q. A Geometric Framework for Rectangular Shape Detection. **IEEE Transactions on Image Processing**, v. 23, n. 9, p. 4139–4149, Sep. 2014. ISSN 1057-7149. Disponível em: <<https://doi.org/10.1109/TIP.2014.2343456>>.
- LI, Q.; ANANTA, S. S. Localisation of Drosophila embryos using active contours in channel spaces. **International journal of computational biology and drug design**, England, v. 7, p. 157–67, 2014. Disponível em: <<https://doi.org/10.1504/IJCBDD.2014.061644>>.
- LI, Q.; GONG, Y. A geometric method for contour extraction of Drosophila embryos. **BMC Systems Biology**, England, v. 11, p. 102, Dec 2017. Disponível em: <<https://doi.org/10.1186/s12918-017-0478-1>>.
- LI, Q.; KAMBHAMETTU, C. Contour Extraction of Drosophila Embryos. **IEEE/ACM Trans. Comput. Biol. Bioinformatics**, IEEE Computer Society Press, Los Alamitos, CA, USA, v. 8, n. 6, p. 1509–1521, nov. 2011. ISSN 1545-5963. Disponível em: <<https://doi.org/10.1109/TCBB.2011.37>>.
- LI, Q.; SILVEIRA, W.; GONG, Y. Localization of gene expression regions in Drosophila embryonic images. In: **2014 IEEE China Summit International Conference on Signal and Information Processing (ChinaSIP)**. [s.n.], 2014. p. 535–539. Disponível em: <<https://doi.org/10.1109/ChinaSIP.2014.6889300>>.
- LI, Q.; XING, G.; SHINDHELM, A. An automatic framework for standardization of Drosophila embryonic images. In: **15th IEEE International Conference on Image Processing**. [s.n.], 2008. p. 1824–1827. ISSN 1522-4880. Disponível em: <<https://doi.org/10.1109/ICIP.2008.4712132>>.
- LOPES, F.; VIEIRA, F.; HOLLOWAY, D.; BISCH, P.; SPIROV, A. Spatial Bistability Generates Hunchback Expression Sharpness in the Drosophila Embryo. **PLoS Comput Biol**, v. 4, n. 9, p. 1–14, 2008. Disponível em: <<https://doi.org/10.1371/journal.pcbi.1000184>>.
- LYNE, R.; SMITH, R.; RUTHERFORD, K.; WAKELING, M.; VARLEY, A.; GUILLIER, F.; JANSSENS, H.; JI, W.; MCLAREN, P.; NORTH, P.; RANA, D.; RILEY, T.; SULLIVAN, J.; WATKINS, X.; WOODBRIDGE, M.; LILLEY, K.; RUSSELL, S.; ASHBURNER, M.; MIZUGUCHI, K.; MICKLEM, G. FlyMine: an integrated database for Drosophila and Anopheles genomics. **Genome Biology**, v. 8, p. R129, 2007. Disponível em: <<https://doi.org/10.1186/gb-2007-8-7-r129>>.
- MACE, D. L.; VARNADO, N.; ZHANG, W.; FRISE, E.; OHLER, U. Extraction and comparison of gene expression patterns from 2D RNA in situ hybridization images. **Bioinformatics (Oxford, England)**, v. 26, p. 761–9, 11 2010. Disponível em: <<https://doi.org/10.1093/bioinformatics/btp658>>.
- MAZUMDAR, A.; MAZUMDAR, M. How one becomes many: Blastoderm cellularization in Drosophila melanogaster. **BioEssays**, v. 24, n. 11, p. 1012–1022, 2002. Disponível em: <<https://doi.org/10.1002/bies.10184>>.

- MEIJERING, E. Cell Segmentation: 50 Years Down the Road [Life Sciences]. **IEEE Signal Process. Mag.**, v. 29, n. 5, p. 140–145, 2012. Disponível em: <<https://doi.org/10.1109/MSP.2012.2204190>>.
- MICHNO, K. **Drosophila Melanogaster as a Model Organism to Study Human Neurodegenerative Diseases**. University of Toronto, 2009. ISBN 9780494723975. Disponível em: <<http://books.google.com.br/books?id=8IY6MwEACAAJ>>.
- MILLER, J.; CHI, T.; KAPAH, P.; KAHN, A. J.; KIM, M. S.; HIRATA, T.; ROMERO, M. F.; DOW, J. A. T.; STOLLER, M. L. Drosophila melanogaster as an emerging translational model of human nephrolithiasis. **The Journal of urology**, United States, v. 190, p. 1648–56, Nov 2013. Disponível em: <<https://doi.org/10.1016/j.juro.2013.03.010>>.
- MIRZOYAN, Z.; SOLLAZZO, M.; ALLOCÀ, M.; VALENZA, A. M.; GRIFONI, D.; BELLOSTA, P. Drosophila melanogaster: A Model Organism to Study Cancer. **Frontiers in Genetics**, v. 10, p. 51, 2019. ISSN 1664-8021. Disponível em: <<https://doi.org/10.3389/fgene.2019.00051>>.
- MORGAN, D. **The cell cycle : principles of control**. London Sunderland, MA: New Science Press Sinauer Associates, 2007. (Primers in Biology). ISBN 9780199206100.
- NETO, F. G. M.; PAULA, I. C. de. Fractal Dimension-Based Drosophila Melanogaster Wing Genotype Classification. In: **Brazilian Conference on Intelligent Systems (BRACIS)**. [s.n.], 2017. p. 258–263. Disponível em: <<https://doi.org/10.1109/BRACIS.2017.48>>.
- NIZAMI, Z. F.; LIU, J.-L.; GALL, J. G. Fluorescent In Situ Hybridization of Nuclear Bodies in Drosophila melanogaster Ovaries. **Methods in molecular biology (Clifton, N.J.)**, United States, v. 1328, p. 137–49, 2015. Disponível em: <[https://doi.org/10.1007/978-1-4939-2851-4\\_10](https://doi.org/10.1007/978-1-4939-2851-4_10)>.
- NUTAKKI, G. Appearance Based Stage Recognition of Drosophila Embryos. **Masters Theses & Specialist Projects**, 02 2010.
- OCHOA-ESPINOSA, A.; YU, D.; TSIRIGOS, A.; STRUFFI, P.; SMALL, S. Anterior-posterior positional information in the absence of a strong Bicoid gradient. **Proceedings of the National Academy of Sciences**, National Academy of Sciences, v. 106, n. 10, p. 3823–3828, 2009. ISSN 0027-8424. Disponível em: <<https://doi.org/10.1073/pnas.0807878105>>.
- PAN, J.; BALAN, A.; XING, E.; TRAINA, A.; FALOUTSOS, C. Automatic mining of fruit fly embryo images. In: **Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining**. [s.n.], 2006. p. 693–698. Disponível em: <<https://doi.org/10.1145/1150402.1150489>>.
- PANDEY, U. B.; NICHOLS, C. D. Human disease models in Drosophila melanogaster and the role of the fly in therapeutic drug discovery. **Pharmacological reviews**, United States, v. 63, p. 411–36, Jun 2011. Disponível em: <<https://doi.org/10.1124/pr.110.003293>>.
- PAULA, C.; TECHIO, V. Immunolocalization of chromosome-associated proteins in plants - Principles and applications. **Botanical Studies**, v. 55, p. 63, 08 2014.

- PENG, H.; LONG, F.; EISEN, M. B.; MYERS, E. W. Clustering gene expression patterns of fly embryos. In: **3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006.** [s.n.], 2006. p. 1144–1147. ISSN 1945-7928. Disponível em: <<https://doi.org/10.1109/ISBI.2006.1625125>>.
- PENG, H.; MYERS, E. W. Comparing in Situ mRNA Expression Patterns of Drosophila Embryos. In: **Proceedings of the Eighth Annual International Conference on Research in Computational Molecular Biology.** New York, NY, USA: ACM, 2004. (RECOMB '04), p. 157–166. ISBN 1-58113-755-9. Disponível em: <<https://doi.org/10.1145/974614.974636>>.
- PIERRE, S. E. S.; PONTING, L.; STEFANCSIK, R.; MCQUILTON, P. FlyBase 102—advanced approaches to interrogating FlyBase. **Nucleic Acids Research**, Oxford University Press (OUP), v. 42, n. D1, p. D780–D788, nov 2013. Disponível em: <<https://doi.org/10.1093/nar/gkt1092>>.
- PISAREV, A.; POUSTELNIKOVA, E.; SAMSONOVA, M.; REINITZ, J. FlyEx, the quantitative atlas on segmentation gene expression at cellular resolution. **Nucleic Acids Research**, v. 37, n. suppl 1, p. D560–D566, 2009. Disponível em: <<https://doi.org/10.1093/nar/gkn717>>.
- POUSTELNIKOVA, E.; PISAREV, A.; BLAGOV, M.; SAMSONOVA, M.; REINITZ, J. A database for management of gene expression data in situ. **Bioinformatics**, v. 20, n. 14, p. 2212–2221, 2004. Disponível em: <<https://doi.org/10.1093/bioinformatics/bth222>>.
- PRUTEANU-MALINICI, I.; MAJOROS, W. H.; OHLER, U. Automated annotation of gene expression image sequences via non-parametric factor analysis and conditional random fields. **Bioinformatics (Oxford, England)**, v. 29, p. i27–35, Jul 2013. Disponível em: <<https://doi.org/10.1093/bioinformatics/btt206>>.
- PUNIYANI, K.; FALOUTSOS, C.; XING, E. P. SPEX2: automated concise extraction of spatial gene expression patterns from fly embryo ISH images. **Bioinformatics**, Oxford University Press (OUP), v. 26, n. 12, p. i47–i56, jun 2010. Disponível em: <<https://doi.org/10.1093/bioinformatics/btq172>>.
- PUNIYANI, K.; XING, E. P. GINI: From ISH Images to Gene Interaction Networks. **PLOS Computational Biology**, Public Library of Science, v. 9, n. 10, p. 1–15, 10 2013. Disponível em: <<https://doi.org/10.1371/journal.pcbi.1003227>>.
- REEVES, G. T.; TRISNADI, N.; TRUONG, T. V.; NAHMAD, M.; KATZ, S.; STATHOPOULOS, A. Dorsal-Ventral Gene Expression in the Drosophila Embryo Reflects the Dynamics and Precision of the Dorsal Nuclear Gradient. **Developmental cell**, v. 22, n. 3, p. 544–557, fev. 2012. ISSN 1878-1551. Disponível em: <<https://doi.org/10.1016/j.devcel.2011.12.007>>.
- ROBINSON, S. W.; HERZYK, P.; DOW, J. A. T.; LEADER, D. P. FlyAtlas: database of gene expression in the tissues of *Drosophila melanogaster*. **Nucleic Acids Research**, England, v. 41, p. D744–50, Jan 2013. Disponível em: <<https://doi.org/10.1093/nar/gks1141>>.
- ROSENFIELD, A.; PFALTZ, J. L. Distance Functions on Digital Pictures. **Pattern Recognition**, v. 1, p. 33–61, 1968.

- SHEN J.; CASTAN, S. An optimal linear operator for step edge detection. **Computer Vision, Graphical Models and Image Processing**, v. 54, n. 2, p. 112–133, 1992. Disponível em: <[https://doi.org/10.1016/1049-9652\(92\)90060-B](https://doi.org/10.1016/1049-9652(92)90060-B)>.
- SKOBEL, M.; KOWAL, M.; KORBICZ, J.; OBUCHOWICZ, A. Cell nuclei segmentation using marker-controlled watershed and bayesian object recognition. In: PIETKA, E.; BADURA, P.; KAWA, J.; WIECLAWEK, W. (Ed.). **Information Technology in Biomedicine**. Cham: Springer International Publishing, 2019. p. 407–418. ISBN 978-3-319-91211-0. Disponível em: <[https://doi.org/10.1007/978-3-319-91211-0\\_36](https://doi.org/10.1007/978-3-319-91211-0_36)>.
- SOUZA, D. J.; CARDOSO, M. A.; BISCH, P. M.; LOPES, F. J. P.; TRAVENÇOLO, B. A. N. A segmentation method for nuclei identification from sagittal images of *Drosophila melanogaster* embryos. In: **21st WSCG International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision Full Papers Proceedings**. Václav Skala-UNION Agency, 2013. p. 133–142. Disponível em: <<https://otik.uk.zcu.cz/xmlui/handle/11025/10603>>.
- SOUZA, D. J.; LOPES, F. J. P.; TRAVENÇOLO, B. A. N. Methods for nuclear segmentation and evaluation of gene expression in sagittal images of *Drosophila melanogaster* embryos. In: **Workshop of Theses and Dissertation - 27th Conference on Graphics, Patterns and Images (SIBGRAPI)**. [S.l.: s.n.], 2014.
- SOUZA, D. J. de; CARDOSO, M. A.; BISCH, P. M.; LOPES, F. J. P.; TRAVENÇOLO, B. A. N. Automated standardization of images of *Drosophila* embryos. **Journal of Visual Communication and Image Representation**, v. 71, p. 102758, 2020. ISSN 1047-3203. Disponível em: <<https://doi.org/10.1016/j.jvcir.2020.102758>>.
- SPIROV, A. V.; TIMAKIN, D. L.; REINITZ, J.; KOSMAN, D. Using of Evolutionary Computations in Image Processing for Quantitative Atlas of *Drosophila* Genes Expression. In: BOERS, E. J. W. (Ed.). **Applications of Evolutionary Computing**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001. p. 374–383. ISBN 978-3-540-45365-9. Disponível em: <[https://doi.org/10.1007/3-540-45365-2\\_39](https://doi.org/10.1007/3-540-45365-2_39)>.
- STAPLETON, M.; LIAO, G.; BROKSTEIN, P.; HONG, L.; CARNINCI, P.; SHIRAKI, T.; HAYASHIZAKI, Y.; CHAMPE, M.; PACLEB, J.; WAN, K.; YU, C.; CARLSON, J.; GEORGE, R.; CELNIKER, S.; RUBIN, G. M. The *Drosophila* gene collection: identification of putative full-length cDNAs for 70% of *D. melanogaster* genes. **Genome Res**, v. 12, n. 8, p. 1294–1300, Aug 2002. Disponível em: <<https://doi.org/10.1101/gr.269102>>.
- STEPHENSON, R.; METCALFE, N. H. *Drosophila melanogaster*: a fly through its history and current use. **The journal of the Royal College of Physicians of Edinburgh**, Scotland, v. 43, p. 70–5, 2013. Disponível em: <<https://doi.org/10.4997/JRCPE.2013.116>>.
- SURKOVA, S.; MYASNIKOVA, E.; JANSSENS, H.; KOZLOV, K.; SAMSONOVA, A.; REINITZ, J.; SAMSONOVA, M. Pipeline for acquisition of quantitative data on segmentation gene expression from confocal images. **Fly (Austin)**, v. 2, n. 2, p. 58–66, 2008. Disponível em: <<https://doi.org/10.4161/fly.6060>>.
- SURKOVA, S.; MYASNIKOVA, E.; KOZLOV, K. N.; PISAREV, A.; REINITZ, J.; SAMSONOVA, M. Quantitative Imaging of Gene Expression in *Drosophila* Embryos.

**Cold Spring Harb Protoc**, 2013. Disponível em: <<https://doi.org/10.1101/pdb.top075101>>.

SURKOVA, S.; SOKOLKOVA, A.; KOZLOV, K.; NUZHDIN, S. V.; SAMSONOVA, M. Quantitative analysis reveals genotype- and domain- specific differences between mRNA and protein expression of segmentation genes in *Drosophila*. **Developmental Biology**, Elsevier BV, v. 448, n. 1, p. 48–58, apr 2019. ISSN 0012-1606.

SURKOVA, S. Y.; GURSKIY, V. V.; REINITZ, J.; SAMSONOVA, M. G. Study of stability mechanisms of embryonic development in fruit fly *Drosophila*. **Developmental Biology**, v. 42, n. 1, 2011. Disponível em: <<https://doi.org/10.1134/S1062360411010115>>.

TAORMINA, G.; FERRANTE, F.; VIENI, S.; GRASSI, N.; RUSSO, A.; MIRISOLA, M. G. Longevity: Lesson from model organisms. **Genes**, MDPI AG, v. 10, n. 7, p. 518, jul 2019. Disponível em: <<https://doi.org/10.3390/genes10070518>>.

TOLWINSKI, N. S. Introduction: *Drosophila* – A Model System for Developmental Biology. **Journal of Developmental Biology**, v. 5, n. 3, 2017. ISSN 2221-3759. Disponível em: <<https://doi.org/10.3390/jdb5030009>>.

TOMANCAK, P.; BEATON, A.; WEISZMANN, R.; KWAN, E.; SHU, S.; LEWIS, S. E.; RICHARDS, S.; ASHBURNER, M.; HARTENSTEIN, V.; CELNIKER, S. E.; RUBIN, G. M. Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. **Genome Biol**, v. 3, n. 12, p. RESEARCH0088, 2002. Disponível em: <<https://doi.org/10.1186/gb-2002-3-12-research0088>>.

TOMANCAK, P.; BERMAN, B. P.; BEATON, A.; WEISZMANN, R.; KWAN, E.; HARTENSTEIN, V.; CELNIKER, S. E.; RUBIN, G. M. Global analysis of patterns of gene expression during *Drosophila* embryogenesis. **Genome Biol**, v. 8, n. 7, p. R145, 2007. Disponível em: <<https://doi.org/10.1186/gb-2007-8-7-r145>>.

TRAVENÇOLO, B. A. N. **Métodos computacionais para a caracterização e análise da relação entre anatomia e expressão gênica em sistemas biológicos**. Tese (Doutorado) — Programa Interunidades em Bioinformática, Universidade de São Paulo (USP), 2007.

TRISNADI, N.; ALTINOK, A.; STATHOPOULOS, A.; REEVES, G. T. Image analysis and empirical modeling of gene and protein expression. **Methods**, v. 62, n. 1, p. 68 – 78, 2013. ISSN 1046-2023. Modeling Gene Expression. Disponível em: <<https://doi.org/10.1016/j.ymeth.2012.09.016>>.

TURNER, T. L.; STEWART, A. D.; FIELDS, A. T.; RICE, W. R.; TARONE, A. M. Population-Based Resequencing of Experimentally Evolved Populations Reveals the Genetic Basis of Body Size Variation in *Drosophila melanogaster*. **PLOS Genetics**, Public Library of Science, v. 7, n. 3, p. 1–10, 03 2011. Disponível em: <<https://doi.org/10.1371/journal.pgen.1001336>>.

WATSON, J. D.; BAKER, T. A.; BELL, S. P. **Biologia Molecular do Gene**. [S.l.]: Artmed, 2015. ISBN 9788582712085.

- WEBER, G. H.; RUBEL, O.; HUANG, M.-Y.; DEPACE, A. H.; FOWLKES, C. C.; KERANEN, S. V. E.; HENDRIKS, C. L. L.; HAGEN, H.; KNOWLES, D. W.; MALIK, J.; BIGGIN, M. D.; HAMANN, B. Visual Exploration of Three-dimensional Gene Expression Using Physical Views and Linked Abstract Views. **IEEE/ACM Transactions on Computational Biology and Bioinformatics**, United States, v. 6, p. 296–309, Apr-Jun 2009. Disponível em: <<https://doi.org/10.1109/TCBB.2007.70249>>.
- WEISSTEIN, E. W. **Chord**. 2020. Disponível em: <<https://mathworld.wolfram.com/Chord.html>>.
- \_\_\_\_\_. **Eccentricity**. 2020. Disponível em: <<https://mathworld.wolfram.com/Eccentricity.html>>.
- WILK, R.; HU, J.; BLOTSKY, D.; KRAUSE, H. M. Diverse and pervasive subcellular distributions for both coding and long noncoding RNAs. **Genes & Development**, United States, v. 30, p. 594–609, Mar 2016. Disponível em: <<https://doi.org/10.1101/gad.276931.115>>.
- WUNDERLICH, Z.; FOWLKES, C. C.; ECKENRODE, K. B.; BRAGDON, M. D. J.; ABIRI, A.; DEPACE, A. H. Quantitative Comparison of the Anterior-Posterior Patterning System in the Embryos of Five *Drosophila* Species. **G3: Genes, Genomes, Genetics**, G3: Genes, Genomes, Genetics, v. 9, n. 7, p. 2171–2182, 2019. Disponível em: <<https://www.g3journal.org/content/9/7/2171>>.
- YE, J.; CHEN, J.; JANARDAN, R.; KUMAR, S. Developmental Stage Annotation of *Drosophila* Gene Expression Pattern Images via an Entire Solution Path for LDA. **ACM Transactions on Knowledge Discovery from Data**, v. 2 1, 2008. Disponível em: <<https://doi.org/10.1145/1342320.1342324>>.
- YUAN, L.; PAN, C.; JI, S.; MCCUTCHAN, M.; ZHOU, Z.-H.; NEWFELD, S. J.; KUMAR, S.; YE, J. Automated Annotation of Developmental Stages of *Drosophila* Embryos in Images Containing Spatial Patterns of Expression. **Bioinformatics (Oxford, England)**, v. 30, 12 2013. Disponível em: <<https://doi.org/10.1093/bioinformatics/btt648>>.
- ZHOU, J.; SCHOR, I. E.; YAO, V.; THEESFELD, C. L.; MARCO-FERRERES, R.; TADYCH, A.; FURLONG, E. E. M.; TROYANSKAYA, O. G. Accurate genome-wide predictions of spatio-temporal gene expression during embryonic development. **PLOS Genetics**, Public Library of Science, v. 15, n. 9, p. 1–20, 09 2019. Disponível em: <<https://doi.org/10.1371/journal.pgen.1008382>>.