

Ecole Polytechnique

EDX

THÈSE

pour obtenir le titre de

Docteur de l'Ecole Polytechnique

Mention : INFORMATIQUES ET SYSTÈMES COMPLEXES

Présentée et soutenue par

Emmanuel FAURE

Reconstruction automatisée du lignage cellulaire spatio-temporel de l'embryon

Thèse dirigée par Paul BOURGINE

préparée au CREA (Ecole Polytechnique et CNRS)

soutenance prévue le 1 Juillet 2009

<i>Directeur :</i>	Paul BOURGINE	-	CREA - Ecole Polytechnique
<i>Rapporteurs :</i>	Jacques DEMONGEOT	-	TIMC - IMAG
	Hugues BERSINI	-	IRIDIA - ULB (Bruxelles, Belgique)
<i>Examineurs :</i>	Isabelle GUYON	-	Clopinet Consulting (San Francisco, USA)
	Nadine PEYRIÉRAS	-	DEPSN - CNRS
	Guy THERAULAZ	-	CRCA - UPS
<i>Invité :</i>	René DOURSAT	-	ISCPIF - Ecole Polytechnique

Ecole Polytechnique

EDX

THÈSE

pour obtenir le titre de

Docteur de l'Ecole Polytechnique

Mention : INFORMATIQUES ET SYSTÈMES COMPLEXES

Présentée et soutenue par

Emmanuel FAURE

Reconstruction automatisée du lignage cellulaire spatio-temporel de l'embryon

Thèse dirigée par Paul BOURGINE

préparée au CREA (Ecole Polytechnique et CNRS)

soutenance prévue le 1 Juillet 2009

<i>Directeur :</i>	Paul BOURGINE	-	CREA - Ecole Polytechnique
<i>Rapporteurs :</i>	Jacques DEMONGEOT	-	TIMC - IMAG
	Hugues BERSINI	-	IRIDIA - ULB (Bruxelles, Belgique)
<i>Examineurs :</i>	Isabelle GUYON	-	Clopinet Consulting (San Francisco, USA)
	Nadine PEYRIÉRAS	-	DEPSN - CNRS
	Guy THERAULAZ	-	CRCA - UPS
<i>Invité :</i>	René DOURSAT	-	ISCPIF - Ecole Polytechnique

Remerciements

à Romane et Naël.

Je voudrais tout d'abord remercier mon directeur de thèse, **Paul Bourguine**. Non pas par convention ou respect qu'il se doit d'être en premier, mais simplement parce que sans lui cette thèse n'aurait jamais vu le jour et la reconstruction du lignage cellulaire aurait pris encore bien des années Son enthousiasme, sa foi en toutes les sciences, son énergie, son savoir, sa pédagogie, sa patience et sa confiance face à la diversité de chacun, ainsi que sa sagesse mêlée à son éternelle jeunesse sont les atouts d'un directeur de recherche exceptionnel, attentif et passionné. Il a toujours été disponible et demandeur pour d'intenses discussions de fond, de forme, mais aussi du reste. Pour tout cela, sa confiance en mon éternelle autonomie, son soutien financier au cours de ma thèse, et son amitié, je le remercie vivement.

Mon second remerciement va à **Nadine Peyriéras** qui a été au cours de mes années de thèses bien plus que notre expert biologiste, mais m'a guidé, écouté et conseillé. Tout ma connaissance d'embryologiste, vient d'elle, et de sa passion à nous faire partager les problématiques les plus fondamentales, les questions les plus motivantes et les meilleurs glaces. Pour le temps consacré à mes cotés et mon anglais à corriger sans cesse, je la remercie très chaleureusement.

Je tiens à remercier particulièrement Mr **Jacques Demongeot** d'avoir été un excellent rapporteur et Directeur de jury. Je voudrais aussi remercier Mr **Hugues Bersini**, Mr **Guy Theraulaz**, Mme **Isabelle Guyon**, et Mr **René Doursat**, qui m'ont fait le plaisir et l'honneur de faire partie de mon jury de thèse.

Je voudrais remercier L'Association pour la recherche sur le cancer de m'avoir aidé financièrement dans les derniers mois de ma Thèse.

Je voudrais ensuite fortement remercier mes collègues de tous les jours qui ont grandement contribué à l'avancé de cette thèse, tant sur le plan scientifique et technique, qu'humain.

Au sein du Centre de Recherche en Epistémologie Appliquée, **Benoît Lombardot** qui m'a réellement aidé à avoir les pieds sur terre en rendant mon enthousiasme débordant en méthodes concrètes et propres, mais aussi pour son amitié quotidienne ; **Thierry Savy**, pour son génie, et la question qui fâche ; **Raphaël Keller**, un vérificateur de première, aux dernières heures ; et **Jérôme Raciazek**, mon ingénieur privilégié.

A Gif sur Yvette, l'équipe des biologistes du DEPSN : **Louise Duloquin** pour son dynamisme, son amitié et ses oursins ; **Benoît** pour essayer de rester digne entourer de tant de filles ; ainsi qu'**Ingrid, Sophie, Vasily & Julien**.

Je voudrais aussi remercier tout particulièrement mes collaborateurs principaux dans le cadre des projet européens Embryomics et BioEmergences. Ce fut des collègues exceptionnels avec lesquels j'ai créé une famille internationale pluridisciplinaire. Au travers d'un partage scientifique intense, chaque meeting fut un instant d'apogée dans l'échange humain. Ils ont su rester attentif et calme face

à mon insistance de standardisation au sein du workflow et m'ont accueilli plus qu'admirablement dans chacun de leur pays.

Dans un désordre préférentiel,

- en Italie, l'équipe d'**Alexandro Sarti** pour m'avoir ouvert les portes de *Bologna* pendant quelques mois. **Matteo** sa kite surf, son partage culturel, son amitié, sa famille et ses funghis. **Camilo** sa patience, même s'il est trop fort au squash. **Sara** pour m'avoir appris l'italienne, ainsi que **Barbara & Cecilia**.
- en Slovaquie, l'équipe de **Karol Mikula** pour m'avoir expliquer que notre président vient des gipsis ; ainsi que nos fidèles mathématiciens **Robert, Mariana, Olga**.
- en Espagne, l'équipe d'**Andres Santos** , notamment **Miguel Luengo Oroz**, un moteur d'idées, de papiers et de cuba libre.
- en Belgique, **Dominique & Dominique**, pour sa patience et son expertise.
- en France, **Pascal** pour être un *interlocuteur privilégié* à l'IN2P3.

Je tiens aussi à remercier chaleureusement l'ensemble des membres de l'Institut des Systèmes Complexes. Son Directeur, **René** pour avoir accepter et peut-être même apprécier notre *squat*. Les sociologues complexes pour des cafés à nous changer les idées, **Jean-Philippe, David, Camille, Carla** (pour nous avoir organisé un superbe échange sud-américain), ainsi tous les autres Postdocs, Thésards et administratifs de l'ISC, qui ont su créer un lieu d'échange *inter-tout* des plus enrichissants.

Je voudrais exprimer tout ma reconnaissance à l'ensemble du personnel administratif de tous les laboratoires que j'ai traversé, pour leur gentillesse et leur compréhension face à des chercheurs perchés, et plus particulièrement **Geneviève, Stéphanie, Noemi, Marie-Jo, Nadiège** et **Amparo**.

Mes remerciements vont aussi à mes amis, mes proches. Je pense bien-sûr à *l'équipa*, **Platko, Ed, Miguel** et **Ad** pour avoir réussi à lier science et amitié depuis le Master de Sciences Cognitives. Mes coloc préférés, passés, présent et futur, particulièrement **Kirsten** pour son brillant anglais aidant, **Lolo** pour arriver à me sortir le cerveau de la science par la *sérialisation* et tout particulièrement **Karim** pour son soutien exceptionnel le jour J, pour être être le papa scientifique en monde *on*, et sinon en mode *off*.

Je pense aussi évidemment à mes amis de toujours et de Toulouse, **Youri, Khalid, Lolo, Barbouns, Luthor**, ainsi que les **Manu**⁶ et les autres³⁰.

Je voudrais faire un énorme merci à **Mama**, pour son aide précieuse dans la relecture, son soutien moral, sa motivation et sa confiance en moi.

Sans oublier bien-sûr mon origine, je remercie vivement ma mère, ma soeur et mon frère. Mais aussi mon oncle *JJF* pour avoir été ma famille parisienne ainsi que tout ceux qui ont été emporté par le vent.

0.1 Résumé

La reconstruction des dynamiques multi-échelles de la morphogenèse des organismes vivants est devenue un enjeu majeur pour la bio-médecine. L'échelle cellulaire est un niveau d'intégration obligé des dynamiques multi-échelles et il s'agit de concevoir des méthodes automatisées de mesure des comportements cellulaires et de reconstruction du lignage cellulaire. Ce travail de thèse s'inscrit dans la réalisation des objectifs des projets CE Embryomics (NEST adventure n°12916) et BioEmergences (NEST Measuring the impossible N°28809).

Cette thèse est fondée sur une approche théorique de la reconstruction du lignage cellulaire et des morphodynamiques cellulaires au cours de l'embryogenèse précoce du téléostéen *Danio rerio* (*zebrafish*) et de l'oursin *Paracentrotus lividus*. Le concept principal est que toute embryogenèse est parfaitement définie géométriquement par le lignage spatio-temporel des cellules. La reconstruction phénoménologique de ce lignage est donc générique et centrale. Il s'agit, à partir d'imagerie 4D *in vivo* de l'organisme entier en microscopie biphotonique à balayage laser, de mettre en oeuvre les stratégies adéquates de traitement d'images afin d'obtenir les mesures pertinentes pour étudier les comportements cellulaires : position des cellules, lignage, forme du noyau, contour de la membrane cellulaire. Ces mesures sont nécessaires à une description synthétique de l'individu et à la définition d'un prototype pour une classe d'individus.

Nous avons conçu et mis en oeuvre une plate-forme expérimentale pour automatiser l'ensemble des mesures sur des embryons sauvages ou mutants ou traités par des drogues. La question de l'identification des mitoses s'avérant particulièrement difficile, nous avons mis en place un système d'apprentissage actif permettant leur reconnaissance. L'identification précise des mitoses nous a permis d'étudier systématiquement les caractéristiques des divisions cellulaires et de mettre en évidence un biais dans l'orientation des divisions cellulaires au cours de la gastrulation du *zebrafish* : le dipôle des cellules filles tend à être orienté perpendiculairement à la trajectoire de la cellule mère en corrélation avec sa vitesse moyenne avant division. Nous avons finalement établi une stratégie de reconstruction phénoménologique du lignage cellulaire fondée sur des méthodes probabilistes et mettant en oeuvre des algorithmes d'estimation et de maximisation de la vraisemblance.

L'analyse statistique de l'homogénéité du lignage cellulaire spatio-temporel permet de mettre en évidence les champs morphogénétiques de l'embryon. Ces reconstructions permettent de mesurer la variabilité entre différents individus d'une même espèce et d'une même classe correspondant à des conditions génétiques et environnementales similaires. Nous présentons en annexe un modèle computationnel du développement du *zebrafish* au cours des phases précoces de l'embryogenèse, fondé sur l'ensemble des caractéristiques mesurées.

0.2 Abstract

The reconstruction of the multi-scale dynamics of living organisms' morphogenesis is a major challenge in the field of biomedicine. The cellular level of organization is required for the multi scale integration of biological processes, and reconstructing its dynamics should be achieved by designing automated methods for building the cell lineage and measuring cell behaviours.

The thesis is based on a theoretical approach of the reconstruction of the cell lineage and cell morphodynamics during early embryogenesis of the teleost *Danio rerio* (*zebrafish*) and the sea urchin *Paracentrotus lividus*. The main concept is that an embryo is precisely defined geometrically by its spatio-temporal cell lineage. The phenomenological reconstruction of the cell lineage tree is generic and central. It is achieved from 4D *in vivo* imaging of the whole organism by bi-photon laser scanning microscopy with appropriate image processing strategies and provides relevant measurements for further studying cellular behaviours specific features : cell shape, position, neighbourhood, movement, division. Getting measurement for cell behaviour is necessary to provide a summary description of the individual and to define a prototype for a class of individuals. This PhD thesis work is part of the objectives of the EC projects Embryomics (NEST adventure No. 12916) and BioEmergences (NEST Measuring the impossible No. 28809).

We designed and implemented an experimental platform to automate the measurement of cell behaviours in wild type or mutant embryos, either treated or not by drugs of interest. The identification of mitoses proved particularly difficult and we showed that an active machine learning system is a powerful tool for this task. The precise identification of mitoses allowed us to systematically study the characteristics of cell divisions and to identify a bias in the orientation of cell division during zebrafish gastrulation : the dipole made by daughter cells tends to be oriented perpendicular to the trajectory of the mother as a function of its average speed. Finally, we achieved the phenomenological reconstruction of the cell lineage based on probabilistic methods and algorithms for estimating and maximizing the reconstruction likelihood.

The analysis of the statistical homogeneity of the cell lineage in space and time should reveal the morphogenetic fields of the embryo. The phenomenological reconstructions are also the basis for measuring the variability between different individuals of the same class corresponding to similar genetic and environmental conditions. Our work opens the way towards theoretical models and computational simulation of embryogenesis. We show our first attempt to simulate the development of zebrafish during the early stages of embryogenesis, based on the measurements obtained from the phenomenological reconstruction.

0.3 Préambule

Cette thèse a été effectuée au CREA (Centre de Recherche en Épistémologie Appliquée) sous la direction de Paul Bourguine. Elle s'inscrit dans le cadre de deux projets Européens, *Embryomics* : "Measuring the impossible" et *BioEmergences* : "In what" and "how much" are individuals similar and different? Nous visons à reconstruire l'arbre de lignage cellulaire de plusieurs organismes et mesurer les différences et les similitudes entre les individus, cela à différentes échelles. Depuis les données d'imagerie obtenues par acquisition en microscopie multi-photons, jusqu'à la modélisation des comportements cellulaires par l'approche des systèmes complexes, cette thèse se situe dans un cadre intrinsèque d'interdisciplinarité. Nous avons implémenté au sein des grilles de calculs des études mathématiques et physiques automatisées afin de mieux comprendre les phénomènes biologiques sous-jacent à l'embryogenèse. Cette thèse se veut être comprise par toute personne venant d'un de ces domaines et s'intéressant à un autre.

Ce manuscrit¹ se découpe en 6 chapitres et 2 annexes.

- ★ **Chapitre 1 : Introduction** : présentation du contexte, de nos motivations et introduction des notions nécessaires à la compréhension des différentes problématiques.
- ★ **Chapitre 2 : Active Machine Learning for Embryogenesis apply on cell division detection** : mise en place d'un système d'apprentissage actif permettant la détection automatique des différentes phases de la division cellulaire. (Article qui sera soumis dans *Journal of Machine Learning Research*. Une version simplifiée a déjà fait l'objet d'une parution dans [Faure et al. \(2007\)](#)).
- ★ **Chapitre 3 : MorphoTrack** : reconstruction phénoménologique du lignage cellulaire comme processus de branchement spatio-temporel.
- ★ **Chapitre 4 : Cell division orientation in zebrafish brain early embryogenesis** : étude des caractéristiques de la division cellulaire chez le poisson zèbre. (Article qui sera soumis dans *Plos Biology*. Des explications supplémentaires spécifiques à la compréhension générale de la thèse sont données en français).
- ★ **Chapitre 5 : Conclusion , perspective** et présentation d'un modèle du développement du poisson zèbre.
- ★ **Annexe A : Workflow Embryomics / BioEmergences** : développement d'une plate-forme automatisée de reconstruction du lignage cellulaire.
- ★ **Annexe B : Reconstructing the Cell Lineage Tree as the Core of the Embryome** : définition de l'"embryome" d'un organisme comme la description dynamique multi-échelle des stades de développement qui corrélient le génotype et le phénotype. (Article qui sera soumis dans *Developmental Cell*).
- ★ **Annexe C : Modèle du développement embryonnaire** : Bases d'un modèle permettant la simulation du développement du poisson zèbre.

¹Les données étant en 4 dimensions, nous conseillons fortement au lecteur de regarder les vidéos associées.

Table des matières

0.1	Résumé	iii
0.2	Abstract	iv
0.3	Préambule	v
1	Introduction	1
1.1	Motivations	2
1.2	Développement du poisson zèbre	5
1.2.1	Un modèle incontournable	5
1.2.2	Le développement embryonnaire du poisson zèbre	7
1.2.3	Carte des territoires présomptifs	14
1.2.4	Champs Morphogénétiques	15
1.2.5	Importance de l'étude clonale	18
1.3	Division cellulaire et morphogénèse	19
1.3.1	Phases du cycle cellulaire	20
1.3.2	Les divisions cellulaires du poisson zèbre.	23
1.4	Imagerie 4D <i>in toto</i>	25
1.4.1	Microscopie à fluorescence	25
1.4.2	Marqueurs Fluorescents	26
1.4.3	Le microscope confocal à balayage laser : CLSM	27
1.4.4	CLSM multi-photons : MLSM	28
1.4.5	Stratégie d'acquisition par parties	30
1.5	L'approche des systèmes complexes	32
1.5.1	<i>Les systèmes complexes sont</i>	32
1.5.2	L'embryon vu comme un système complexe	33
1.5.3	Modèles mathématiques	37
1.6	Reconstruction automatisée	39
1.6.1	Notion de <i>workflow</i>	39
1.6.2	Un workflow pour la reconstruction de l'embryogenèse	39
1.7	L'apprentissage actif de l'embryogénèse	40
1.7.1	Qu'est-ce que l'apprentissage actif?	40
1.7.2	Les impacts de l'apprentissage actif	42
1.7.3	SVM : Machine à Vecteurs de Support ou Séparateurs à Vaste Marge	43
1.7.4	Apprentissage actif de l'embryon	44
1.7.5	Reconnaissance automatique des cellules en divisions	45
1.8	Reconstruction phénoménologique de l'arbre de lignage cellulaire	47
1.8.1	Suivi des particules par détection	48
1.8.2	Suivi par modèle évolutif	49
1.8.3	Suivi des cellules du poisson zèbre	50

2	Active Machine Learning for Embryogenesis apply on cell division detection	53
2.1	Introduction	54
2.1.1	Cell phase detection	55
2.1.2	Active machine learning apply on complex system	56
2.2	Active machine learning in the Embryomics project	57
2.2.1	Architecture of the automated Embryomics workflow	57
2.2.2	Assisting the Embryomics workflow with active machine learning	57
2.2.3	Annotation tool for supervised training	58
2.3	Active machine learning applied to cell division detection	59
2.3.1	Labels	59
2.3.2	Feature space of individual cells	59
2.3.3	The AML mitotic cell classifier	59
2.4	Results And Analysis	63
2.4.1	Measures	64
2.4.2	Parameters MDB and STM	64
2.4.3	Result depending on each step of the AML	65
2.5	Perspective and Discussion	66
3	MorphoTrack	69
3.1	Concept	70
3.2	Description	70
3.2.1	Problématique	70
3.2.2	Mesure de l'algorithme de <i>tracking</i>	72
3.2.3	Création d'un premier <i>tracking</i> par registration	75
3.3	Définition de <i>MorphoTrack</i>	76
3.3.1	L'algorithme EM	76
3.3.2	Mécanique des fluides complexes	79
3.3.3	Mouvement Brownien et processus stochastique	80
3.3.4	Ensemble statistiquement homogène	81
3.4	M- Maximisation	81
3.4.1	Définition des cellules et de leur <i>features</i>	81
3.4.2	Notion de voisinage cellulaire	82
3.4.3	Définition de la matrice de variance - covariance du mouvement cellulaire	83
3.4.4	Définition de la matrice de déformation du voisinage associée à chaque cellule	83
3.4.5	Détection des frontières	85
3.4.6	Cellules en division	86
3.4.7	Statistiques sur les mesures	88
3.5	E - Estimation	89
3.5.1	f , fonction de <i>tracking</i>	89
3.5.2	Fenêtre spatio-temporelle autour d'une cellule	89

3.5.3	Mesure de probabilités associées à f	90
3.5.4	Probabilité associée à une branche de l'arbre	91
3.5.5	Définition de la loi de probabilité associée	92
3.5.6	Probabilités associées au <i>tracking forward</i>	93
3.5.7	Probabilités associées au <i>tracking backward</i>	97
3.6	Résultats	97
3.7	Le langage MorphoTrack	103
3.8	Interprétation et Conclusion	103
3.9	Démonstration	105
3.10	Symboles	105
4	Cell division orientation in zebrafish brain early embryogenesis : a systematic analysis based on <i>in toto</i> imaging and automated cell tracking	107
4.1	Results	110
4.1.1	Cell behaviours analysis from the phenomenological reconstruction of 4D image data sets	110
4.1.2	Cell morphodynamics characteristics and their correlation with the cell layers topology	111
4.1.3	Quantitative measurements defining cell division characteristic features	111
4.1.4	Cell division plane orientation is biased during gastrulation and correlates with mother cell average speed	112
4.2	Discussion	112
4.3	Figure	113
4.3.1	Figure1	113
4.3.2	Figure2	115
4.3.3	Figure3	117
4.3.4	Figure4	119
4.4	Supplementary Materials	121
4.4.1	The different embryos	121
4.4.2	Setting up a spatial and temporal frame of embryogenesis	121
4.4.3	Measuring angles	126
4.4.4	Study of cell movement relative to the layer	126
4.4.5	Study of neighborhood deformation	127
4.4.6	Cells undergoing a division	127
4.4.7	Model of cellular division	128
4.4.8	Mitosis rates	129
5	Conclusion, Perspectives	131
A	Workflow Embryomics / BioEmergences	137
A.1	Un workflow pour l'embryogenèse	138
A.1.1	Comparaison d'embryons	139
A.1.2	Apprentissage des embryons et généralisation du savoir	139

A.1.3	Quantité de données	139
A.1.4	Plateforme Européenne	139
A.2	Workflow traitement des données	140
A.2.1	Filtrage des Images	140
A.2.2	Détection des centres	142
A.2.3	Mesure multi échelle des mouvements morphogénétiques	142
A.2.4	Segmentation des cellules	145
A.2.5	Détection des divisions cellulaires	148
A.2.6	Reconstruction du lignage cellulaire	149
A.3	Architecture du workflow	149
A.3.1	Implémentation	150
A.3.2	Automatisation	150
A.3.3	Utilisation	151
A.4	Ressources	151
A.4.1	Puissance de calcul	151
A.4.2	Format des données	152
A.4.3	Meta-langage	153
A.4.4	Stockage des données	154
A.4.5	Denali	155
A.5	Workflow automatique	155
A.5.1	Définition de l'expérience	155
A.5.2	Description et exécution des algorithmes	155
A.5.3	Définition des <i>pipelines</i>	158
A.5.4	Gestion des <i>jobs</i> et des erreurs	158
A.6	Conclusion	160
B	Embryomics : Reconstructing the Cell Lineage Tree as the Core of the Embryome	163
B.1	Methods	169
B.1.1	Embryo staining and mounting.	169
B.1.2	Image acquisition.	169
B.1.3	Image filtering.	170
B.1.4	Nuclei centre detection.	170
B.1.5	Segmentation of cellular nuclei and membranes.	170
B.1.6	Cell tracking.	171
B.1.7	Tracking validation.	171
B.1.8	Interactive Embryome Visualisation Platform.	172
B.1.9	Figures	173
B.2	Supplementary material	181
B.2.1	Table S1 Developmental table of the imaged embryo ID 070418a	181
B.2.2	Movies	181

C	Modèle du développement embryonnaire	185
C.1	Concept	185
C.2	Définition d'une cellule	185
C.3	Membrane cellulaire	185
C.4	Stochasticité	186
C.5	Voisinage	186
C.6	Vitellus	186
C.7	Taille des cellules	186
C.8	Couche cellulaire	186
C.9	Mouvement cellulaire	187
C.10	Division cellulaire	187
	Nomenclature	190
	Bibliographie	191

Introduction

Sommaire

1.1 Motivations	2
1.2 Développement du poisson zèbre	5
1.2.1 Un modèle incontournable	5
1.2.2 Le développement embryonnaire du poisson zèbre	7
1.2.3 Carte des territoires présomptifs	14
1.2.4 Champs Morphogénétiques	15
1.2.5 Importance de l'étude clonale	18
1.3 Division cellulaire et morphogenèse	19
1.3.1 Phases du cycle cellulaire	20
1.3.2 Les divisions cellulaires du poisson zèbre.	23
1.4 Imagerie 4D in toto	25
1.4.1 Microscopie à fluorescence	25
1.4.2 Marqueurs Fluorescents	26
1.4.3 Le microscope confocal à balayage laser : CLSM	27
1.4.4 CLSM multi-photons : MLSM	28
1.4.5 Stratégie d'acquisition par parties	30
1.5 L'approche des systèmes complexes	32
1.5.1 <i>Les systèmes complexes sont</i>	32
1.5.2 L'embryon vu comme un système complexe	33
1.5.3 Modèles mathématiques	37
1.6 Reconstruction automatisée	39
1.6.1 Notion de <i>workflow</i>	39
1.6.2 Un workflow pour la reconstruction de l'embryogenèse	39
1.7 L'apprentissage actif de l'embryogenèse	40
1.7.1 Qu'est-ce que l'apprentissage actif?	40
1.7.2 Les impacts de l'apprentissage actif	42
1.7.3 SVM : Machine à Vecteurs de Support ou Séparateurs à Vaste Marge	43
1.7.4 Apprentissage actif de l'embryon	44
1.7.5 Reconnaissance automatique des cellules en divisions	45
1.8 Reconstruction phénoménologique de l'arbre de lignage cellulaire	47
1.8.1 Suivi des particules par détection	48
1.8.2 Suivi par modèle évolutif	49
1.8.3 Suivi des cellules du poisson zèbre	50

1.1 Motivations

Du quartz à l'univers, en passant par l'Homme, l'étude de l'organisation de chacune des échelles spatiales motivera toujours une quête infinie de la nécessité et de la curiosité scientifique. De la même façon, depuis le big-bang jusqu'à la perception futuriste de la science fiction, l'échelle temporelle restera incessamment étudiée en partant de ses extrêmes limites, peut-être infinies, jusqu'à son découpage par plus petit morceau. A travers cette caractérisation spatiale et temporelle, l'étude dissociée de chacun des niveaux de ces échelles a rapidement atteint une limite. Une importante communauté s'intéresse aux principes d'organisation entre ces échelles, et à l'intérieur de chacune d'entre elles, aux interactions *bottom-up* (vers le haut) et *top-down* (vers le bas) qui aboutissent à des *phénomènes émergents*.

La reconstruction et la caractérisation quantitative de l'organogenèse au cours du développement des espèces vivantes est devenue un domaine d'intérêt majeur pour la bio-médecine. Un des principaux challenges de l'ère post-génomique est la compréhension des dynamiques sous-jacentes à l'embryogenèse animale : comment est-ce qu'une séquence linéaire d'ADN traduit un organisme multicellulaire permettant de coder ainsi une immense diversité de cellules réparties dans l'espace et le temps ? Quels sont les rôles des environnements internes et externes impliqués dans les processus de structuration ? Le défi d'intégration de l'ensemble des informations nécessite une approche multi-échelles permettant l'interaction de différents facteurs sur l'ensemble des niveaux : génétique, cellulaire, organique, individuel (et social ?) (Hunter *et al.* (2005)). Malgré les contraintes d'interactions complexes engendrant des différences infinies d'individus, la conception naturelle de l'organisme se fait presque systématiquement de façon uniforme.

La discrétisation biologique de l'ensemble des échelles constituant les organismes vivants, a conduit au cours de l'histoire de la science à une inévitable création de différentes communautés scientifiques distinctes. Parallèlement à ce découpage associé *échelle-science*, les interactions multi-échelles naturelles ont fait émerger des collaborations *polyscientifiques* au sens où leur interdisciplinarité est intrinsèque et endogène, imposée par la nature même des structures étudiées.

Ainsi le vaste champ d'étude permettant la reconstruction et l'analyse de l'embryogenèse, comprend un ensemble de disciplines variées : la biologie, les mathématiques, la physique, l'informatique et l'ingénierie. Ces différents domaines doivent ainsi être combinés et supervisés par une approche épistémique permettant ainsi d'approcher la problématique de manière systématique.

Le développement des techniques d'imagerie permettant l'observation *in-vivo* au niveau cellulaire (Megason & Fraser (2003); Cooper *et al.* (1999); Fernández-González & Muñoz-Barrutia (2006)) est, sans doute, l'une des récentes avancées capitales en biologie du développement. Les techniques de microscopie à balayage laser, combinées avec le développement rapide de marqueurs fluorescents, per-

mettent d'obtenir des films en trois dimensions du développement de l'organisme vivant. Ainsi, il est possible à l'échelle du micro-mètre de mettre en évidence différents types de structures au sein de l'imagerie. Certains marqueurs jouent des rôles prépondérants dans la localisation spatio-temporelle de caractéristiques cellulaires spécifiques comme le noyau ou la membrane plasmique. Cela offre la possibilité d'étudier l'organisme au niveau de la cellule individuelle pendant toute l'embryogenèse animale.

A cette échelle une description du paysage épigénétique (Goldberg *et al.* (2007)) peut être réalisée par le biais d'un arbre de lignage cellulaire. Des mesures descriptives et quantitatives de ces images permettent d'approfondir la caractérisation du développement embryonnaire de l'individu et offre la possibilité de comparer différents organismes.

Dans une première phase, l'énorme quantité de données recueillies par ces techniques d'imagerie doit être traitée de façon robuste et efficace. Le développement d'un ensemble de stratégies, de méthodes et d'algorithmes permettant d'analyser des images en 4 dimensions (3D + temps) de l'embryogenèse est une étape cruciale aboutissant à la reconstruction de l'arbre de lignage cellulaire de l'embryon. Ces méthodes permettent d'identifier des événements spécifiques tels que la mitose ou la mort cellulaire, et d'en mesurer des paramètres tels que le taux de prolifération cellulaire ou la forme de la membrane plasmique.

Ces informations jouent un rôle fondamental dans l'étude des populations des cellules souches, dans la caractérisation du comportement cellulaire en réponse à des drogues et dans la formation de processus cancérogènes. L'assimilation de l'étude de l'oncogenèse à l'embryogenèse repose sur des principes de mécanismes (Amatruda *et al.* (2002)) présentant un grand nombre de caractéristiques similaires, comme l'importance du taux de divisions cellulaires ou du processus d'angiogenèse (Chan *et al.* (2002)).

Le dépistage par des drogues spécifiques, de cellules au comportement spatio-temporel anormal au cours de l'embryogenèse peut servir à identifier des médicaments contre le cancer. En outre, l'étude des processus de l'embryogenèse nous aide à mieux comprendre les principes de dégénérescence, ce qui pourrait nous permettre de les arrêter ou de les inverser (la médecine régénérative).

Les méthodes, les algorithmes et plus généralement, les différentes stratégies envisagées, présentés dans cette thèse seront appliqués aux données du poisson zèbre (*Danio rerio*) à un stade de développement précoce de l'embryon. L'accessibilité, la rapidité du développement et le nombre d'embryons disponibles font du poisson zèbre un modèle de choix. De plus, en raison de sa proximité phylogénétique, il a été largement validé comme un puissant modèle pour les études relatives aux humains dans le traitement contre le cancer.

Aujourd'hui, le modèle du poisson zèbre est un référent en biologie du développement, dans l'étude du cancer et des maladies du comportement des cellules souches. Grâce à la transparence de ses tissus, il est possible d'acquérir *in vivo* et en haute résolution des images de noyaux de cellules, de membranes cellulaires ou de l'expression génétique au cours du développement de l'embryon. De plus,

l'industrie pharmaceutique a augmenté exponentiellement son investissement au cours des dernières années afin de trouver de nouveaux agents thérapeutiques en utilisant le poisson zèbre (Goldsmith (2004)).

Néanmoins, l'énorme quantité de données fournies par ce type d'imagerie, partant de une et pouvant atteindre 50.000 cellules, nécessite, de façon automatisée, le développement de nouvelles stratégies qui permettent non seulement de traiter les données mais également de les interpréter. Pour cette raison, l'objectif de cette thèse est de proposer de telles stratégies quantitatives permettant des mesures numériques automatisées. Une étape importante est de caractériser et de mesurer les similitudes et les différences de développement entre les différents embryons, qu'ils soient sauvages (*wildtype*), mutants ou touchés par la drogue. A plus long terme cette systématisation peut contribuer à la mise en place d'une médecine personnalisée adaptée. Un protocole de traitement pour deux individus, même s'ils sont des clones, peut-être différent car ils ne sont pas complètement identiques en raison de leurs histoires passées, des interactions avec leurs environnements et d'un certain comportement stochastique de leurs processus internes.

À l'avenir, les traitements médicaux seront fonction de certaines prédispositions génétiques et épigénétiques du patient, ce qui permettra d'optimiser les traitements individuels et les plans de prévention dans le cadre de santé publique.

Nous présentons dans ce chapitre introductif le contexte du projet. Nous introduisons au fur et à mesure les problématiques liées à nos études ainsi que les solutions et les méthodes proposées dans les différents chapitres.

Nous expliquons en premier le choix du poisson zèbre comme un modèle incontournable des vertébrés puis introduisons ensuite son développement embryonnaire et les notions nécessaires à sa compréhension et son étude : la carte des territoires présomptifs, les champs morphogénétiques et la notion de clone cellulaire.

Nous nous focalisons ensuite sur la division cellulaire qui est un des éléments clés à la reconstruction du lignage cellulaire. Nous expliquons son importance dans l'étude de la morphogenèse et nous introduisons brièvement les différentes phases du cycle cellulaire. Nous faisons ensuite un état de l'art des différentes études menées sur la compréhension des mécanismes de la division cellulaire du poisson zèbre.

Nous introduisons ensuite les différents éléments permettant de faire l'imagerie en 4 dimensions d'un embryon : la microscopie à fluorescence et les marqueurs fluorescents. Nous expliquons le choix de la stratégie d'acquisition par parties permettant de reconstruire en totalité le lignage cellulaire du poisson zèbre.

Après avoir ainsi introduit les notions nécessaires à la compréhension des choix biologiques effectués pour cette thèse, nous présentons l'approche générale, à travers les *systèmes complexes*, menée sur l'embryogenèse. Nous en profitons pour introduire les différents modèles mathématiques nécessaires à la reconstruction théorique de la morphogenèse d'organismes complexes. Nous définissons ensuite la nécessité d'une automatisation de la reconstruction de l'embryon.

Au vue des différentes caractéristiques composant le développement embryonnaire, nous présentons un mode d'apprentissage actif comme étant au coeur de cette reconstruction et développons son application dans le cadre de la reconnaissance automatique des phases du cycle cellulaire.

Et enfin nous finissons par introduire les différents travaux effectués permettant d'accéder à la reconstruction du lignage cellulaire d'un embryon, à travers le suivi spatio-temporel des cellules.

Tout processus de morphogenèse multicellulaire peut être géométriquement et, au delà, morphodynamiquement représenté comme un lignage cellulaire spatio-temporel. Cette thèse prend tout son sens au chapitre 3 lorsque nous définirons précisément un lignage cellulaire spatiotemporel et les dynamiques cellulaires dans leur espace d'état. De manière intuitive, si l'on définit un lignage cellulaire spatiotemporel comme la donnée à chaque instant - en autre - de la position des cellules et des divisions cellulaires, la caractérisation géométrique à chaque instant est sans ambiguïté ainsi que les vecteurs de vitesses stochastiques et toutes les dynamiques de formes qui s'en suivent.

1.2 Développement du poisson zèbre

1.2.1 Un modèle incontournable

L'étude de cellules en cultures extraites de l'organisme vivant n'est pas suffisante pour comprendre les processus de l'embryogenèse au niveau cellulaire. Il devient donc important d'avoir des modèles référents à la morphogenèse en biologie du développement. Un modèle correspond à un certain groupe d'organismes, et il est choisi pour une étude spécifique. Il convient aussi que les modèles subviennent à un certain nombre de pré-requis pratique : la possibilité de fonctionner en laboratoire, des critères économiques et des critères liés à l'éthique. Les animaux modèles les plus étudiés sont : la souris ; la mouche (*Drosophila melanogaster*), une larve animale appelé nématode (*Caenorhabditis elegans*) ; l'oursin ; un animal marin d'organisation relativement rudimentaire et assez semblables d'apparence aux poissons : l'*Amphioxus*, et le poisson zèbre (*Danio rerio*).

C'est ainsi que des modèles d'invertébrés ont permis d'identifier certains gènes impliqués dans le processus du développement. Leur étude a été nécessaire pour découvrir et comprendre les processus généraux du développement chez leurs homologues vertébrés. Mais il est devenu indispensable de disposer d'un modèle de vertébrés permettant d'étudier certaines fonctions bien moins complexe chez les insectes et les vers que chez l'homme, comme le système vasculaire ou le système nerveux. George Streisinger s'intéressait au bactériophage T4 quand, à la fin des années 60, il décida d'étudier la génétique du développement d'un modèle vertébré. Il choisit alors un poisson bien connu des aquariophiles : le poisson zèbre. Streisinger isola les premiers mutants, en produisant des individus homozygotes

diploïdes par fertilisation d'œufs avec du sperme inactivé aux rayons ultraviolets, permettant ainsi de révéler des mutations récessives (Streisinger *et al.* (1981)).

Aujourd'hui, le poisson zèbre (que nous appelons souvent par anglicisme, le *zebrafish*) est devenu un modèle de la génétique à part entière (Shafizadeh & Paw (2004)). La caractérisation des mutants et de leurs gènes affectés, le développement de nouveaux outils génétiques et les nouvelles techniques d'imagerie ont permis une compréhension plus précise des mécanismes dynamiques précoces impliquant l'organisation de ce vertébré.

Poisson téléostéen de la famille des Cyprinidés (comprenant notamment la carpe et le poisson rouge), le poisson zèbre est originaire de l'Inde et mesure environ 5 centimètres de long. Il s'élevé dans une eau douce, à une température optimale de 28.5°C (à l'état sauvage, on le retrouve principalement dans le Gange). Les femelles peuvent aisément être distinguées des mâles car elles présentent un abdomen rebondi contenant des œufs et possèdent des nageoires verdâtres contrairement aux nageoires orangées des mâles.

Le poisson zèbre est devenu un modèle de choix en raison de ces nombreux avantages. En effet, il présente des caractéristiques spécifiques des modèles invertébrés : les cohortes sont nombreuses, les femelles pondent quelques centaines d'œufs chaque semaine, le développement de l'embryon est externe à l'œuf, ce qui permet une observation plus aisée et une manipulation des embryons facilitée. Le développement de l'embryon est par ailleurs très rapide (environ 3 jours post fécondation). Mais le poisson zèbre présente également l'avantage de posséder le même développement direct que les mammifères, c'est-à-dire sans métamorphose (Metscher & Ahlberg (1999)). L'alevin éclôt à 48 heures post fécondation (hpf) et la résorption de la chorde et du sac vitellin se produit à 20 jours. A 30 jours, il possède la morphologie d'un poisson adulte. Le stade juvénile se poursuit jusqu'à 3 mois, âge auquel la maturité sexuelle est atteinte (Driever *et al.* (1999)). De plus, la transparence du poisson zèbre permet d'observer le développement des organes sur des animaux vivants. La fertilité des téléostéens diminue à l'âge de deux ans et la durée de vie moyenne d'un *Danio rerio* est de trois à quatre ans en laboratoire. Un autre avantage de l'utilisation de ce modèle vient de son appartenance aux vertébrés diploïdes (et non tri ou tétraploïdes, qui rendrait l'analyse génétique difficile). Le programme de séquençage du génome est achevé et la phase d'assemblage du génome est bien amorcée. Celui-ci, réparti en 25 paires de chromosomes, est d'environ 1700 Méga Bases (=10⁶ bases). Toutes ces caractéristiques permettent l'étude de lignées et rendent les études génétiques (notamment la mutagenèse) bien plus aisée que pour d'autres vertébrés. Le poisson zèbre est un modèle devenu incontournable lorsque l'on étudie par exemple le développement de tissus complexes comme le rein ou encore le cœur. En effet, le cœur embryonnaire de poisson zèbre ressemble au cœur humain à 3 semaines de gestation et son rein pronéphrétique est un organe relativement simple qui consiste en 2 néphrons ayant des glomérules fusionnés, et deux tubes pronéphrétiques pairs et bilatéraux. Lors de l'étude de l'hématopoïèse, les poissons zèbres mutants, du à leur croissance exponentielle, montrent des phénotypes étonnamment similaires

à ceux observés dans des maladies humaines comme le cancer (Goessling *et al.* (2007)), et les maladies toxicologiques (Stern & Zon (2003)). Un poisson zèbre sauvage ne produit pas de pigment avant 24hpf (Pickart *et al.* (2004)). On a récemment découvert une lignée transgénique qui ne produit pas du tout de pigment, même chez les adultes. La transparence du corps adulte permet alors la visualisation des cellules souches du sang, des cellules cancéreuses et des métastases. Cela permet de fournir un important support pour l'étude des maladies humaines comme la leucémie (White *et al.* (2008) ; Dooley & Zon (2000)). Il semble donc évidemment que les caractéristiques intrinsèques du poisson zèbre combinées aux technologies récentes d'imagerie, aux outils d'ingénierie en génétique, et à une grande variété de marqueurs fluorescents, fournissent un dispositif très compétitif pour l'étude des vertébrés.

1.2.2 Le développement embryonnaire du poisson zèbre

Kimmel *et al.* (1995) ont établi une liste exhaustive des stades du développement de l'embryon du poisson zèbre. Ils décrivent huit périodes depuis le zygote (0hpf) jusqu'au début de la larve (72hpf) et chaque période est subdivisée en différentes étapes. La description des étapes est basée sur les caractéristiques morphologiques et correspondent à des moments précis et identifiables dans le développement. Nous pouvons voir sur la figure 1.1 une représentation schématique du développement de l'embryon au cours du temps, et le film *Zebrafish* fait par Karlstrom (1996) montre le développement du poisson zèbre vu du plan médio-latéral.

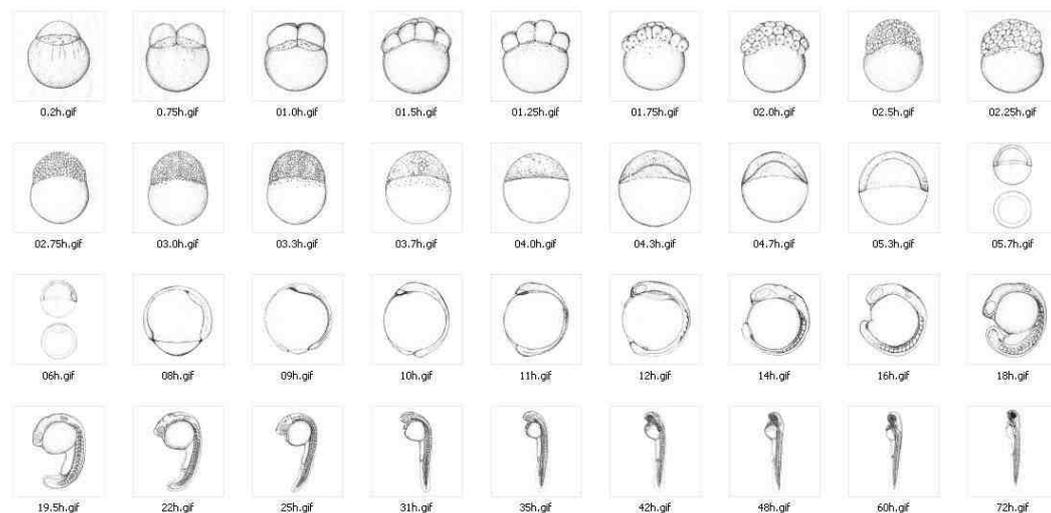


FIG. 1.1 – Représentation schématique du développement de l'embryon d'un poisson zèbre au cours du temps

Au cours de l'embryogenèse, l'embryon récupère ses éléments nutritifs depuis le *vitellus*. C'est un tissu de forme sphérique constituant une réserve de

protéines, semblable au jaune d'œuf. Un des éléments constituant le référentiel du développement est donné par l'axe antéro-postérieur (AP) qui part de la tête de l'embryon jusqu'à sa queue. L'autre axe, dorso-ventral (DV), est formé chez l'embryon adulte. Il apparaît en premier pendant la période de gastrulation. Il passe par l'arrière de la poitrine, dans un plan sagittal et de façon perpendiculaire à l'axe AP (Schier & Talbot (2005)).

Nous allons introduire et illustrer les principales périodes de développement.

1.2.2.1 Zygote (0h)

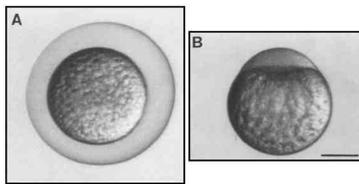


FIG. 1.2 – De Kimmel *et al.* (1995). Période Zygote : A : extrait de son chorion quelques minutes après fertilisation . B : vue du pôle animal environ 10 min après fécondation. Le *yolk* sans cytoplasme a commencé à se séparer de l'animal. Echelle : 250 μm .

La période du zygote correspond à l'œuf tout juste fécondé jusqu'à la première division qui arrive 40 minutes environ après la fécondation. Le zygote a un diamètre d'environ 700 μm .

1.2.2.2 Clivage ($\frac{3}{4}$ - $2\frac{1}{4}h$)

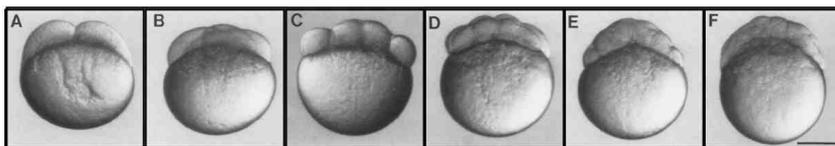


FIG. 1.3 – De Kimmel *et al.* (1995). Période de Clivage : Embryons de face à l'exception de B où il est tourné vers l'axe l'animal-végétal d'environ 45°. A : 2 cellules (0,75 h). B : 4 cellules (1h). C : 8 cellules (1,25 h). D : 16 cellules (1,5 h). E : 32 cellules (1.75 h). F : 64 cellules (2 h). Echelle := 250 μm .

15 minutes environ après la première division, les cellules (blastomères) se divisent à intervalles réguliers. Cette période de six clivages organise l'embryon en motifs réguliers. Au stade de 64 cellules (fin du clivage), certaines cellules sont entièrement recouvertes par d'autres cellules. Les blastomères restent sur la surface de la sphère et forme une couche enveloppante, l'épithélium (EVL).

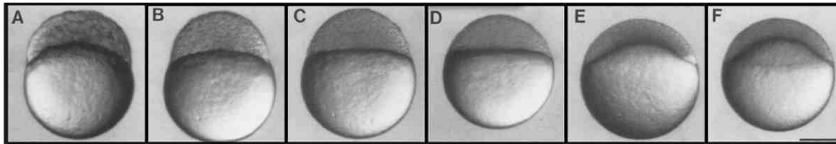
1.2.2.3 *Blastula* ($2\frac{1}{4}$ - $5\frac{1}{4}$ h)

FIG. 1.4 – De Kimmel *et al.* (1995). Période de Blastula : Vues de face. A 256 cellules (2.5 h). B : stade haut (3.3 h). C : transition entre le stade haut et l'oblong (3.5 h). D : transition entre le stade oblong et la sphère (3.8 h). E : stade du dome (4.3 h). F : stade à 30% de l'epibolie (4.7 h). Echelle := 250 μ m.

Cette période part de 128 cellules jusqu'au moment de l'apparition de la gastrulation (14^{ème} cycle de division). Au cours de cette période, les plans de division deviennent de plus en plus indéterminés et les réarrangements cellulaires se font de façon plus irréguliers. Toutes les cellules communiquent entre elles et avec le vitellus par l'intermédiaire de ponts cytoplasmiques : les divisions sont *métasynchrones* et se produisent selon une vague radiale qui part du haut de l'embryon. Vers le 10^{ème} cycle cellulaire, les cellules deviennent mobiles, les cycles s'allongent et marque le début de la transition vers la mi-blastula (T.M.B. pour Transition Mi-Blastuléenne). A ce stade les cellules les plus marginales fusionnent avec le vitellus pour former le Syncytium Vitellin (SV), un tissu extra-embryonnaire spécialisé dans les échanges avec le vitellus. Les mitoses deviennent alors asynchrones, avec un temps moyen de division de plus en plus long, et cela marque l'activation du génome zygotique (Kane & Kimmel (1993)).

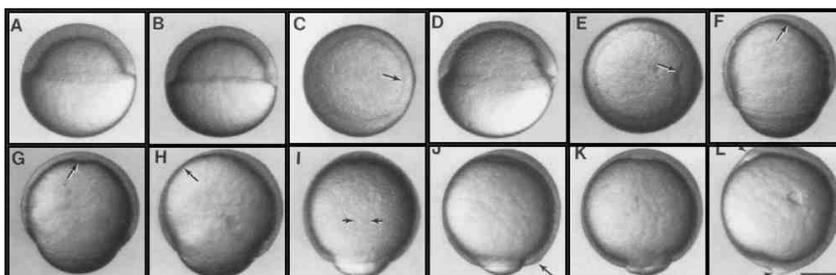
1.2.2.4 *Gastrulation* ($5\frac{1}{4}$ - 10h)

FIG. 1.5 – De Kimmel *et al.* (1995). Période de Gastrulation : Vues du côté gauche :A,B,D,F,G,H,J,L. A : 50% du stade epibolie (5,25 h). B : stade du germe de l'anneau (5,7 h). C : Vue du pôle animal du stade du germe de l'anneau. D : stade shield (6 h). E : vue du pôle animale du shield. F : 70% du stade epibolie (7,7 h). G : 70% du stade epibolie, vue ventrale. H : 75% du stade epibolie (8 h). I : 80% du stade epibolie (8,4 h), vue dorsale. J : 90% du stade epibolie (9 h). K : 90 du stade epibolie, vue ventrale. L : stade du bud (10 h). Echelle := 250 μ m.

La gastrulation est une étape cruciale pour la mise en place des feuillets embryonnaires. Elle est caractérisée par l'apparition de mouvements morphogénétiques complexes (Warga & Kimmel (1990)). Plus précisément, le déclenchement de l'involution des blastomères marginaux à 50% d'épibolie marque le début de cette période. On observe alors un anneau marginal plus épais, formé d'un feuillet de cellules ayant involuées, l'hypoblaste, et d'un feuillet de cellules restant à la surface, l'épiblaste. Après 6 heures de développement, se forme le *shield*, que l'on peut traduire par *écusson embryonnaire* et qui marque le futur axe dorsal de l'embryon. Il est le reflet d'un troisième phénomène appelé la convergence-extension. Les cellules ventrales et latérales des deux feuillets convergent vers le côté dorsal, s'y intercalent, provoquant son épaissement et son extension. Ces trois mouvements, épibolie, involution et convergence-extension, aboutissent à un embryon en fin de gastrulation qui recouvre complètement le vitellus. L'embryon est organisé selon un axe antéro-postérieur concentré du côté dorsal, la future tête étant située au pôle animal et la queue au pôle végétal. Les principaux facteurs sont les interactions entre les tissus, l'adhérence cellulaire et le réarrangement du cytosquelette (Solnica-Krezel (2006), Myers *et al.* (2002)). Durant la gastrulation, on assiste à la détermination des cellules dans un type cellulaire donné en fonction de leur position dans la gastrula. Cela a, en particulier, été montré par des expériences de transplantations (Ho (1992)), et une carte des territoires présomptifs a été détaillée grâce à des expériences de suivi des cellules avec des colorants vitaux (Kimmel *et al.* (1990)). Ainsi, à 50% d'épibolie, on sait que les territoires les plus marginaux, qui formeront l'hypoblaste par involution, donneront naissance au mésoderme et à l'endoderme. L'épiblaste dérivant de la partie animale de l'embryon à 50% d'épibolie, correspond au futur ectoderme qui se différenciera en épiderme du côté ventral et en neurectoderme du côté dorsal. Ainsi, une fois la gastrulation finie, on peut considérer que le plan d'organisation précis du futur poisson est déterminé et à partir de là se met en place la différenciation des organes.

1.2.2.5 *Segmentation* (10 - 24h)

Une riche variété de mouvements morphogénétiques s'effectuent durant la période de segmentation. Autour de 10 hpf, on observe la première séparation entre le premier et le second somite. Les somites continueront à apparaître à raison d'environ un toutes les demi-heures, d'abord au niveau du tronc, puis dans la queue en élongation. Les deux axes AP et DV se discernent sans aucune ambiguïté (Saga & Takeda (2001); Stickney *et al.* (2000)). Le tronc du poisson adulte est composé de blocs de muscles, de cartilage et d'os. Ces blocs, arrangés en série de façon identique aux vertèbres humaines, sont issus des somites qui se forment au cours de l'embryogenèse. Grâce à leur formation précise dans le temps, elles permettent de définir les différents stades de développement pendant la segmentation. Une ébauche d'oeil est visible à 11 hpf, et la vésicule otique apparaît aux alentours de 17 hpf. Pendant que la somitogénèse continue, le tronc commence à se décoller du vitellus et la queue s'étend. A partir de la 14ème étape des somites (16hpf),

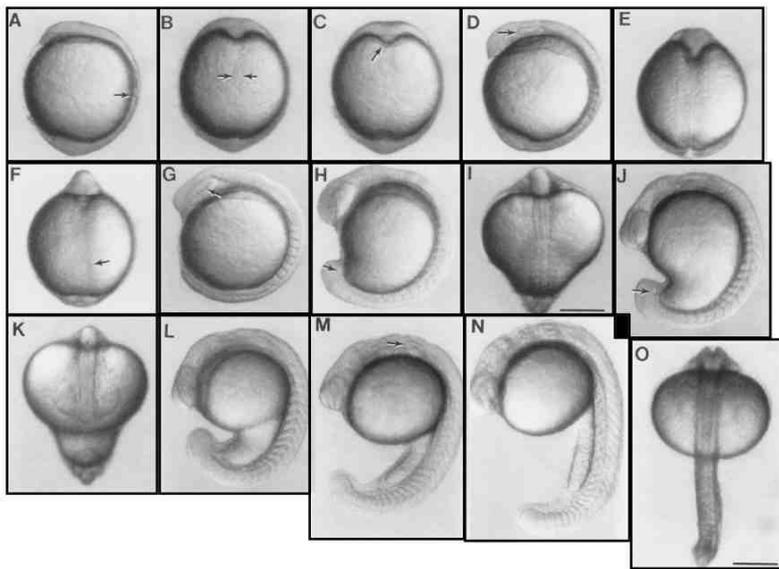


FIG. 1.6 – De *Kimmel et al.* (1995). Période de Gastrulation : Tous sont vues du côté gauche sauf contre-indication. A : stade 2 somites (10,7 h). B : stade 2 somites, vue dorsale. C : stade 2 somites, vue ventrale. D : stade 4 somites (11,3 h). E : stade 4 somites, vue dorsale. F : stade 5 somites (11,7 h), vue ventrale. G : stade 8 somites (13 h). H : stade 13 somites (15,5 h). I : stade 14 somites (16 h), vue dorsale. J : stade 15 somites (16,5 h). K : stade 15 somites, vue dorsale. L : stade 17 somites (17,5 h). M : stade 20 somites (19 h). N : stade 25 somites (21,5 h). O : stade 25 somites, vue dorsale. Echelle := 250 μm .

la queue et la tête d'embryon se redressent complètement, ce qui fait croître la longueur de l'embryon à un taux de $125 \mu\text{m}$ par heure. Au cours de la période de segmentation, se produit également une partie importante du développement morphologique du système nerveux.

1.2.2.6 *Pharyngula* (24 - 48h)

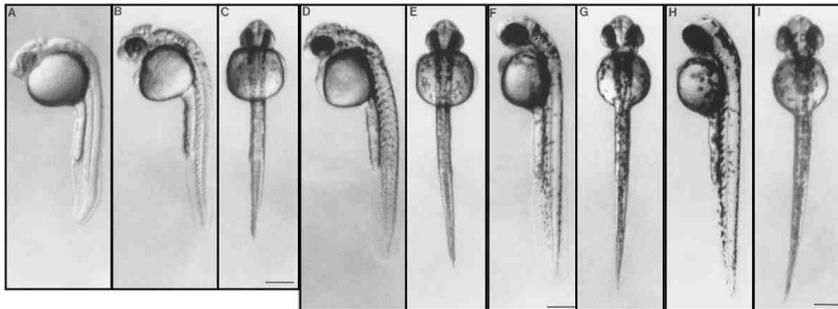


FIG. 1.7 – De Kimmel *et al.* (1995). Période de Pharyngula : Sur le même embryon, à chaque étape vue du côté gauche puis vue dorsale (à l'exception de A) A : vue de côté gauche au stade prim-5 (24 h). B, C : stade prim-12 (28 h). D, E : stade prim-20 (33 h). F, G : stade prim-25 (36 h). H, I : stade haute-pec (42h). Echelle := $250 \mu\text{m}$.

Cette période est conservée dans l'évolution de la plupart des des embryons vertébrés similaires. L'ensemble des somites étant bien formé ainsi que le système nerveux, le poisson zèbre est déjà à ce stade un animal bien défini. Le cœur, divisé en quatre parties, commence à battre, et le sang commence à s'écouler. Au cours de cette période, les cellules pigmentaires du poisson zèbre se différencient et commencent à perdre leur transparence. La figure 1.10 représente une variété de pigmentation.

1.2.2.7 *Eclosion* (48 - 72h)

L'embryon continue de croître et la morphogenèse de nombreux organes, pour la plupart entièrement terminée, se ralentit. Certains derniers organes finissent de se développer comme l'intestin, les nageoires pectorales, les branchies et les mâchoires. La morphogenèse de la mâchoire migre de la bouche vers une position antérieure.

1.2.2.8 *Larve*

L'ensemble de ces étapes conduit à l'obtention d'une larve qui éclot entre le deuxième et le troisième jour, qui équivaut plus ou moins à un adulte en miniature, doté d'organes sensitifs, de muscles et de nageoires lui permettant de nager et de se nourrir.

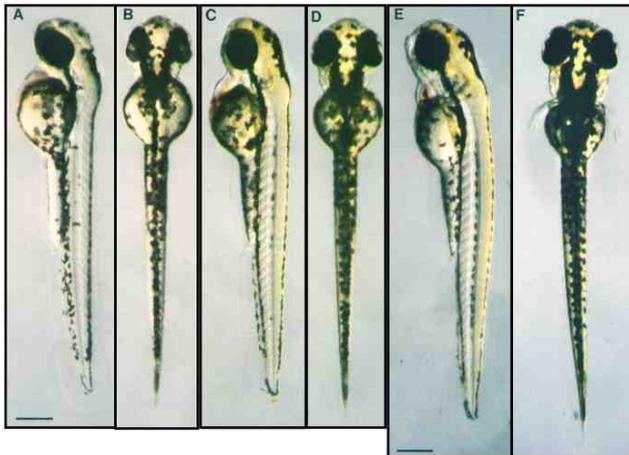


FIG. 1.8 – De Kimmel *et al.* (1995). Période d'éclosion : Vue du côté gauche puis vue dorsale d'un même embryon A, B : stade long-pec (48 h). C, D : stade pec-fin (60 h). E, F : stade saillie-bouche (72 h). Echelle := 250 μm .

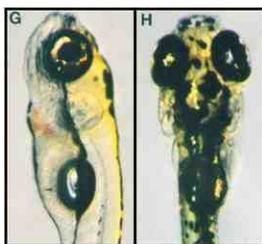


FIG. 1.9 – De Kimmel *et al.* (1995). Période de larve : Vue du côté gauche puis vue dorsale d'un même embryon. G, H : le début de la larve (120 h). Echelle := 250 μm .



FIG. 1.10 – Différence de pigmentation dans la famille du poisson zèbre à l'âge adulte

1.2.3 Carte des territoires présomptifs

Pendant le développement des embryons vertébrés, la descendance d'une cellule peut contribuer à un ou plusieurs organes. Même si la reconstruction totale, ni même partielle, de l'arbre de lignage cellulaire n'a encore pu être faite, plusieurs expériences ont déjà été effectuées pour approximer une carte des territoires présomptifs (en anglais *fatemap*). A un instant donné dans le développement, un marqueur colorant fluorescent, persistant au cours des générations, est injecté dans certaines cellules. L'enregistrement du sous arbre *illuminé* permet de voir le devenir de leurs descendances au cours du temps. Ce type de protocole peut prendre beaucoup de temps et dépend de la proportion de l'échantillon de cellules injectées, en fonction de la région à observer. Néanmoins cela permet de comprendre certains comportements qualitatifs. Un clone dans ce contexte représente tous les descendants d'une seule cellule à partir d'un temps donné. Dans le cadre d'étude de l'organogenèse, se pose alors une question fondamentale : *Est-ce qu'une seule cellule placée à un instant précis dans le temps et dans l'espace génère à elle seule la totalité de l'organe ?* On doit donc pouvoir distinguer les organes d'origine monoclonale de ceux d'origine polyclonale produits par la lignée des multiples cellules.

Kimmel a établi en 1990 (Kimmel *et al.* (1990)) la première carte des territoires présomptifs, depuis le stade tardif de blastula jusqu'à la gastrulation. Il a démontré que les cellules de l'EVL sont presque limitées au périoderme et qu'une couche profonde de cellules génère et organise la carte avec des limites assez floues.

Une autre étude (Strehlow *et al.* (1994)) établit la carte depuis le stade de 16 cellules. Chacun des 16 blastomères a été suivi à maintes reprises jusqu'au stade de 26hpf permettant ainsi de mesurer le nombre de cellules de chaque clone, contribuant à 32 tissus différents. Aucun blastomère n'a contribué à un seul et unique organe et plus de la moitié des blastomères ont contribué à plus de la moitié des tissus. Les cartes attribuées à chacun d'entre eux ne sont ni similaires ni symétriques, ce qui montre l'importance de la position spatiale.

Dans Woo & Fraser (1995), une carte du système nerveux antérieur du poisson zèbre est illustrée (Figure 1.11) depuis 6hpf jusqu'à la fin de la période de gastrulation (10hpf). Malgré le déplacement et le réaménagement des cellules, la carte obtenue est ordonnée et cohérente. Seules les cellules qui se trouvaient dans des régions frontalières à 6hpf ont deux filles dans différentes régions. Il est intéressant de constater que le système neural et la rétine appartiennent à un domaine commun contiguë. Durant la morphogenèse, ce domaine semble se séparer en deux régions sous l'action des progéniteurs diencéphaliques. Par conséquent, les mutations engendrant un poisson zèbre cyclope pourraient jouer le rôle d'empêcher les mouvements de ces progéniteurs.

Concernant la formation des yeux, on trouve dans England *et al.* (2006) une carte plus détaillée qui comprend la dynamique des réarrangements cellulaires depuis 8hpf jusqu'à 15hpf. Ils utilisent les champs de vecteurs des déplacements

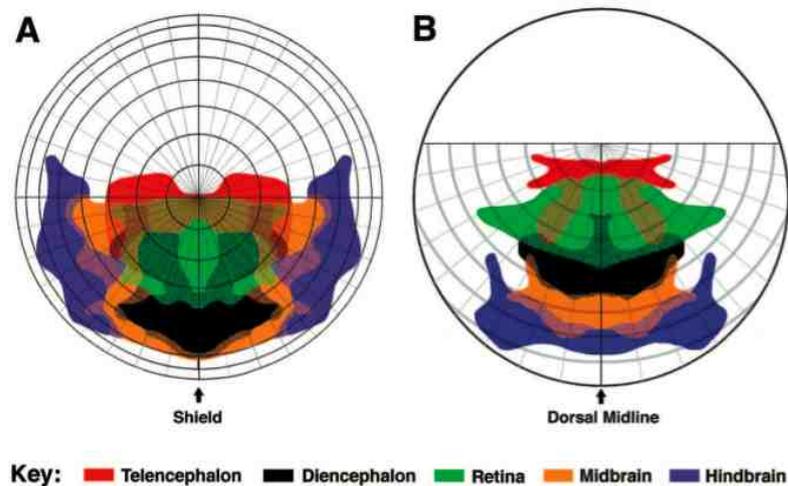


FIG. 1.11 – Organisation du système nerveux central du poisson zèbre. Illustration de la carte des territoires présomptifs, à 6hpf (A) et 10hpf (B). Les domaines occupés par les progéniteurs de chaque subdivision du cerveau sont codés avec leur couleur représentative, comme indiqué en anglais dans la légende. Les aires de recouvrement entre deux domaines sont représentées par un mélange de couleurs. (d'après [Woo & Fraser \(1995\)](#))

extraits de la microscopie à laser fluorescent.

Un autre exemple ([Ward *et al.* \(2007\)](#)) de l'établissement de la carte des territoires au cours de la période de gastrulation (6 - 10hpf,) se focalise sur la morphogenèse du foie et du pancréas. Il existe des progéniteurs de ces organes dans l'ensemble des 4 quadrants avec aucune asymétrie particulière. Il est intéressant de constater que les clones qui ont uniquement contribué au foie ont tendance à être positionnés dans la partie ventrale de l'embryon, alors que les clones qui ont uniquement contribué au pancréas ont tendance à être dans la face dorsale.

Une autre expérience ([Shih & Fraser \(1996\)](#)), se focalise sur les cellules qui composent le *shield*, qui correspond au léger épaissement au niveau de la fine pointe de l'anneau du blastoderme. Lorsque ces cellules sont retirées ou greffées de l'autre côté de l'embryon à 50% de l'épibolie, il en résulte une malformation ou une duplication des axes de l'embryon. Cette expérience soutient l'idée que le *shield* joue un rôle d'organisation capital et est nécessaire à la formation des axes principaux durant l'embryogenèse.

1.2.4 Champs Morphogénétiques

L'apparente modularité des organismes vivants et la stabilité de certains groupes de cellules lors du développement en réponse à des greffes et des transplantations a amené les biologistes à définir le concept de champs morphogénétiques ([Gilbert & Raunio \(1997b\)](#)).

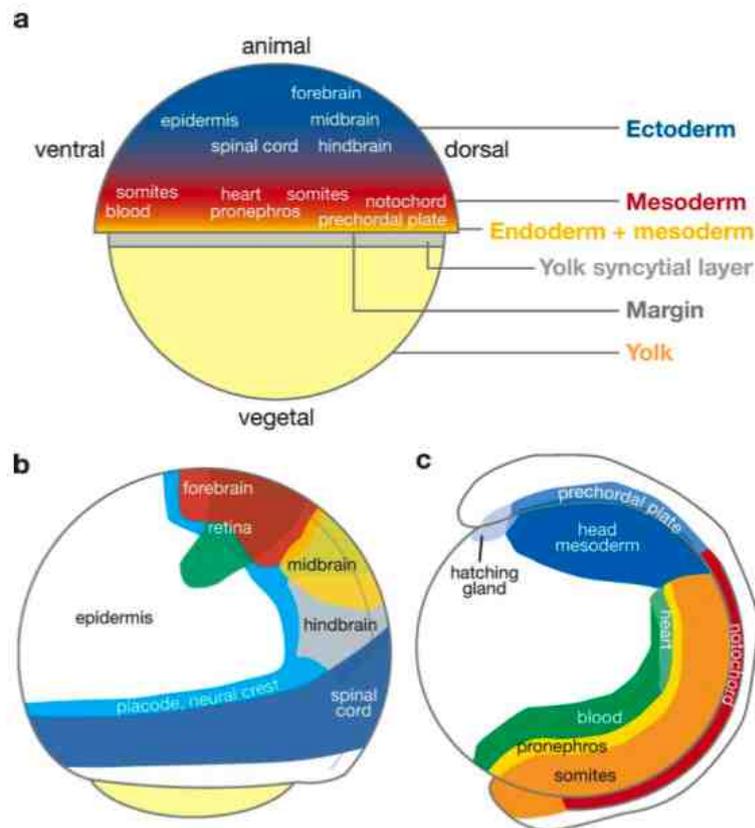


FIG. 1.12 – Carte des territoires présomptifs du poisson zèbre d’après Schier & Talbot (2005). (a) à 50% du stade de l’épibolie, le début de la gastrulation. Vue latérale, dorsale à droite, pôle animal vers le haut. Les couches germinales sont disposées le long de l’axe animal-végétal. Les différents mésodermes et ectodermes sont disposés le long de l’axe dorso-ventral. (b) Évolution de la carte de l’ectodermes à 90% de l’épibolie. Vue latérale, dorsale à droite, pôle animal et antérieur vers le haut. (c) Carte des territoires présomptifs au début du stade mésoblaste. Vue latérale, dorsale à droite, pôle animal et antérieur vers le haut.

Le concept de champs morphogénétiques en biologie du développement a été introduit au début du 19^{ème} siècle par Alexander G. Gurwitsch (*Belousov et al. (1997)*). Les données expérimentales ont, elles, été menées en premier par Ross Granville Harrison's par transplantation de fragments d'embryon dans des régions différentes (*Robertis et al. (1991)*). Harrison a indentifié des *champs* de cellules produisant des organes, comme les membres, la queue et les branchies, et a montré que ces champs pouvaient être fragmentés. Il a donc considéré la notion de *champs* de cellules, plutôt que des cellules individuelles, formant des structures impliquées dans le développement futur de certains organes.

Le concept de champs morphogénétiques a ensuite été développé par Paul Weiss (*Weiss (1939)*). Selon lui, un champ morphogénétique peut être décrit comme un groupe de cellules dont la position respecte les mêmes frontières que la carte des territoires présomptifs.

Scott Gilbert (*Gilbert et al. (1996)*) propose que ce concept de de champs morphogénétiques est à mi-chemin entre les gènes et l'évolution. Les gènes agiraient sur les champs, qui eux même agiraient ensuite sur le développement de l'organisme.

Bolker Jessica (*Bolker (2000)*) décrit les champs morphogénétiques pas seulement comme un début de structures ou d'organes, mais comme des entités dynamiques, avec leurs propres processus de développement local, qui seraient le cœur des phénomènes émergents dans l'évolution du développement ("*Evo-Devo*"). Les questions de l'embryologie comparée sont aujourd'hui reprises dans le contexte de l'Evo-Devo. Cette discipline récente est née de la génétique du développement et utilise essentiellement des stratégies descriptives fondées sur la comparaison des domaines d'expression de gènes orthologues. Au delà des limites actuelles de ses méthodes et de ses concepts, l'Evo-Devo laisse entrevoir la possibilité d'une refondation de la théorie de l'évolution prenant en compte les processus morphogénétiques tels que les explicite la biologie du développement.

Une autre approche des champs morphogénétiques pourrait être le concept de « domaines progéniteurs » (*Davidson (1993)*), représentant tous les descendants d'une cellule. Cependant cette définition ne correspondrait pas strictement au phénomène observé.

La notion de champs morphogénétiques reste aujourd'hui encore assez floue. Comment est-ce que la création de structures se fait au sein de l'embryon? L'origine des différenciations cellulaires, probablement liée à l'expression de telle ou telle voie de signalisation, reste incertaine. La biologie du développement est un champ très vaste, et il semble que les aspects mécaniques du développement et en particulier l'intégration des comportements cellulaires au niveau macroscopique sont encore mal compris.

Nous nous restreindrons donc à une définition assez générale :

Un champ morphogénétique est une population de cellules :

- dont les frontières sont déterminées par des discontinuités de champ de vecteurs,
- ayant les mêmes caractéristiques pour les sous arbres de leur descendance

– et un comportement morphodynamique cohérent.

1.2.5 Importance de l'étude clonale

Pendant le développement, les cellules embryonnaires affichent une grande variété de comportements fondée sur des combinaisons variées d'opérations cellulaires élémentaires : la mobilité, l'adhésivité, la polarisation, le changement de forme, la division et la mort cellulaire (Gilbert & Raunio (1997a)). Cela conduit à la formation de structures embryonnaires qui sont souvent transitoires pour aboutir à une carte des territoires présomptifs (cf 1.2.3). Le défi actuel pour les embryologistes est de déterminer comment ces deux aspects distincts du développement de l'embryon co-existent.

Le comportement des cellules et l'attribution de la carte des territoires à un groupe cellulaire pourraient être deux aspects distincts contrôlés indépendamment et en parallèle. Toutefois, la progression vers un devenir cellulaire spécifique et la sélection d'un comportement particulier peuvent aussi dépendre d'un même processus. La première théorie de *co-contrôle*, bien qu'élémentaire, ne tient pas compte de l'historique et des caractéristiques contingentes de l'évolution qui contraignent le développement. D'un point de vue évolutif, le développement de comportements cellulaires de plus en plus complexes soulève des questions sur les avantages des sélections associées (Raff (1996)). Il est possible que les comportements cellulaires donnent accès à de nouvelles interactions cellulaires, renforçant ainsi la plasticité du développement. Cela est particulièrement plausible chez les vertébrés, contrairement aux organismes aux comportements cellulaires peu élaborés qui ont une invariance dans l'arbre de lignage cellulaire.

Les mouvements qui se produisent au cours de la gastrulation, ont toujours fasciné les embryologistes (Stern (2004)). Ils ont été étudiés chez les amphibiens (Keller (2002)), les poussins (Psychoyos & Stern (1996)) et bien-sûr le poisson zèbre (Myers *et al.* (2002)), en dépit d'une mauvaise connaissance des mécanismes moléculaires sous-jacents. La description des comportements cellulaires devrait déterminer la façon dont les cellules sont impliquées dans la formation de structures embryonnaires et, surtout, permettrait de comprendre la coordination entre les aspects génétiques et le devenir cellulaire.

Il n'est pas encore possible aujourd'hui de générer des reconstructions 4D complètes d'un embryon de vertébrés en raison du grand nombre de cellules impliquées et parce que le développement est un processus à long terme. De plus, les embryons sont souvent mobiles et il est difficile d'interférer sur la mobilité sans interférer sur le développement. L'analyse des comportements cellulaires est principalement effectuée par l'étiquetage de groupe de cellules ou, plus rarement, par l'observation de cellules et leur descendance (Gardner (1985)). Cet étiquetage permet une analyse globale de leur devenir, de leur prolifération et des mouvements des tissus, mais il ne donne aucune information sur les comportements individuels, ce qui est souvent nécessaire pour déterminer les paramètres intrinsèques au développement. L'analyse des clones est basée sur l'étude de clones issus de

cellules uniques. Il donne de très précieuses informations sur les caractéristiques cellulaires et permet de générer des lignées de cellules embryonnaires. Il peut permettre de détecter l'apoptose (Simpson (1979)) et de fournir des informations sur le taux et le mode de croissance (soit orientée ou isotrope). Il est en outre facile grâce à l'analyse clonale de détecter les régions (Mathis & Nicolas (2000)) ou les compartiments (Mathis *et al.* (1999)) d'une structure et d'étudier l'implication des cellules souches dans leurs formations (Nicolas *et al.* (1996)).

1.3 Division cellulaire et morphogénèse

La détection de la division cellulaire lors du développement de l'embryon est une des principales difficultés dans la reconstruction du lignage cellulaire (cf chapitre 4). Afin de mieux la détecter, il est nécessaire de mieux comprendre ce phénomène extrêmement important dans processus de l'embryogénèse.

La division cellulaire est le processus fondamental par lequel une cellule mère donne deux cellules filles *identiques* entre elles et à la cellule dont elles dérivent. Le cycle cellulaire est essentiellement constitué de l'interphase, au cours de laquelle les chromosomes sont répliqués, et de la mitose, au cours de laquelle les chromosomes se répartissent entre les deux cellules filles. À côté de ce cycle de division chromosomique, il existe un cycle cytoplasmique, un cycle centrosomique, un cycle nucléolaire et le cycle de l'appareil de Golgi. Tous ces événements sont interdépendants et parfaitement couplés (Malumbres & Barbacid (2001); Meijer *et al.* (2003)). La division cellulaire est un processus essentiel au développement embryonnaire mais également vital pendant toute la vie de l'organisme adulte. Le mécanisme de division cellulaire, extrêmement complexe, est régulé par un grand nombre de protéines intervenant très transitoirement et dans un ordre défini, permettant ainsi la succession précise des différentes étapes du cycle cellulaire. Il arrive occasionnellement qu'une anomalie de régulation de la division apparaisse, apparition favorisée par les produits cancérigènes. La cellule devient alors cancéreuse et va se développer librement en une tumeur qui, par métastase, va quitter son tissu de départ et se disséminer dans tout l'organisme, entraînant des conséquences dramatiques pour l'individu. On comprend alors l'intérêt d'étudier les régulateurs de la division cellulaire, dont les anomalies sont souvent à l'origine des cancers.

La division cellulaire a été découverte en 1875 chez les plantes par Edward Strasburger (Strasburger (1880)). Flemming, en 1878, montre qu'elle existe aussi chez les animaux et appelle chromatine la substance dont sont faites les particules "colorantes" du noyau (Flemming (1878)). Waldeyer en 1888 donne le nom de chromosome à ces particules (Waldeyer (1888)). La division cellulaire a donc été décrite il y a plus de 150 ans et a fait l'objet de très nombreuses études en raison de son caractère fondamental. En revanche, sa régulation intracellulaire n'est connue que depuis quelques décennies, grâce à des travaux entrepris sur des modèles biologiques très variés. La complémentarité d'approche et la convergence

des résultats ont finalement abouti à un modèle général de contrôle intracellulaire de la division. Ces travaux ont été couronnés par l'attribution, en octobre 2001, du prix Nobel de physiologie et de médecine à trois acteurs essentiels du décryptage des mécanismes régulant le cycle cellulaire, Tim Hunt, Paul Nurse et Leyland Hartwell. Le cycle cellulaire reste un grand sujet d'étude en biologie cellulaire et moléculaire comme en témoignent la dizaine d'articles publiés chaque jour dans ce domaine, les multiples applications possibles de la compréhension de la division cellulaire et l'intérêt croissant de l'industrie pharmaceutique.

En avril 2006, Potapova et al affirment avoir découvert un processus inversant la division cellulaire, ce qui pourrait conduire à de nouveaux procédés de lutte contre le cancer mais aussi à de nouvelles armes empêchant la cicatrisation des plaies (Potapova *et al.* (2006)).

1.3.1 Phases du cycle cellulaire

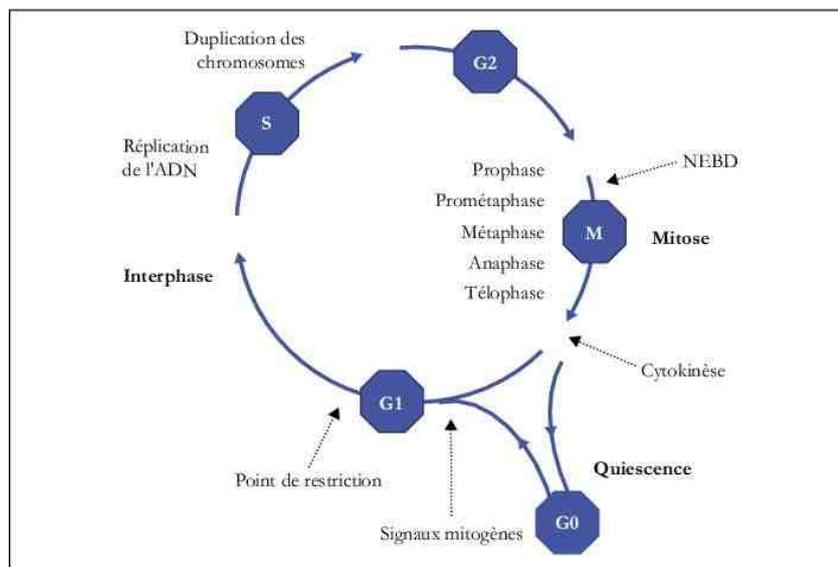


FIG. 1.13 – Phases et événements majeurs du cycle cellulaire.

Lorsqu'elles ne se divisent pas, les cellules sont dites en quiescence ou en phase G0. Sous l'effet de signaux mitogènes, elles s'engagent dans un cycle de division (figure 1.13). Le cycle cellulaire est classiquement divisé en quatre phases, G1, S, G2, M. Au cours de la phase G1 (de gap = intervalle), les cellules passent par le point de restriction, une sorte de point de non-retour à partir duquel le cycle est irréversiblement engagé et l'entrée en division ne dépend plus de la présence des facteurs mitogènes. La phase G1 est préparatoire à la phase S au cours de laquelle l'ADN est répliqué. La phase G2 précède la mitose (phase M), au cours de laquelle les chromosomes dédoublés sont répartis dans les deux cellules filles, grâce au fuseau de division. L'interphase correspond aux phases G1, S et G2 avant la mitose (M).

Les événements qui surviennent durant la mitose sont parfaitement visibles au microscope en raison de la condensation de la chromatine en chromosomes. La mitose est la distribution des deux groupes de chromosomes en deux noyaux séparés et égaux. Elle aboutit à la réplication exacte de l'information génétique. Pour des raisons de commodité, les biologistes divisent le processus en quatre phases : la prophase, la métaphase, l'anaphase et la télophase (figure 1.14) . En réalité, la mitose est un processus continu au cours duquel les différentes phases se chevauchent de façon presque imperceptible.

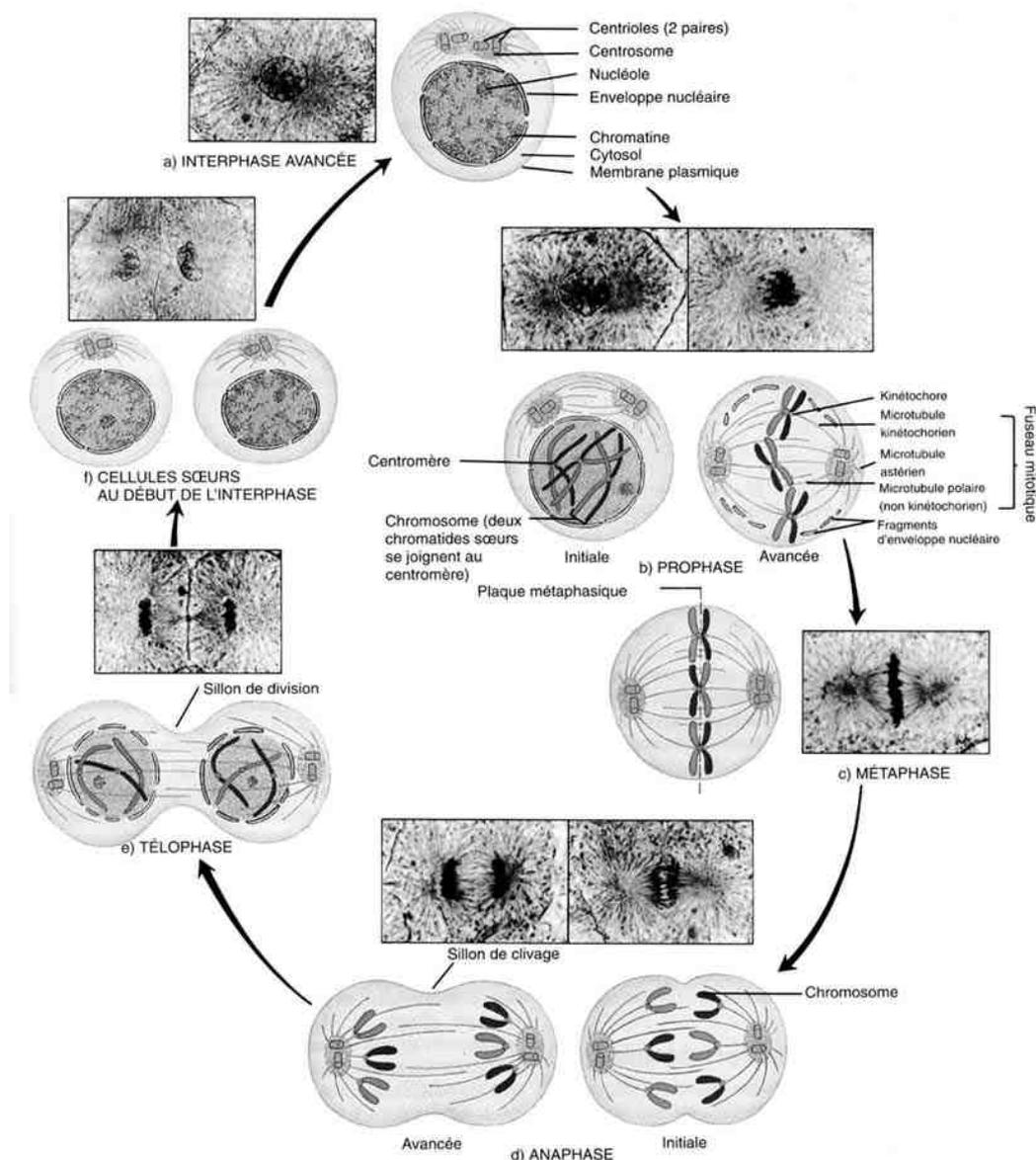


FIG. 1.14 – Les différentes phases de la mitose.

- **La prophase.** La première étape de la mitose est appelée prophase. Au début de la prophase, la chromatine se condense et se raccourcit sous forme de chromosomes. Cette condensation est nécessaire, car de longs brins pourraient s'enchevêtrer, ce qui entraverait leurs mouvements. Comme la réplification de l'ADN a eu lieu pendant l'interphase, chaque chromosome de la prophase contient une paire de molécules d'ADN à double brins identiques appelées chromatides. Chaque paire de chromatides est maintenue ensemble grâce à une petite structure sphérique, le centromère, indispensable à une ségrégation chromosomique adéquate. Ultérieurement dans la prophase, la synthèse de l'ARN est, pendant ce temps, temporairement arrêtée ; l'enveloppe nucléaire se rompt et elle est absorbée dans le cytosol. De plus, un des centrosomes et ses centrioles se déplacent vers le pôle (extrémité) opposé de la cellule. Simultanément, les centrosomes commencent à former le fuseau mitotique, un assemblage de microtubules responsables du mouvement des chromosomes et qui a la forme d'un ballon. L'allongement des microtubules entre les centrosomes pousse ces derniers jusqu'aux pôles de la cellule de sorte que le fuseau s'étend de pôle en pôle.
- **La métaphase.** Au cours de la métaphase, soit la deuxième étape de la mitose, les centromères des paires de chromosomes s'alignent exactement au centre du fuseau mitotique. Cette région centrale est appelée plaque équatoriale.
- **L'anaphase.** La troisième étape de la mitose, l'anaphase, est caractérisée par la division et la séparation des centromères, de même que par le déplacement des deux chromatides sœurs de chaque paire vers les pôles opposés de la cellule. Une fois séparées, les chromatides sœurs sont appelées chromosomes filles. Le mouvement des chromosomes est dû au raccourcissement des microtubules. À mesure que les chromosomes se déplacent durant l'anaphase, ils prennent la forme d'un V, alors que les centromères ouvrent la voie et semblent charrier les parties traînantes des chromosomes vers les pôles opposés de la cellule.
- **La télophase.** La dernière étape de la mitose commence dès que cesse le mouvement chromosomique. Pendant la télophase, les groupes identiques de chromosomes aux pôles opposés de la cellule se déroulent et reprennent leur forme filiforme de chromatine. Une nouvelle enveloppe nucléaire se reforme autour de chaque masse de chromatine ; finalement, le fuseau mitotique se désagrège.

La cytokinèse, correspondant à la séparation physique des deux cellules filles, achève la division de la cellule. Lorsque les cellules cessent toute prolifération, sous l'effet de signaux antimitogènes ou suite à la disparition des agents mitogènes, elles quittent le cycle cellulaire et retournent en phase de quiescence. Les quatre phases du cycle cellulaire s'enchaînent de façon coordonnée, chaque phase ne pouvant commencer que lorsque la précédente s'est déroulée correctement. En effet de nombreux mécanismes de contrôle assurent une sorte de « contrôle qualité » à chaque étape et bloquent le déroulement du cycle lorsqu'une anomalie est

détectée (endommagement de l'ADN, ADN répliquée de façon incomplète, chromosomes non attachés au fuseau mitotique) (Stewart *et al.* (2003)).

1.3.2 Les divisions cellulaires du poisson zèbre.

Au cours de l'embryogenèse du poisson zèbre, les divisions ne se font pas de façon aléatoire (Adams (1996)). Les mécanismes qui contrôlent l'orientation de la division cellulaire *in vivo* ne sont cependant pas connus.

En 1925, dans Wilson (1925), EP Wilson postule que l'axe du fuseau mitotique s'aligne avec le plus grand axe de la masse protoplasmique, et que la division a tendance à le couper transversalement. Cet auteur souligne cependant que d'autres facteurs mécaniques, comme la pression, la forme, etc.. pourraient rendre plus subtiles et plus complexes les opérations impliquées dans le développement général de l'organisme. Les premières divisions du développement de l'embryon sont toutes orientées perpendiculairement à la division précédente, car les centrosomes migrent de la même façon au cours la formation du fuseau mitotique (Hertzler (1992)). Cosllo, en 1961, souligne l'importance de l'orientation du fuseau dans le processus de division au cours de la morphogénèse. (Costello (1961)).

La corrélation entre l'allongement de la forme des cellules et l'orientation de la division décrite par les biologistes cellulaires reste un paradigme dans les études sur le développement. En outre, l'analyse du développement précoce de l'embryon et de la morphogénèse des tissus indiquent des différences spatiales dans les processus de divisions. Cela suggère qu'une organisation sous-jacente des cellules influe dans le processus du développement (Strome (1993)). Les cellules, en sécrétant leur matrice extra-cellulaire, définissent leurs caractéristiques d'adhérence. Leurs formes représentent donc une manifestation des ces tensions adhésives. Théry et al (Théry & Bornens (2006)) suggèrent que ces structures guident l'orientation du fuseau et par conséquent les positions des cellules filles.

L'orientation de la division cellulaire est une partie intégrante du développement dans les processus allant de la division asymétrique à l'histoire clonale de la cellule dans l'élaboration des organes (Ahringer (2003); Sausedo *et al.* (1997)). De nombreuses équipes ont mis en évidence le rôle de l'orientation de la division dans l'élongation du tissu (Schoenwolf & Alvarez (1989); Wei & Mikawa (2000)). Gond et al (Gong *et al.* (2004)) ont analysé les divisions mitotiques du poisson zèbre durant la période de gastrulation et ont trouvé que les cellules dans les tissus dorsaux se divisent de façon préférentielle le long de l'axe animal-végétatif.

Au cours du développement précoce du poisson zèbre on observe des *vagues* de division même si elles deviennent de moins en moins synchrones (Kimmel *et al.* (1995)). Kimmel et al (Kimmel (1994)) ont montré que les cellules qui contribuent à la partie postérieure du cerveau et à la moelle épinière, se divisent préférentiellement parallèlement à l'axe antéro-postérieur (AP) de l'embryon au cours des vagues de divisions 14 et 15, puis se réalignent sur l'axe médio-latéral (ML) à la division 16, qui a lieu au cours de la neurulation. D'autres équipes ont confirmé l'orientation préférentielle des divisions selon des axes ML et AP en fonction de

leurs positions spatiales (Woo & Fraser (1995) ; Concha & Adams (1998)).

Concha et al (Concha & Adams (1998)) ont analysé les différents mouvements et les formes des cellules et les ont corrélés avec l'orientation de la division sous la couche superficielle du poisson zèbre (EVL) pendant la gastrulation et la neurulation (figure 1.15). Ils émettent l'hypothèse que les propriétés intrinsèques de la cellule pourraient déterminer leur angle de division, mais qu'actuellement les paramètres disponibles ne suffisent pas pour établir un stéréotype de division cellulaire. Ils suggèrent donc qu'il existe d'autres principes d'organisations qui doivent contrôler l'orientation de la division.

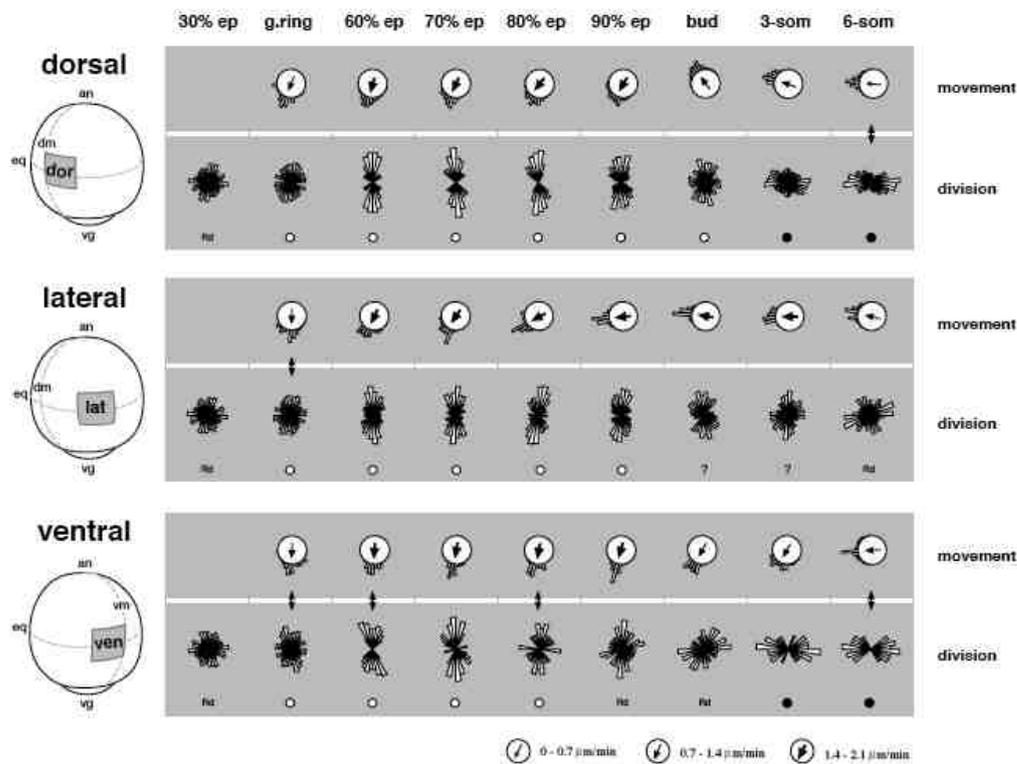


FIG. 1.15 – Les mouvements cellulaires et l'orientation des divisions sont fortement alignés dans l'épiblaste au cours de la gastrulation et au début de la neurulation. Ils sont représentés au niveau de l'équateur pour trois régions : dorsale, latérale et ventrale. Les mouvements (en haut) et des divisions (en bas) sont reportées sur un histogramme circulaire à environ une heure d'intervalle. La moyenne d'orientation du mouvement est indiquée par une flèche, son épaisseur indique la vitesse moyenne (de Concha & Adams (1998)).

Quand la plaque neural commence à se différencier, il y a une transition progressive dans la direction des divisions cellulaires depuis le pôle animal vers l'équateur. L'orientation des divisions de l'épiderme ventral et de la plaque neurale dorsale se fait le long de l'axe ML (Papan & Campos-Ortega (1994)). Geldmacher-

Voss et al ([Geldmacher-Voss et al. \(2003\)](#)) ont observé que dans la plaque et le tube neural, les divisions cellulaires sont orientées parallèlement au plan de l'épithélium neural, tandis que plus tardivement dans la *quille* neurale, les cellules se divisent perpendiculairement à celui-ci. Ce changement dans l'orientation de la mitose est due à une rotation de 90° du fuseau mitotique et les conditions de cette réorientation font l'objet d'un débat.

Poggi et al ([Poggi et al. \(2005\)](#)) examinent l'orientation des divisions cellulaires dans la rétine du poisson zèbre. Ils trouvent que ces cellules se divisent une seule fois le long de la circonférence de l'oeil et qu'une des cellules filles devient une cellule ganglionnaire rétinienne (GRC).

Ahringer et al ([Ahringer \(2003\)](#)) suggèrent, pendant la formation de l'épithélium, de corrélérer la polarité cellulaire à la position du fuseau mitotique pour étudier l'orientation des cellules filles.

L'ensemble des expériences menées sur les orientations des divisions se cantonnent à mesurer un petit nombre de cellules dans des régions très localisées. Elle n'ont pu observer en général que les cellules sur la surface de l'embryon et donc ne peuvent pas voir si la cellule se divise en profondeur. L'observation visuelle d'un angle en 3 dimensions amène au biais d'étude de l'angle solide qui n'a pas la même distribution de probabilités qu'en 2 dimensions. Ces études restent donc très qualitatives, hétérogènes au plan spatio-temporel avec des stratégies de mesures diverses qui rendent difficiles les comparaisons et la généralisation.

Nous proposons, dans le chapitre 4, un protocole standard et automatisé de mesure des caractéristiques de la division cellulaire au cours de la morphogénèse, fondé sur des combinaisons d'algorithmes utilisant les positions cellulaires et leurs trajectoires définies par rapport à un référentiel. Au pôle animal, nos études indiquent qu'au cours de la gastrulation, dans toute l'épaisseur du blastoderme l'orientation de la division est corrélée avec la direction de la vitesse de déplacement et la déformation du voisinage.

1.4 Imagerie 4D *in toto*

L'imagerie optique constitue un outil primordial pour la biologie. Cependant, la forte diffusion de la lumière par les tissus épais empêche leur observation à l'aide des techniques optiques classiques. Récemment, de nouvelles microscopies à balayage basées sur des phénomènes optiques non linéaires ont été développées, comme les microscopies à fluorescence biphotonique. Elles permettent l'imagerie en profondeur des tissus épais et l'application *in vivo* des techniques optiques jusque là limité aux échantillons *in vitro*.

1.4.1 Microscopie à fluorescence

La microscopie à fluorescence exploite la capacité qu'ont certaines molécules d'émettre de la lumière quand on les éclaire avec une lumière de longueur d'onde

supérieure. La lumière émise par une source de lumière blanche est filtrée pour isoler la longueur d'onde qui va exciter la préparation puis focalisée sur la zone d'observation par l'objectif. La lumière émise est captée par l'objectif, filtrée pour isoler les longueurs d'ondes parasites qui pourraient brouiller le signal (comme la longueur d'onde d'excitation par exemple) puis observée. Le but est de pouvoir suivre l'évolution *in situ* d'espèces biologiques préalablement marquées par des fluorophores endogènes ou exogènes. Le système permet de sonder des volumes de tissus avec une résolution submicronique. La fluorescence est la propriété que possèdent certains corps d'émettre de la lumière après avoir absorbé des photons de plus forte énergie (longueur d'onde plus courte). Nous allons présenter ici les microscopes utilisées dans cette thèse, mais il en existe bien plus comme par exemple les microscopes de récupération de la deuxième et troisième génération d'harmoniques, ainsi que les microscopes électroniques pouvant atteindre de très grandes résolutions (de l'ordre du nanomètre).

1.4.2 Marqueurs Fluorescents

Il est possible de se servir du phénomène de fluorescence pour faire des images d'un objet (matière organique) en microscopie, s'il est différencié en régions dont certaines sont fluorescentes et d'autres pas. Ceci témoigne d'une différence de composition chimique entre ces régions. Si l'objet n'est pas naturellement fluorescent, une façon de le différencier en régions fluorescentes est de lui injecter un marqueur fluorescent qui diffuse dans l'échantillon et se fixe préférentiellement dans certains sites chimiques. Ces marqueurs sont appelés fluorochromes. Dans une image, l'intensité variable de l'émission fluorescente selon la localisation dans l'objet témoigne des différences dans la diffusion et la fixation des marqueurs à l'intérieur de l'objet.

En 1961, Shimomura et Johnson ([Shimomura & Johnson \(1972\)](#)) ont découvert que la source fluorescente verte de la méduse *Aequorea victoria* est une protéine (GFP pour *Green Fluorescent Protein*) produite naturellement. En 1992 Prasher et al. ([Prasher et al. \(1992\)](#)) ont réussi à la cloner et ont déterminé sa séquence génétique. Mais la *Green Revolution* ([Stearns \(1995\)](#)) a réellement commencé en 1994 ([Chalfie et al. \(1994\)](#)), quand cette protéine a été exprimée dans d'autres organismes en modifiant leurs génomes. Cette découverte a eu l'impact le plus important sur la recherche en biologie de ces dernières décennies. Elle permet aux biologistes de marquer par fluorescence presque toutes les protéines en fusionnant le gène de la protéine d'intérêt à celui de la GFP qui contient toutes les informations nécessaires à la synthèse de fluorophore. M. Chalfie, O. Shimomura et R.Y. Tsien ont reçu le prix Nobel de Chimie en 2008 pour leur découverte et le développement cette protéine.

Depuis cette protéine GFP à l'état sauvage (wpGFP), de nombreuses variétés de protéines fluorescentes similaires ont été conçues et synthétisées en laboratoire ([Lippincott-Schwartz & Patterson \(2003\)](#); [Chudakov et al. \(2005\)](#)). De plus, la découverte de protéines similaires à la GFP, et donc fluorescentes à l'état naturel sur des coraux (*Anthozoa*), a considérablement élargi la gamme de couleurs dis-

ponibles et aujourd'hui de la bases de données contient plus de 30 différents fluorophores (Miyawaki *et al.* (2003)). Les couleurs obtenues ont différents spectres d'excitation et d'émission, une capacité de fluorescence plus élevée ainsi qu'une plus grande résistance que la GFP. Il est primordial dans ce genre d'expérience que le marqueur fluorescent ne perturbe pas l'intégrité et la fonctionnalité de la protéine marquée. Le facteur le plus limitant de l'utilisation des fluorophores est le photoblanchiment (*photobleaching effect*) (Klonis (2002)). Les fluorophores perdent leur luminescence au cours du temps, ce qui diminue l'intensité de l'excitation, cela limite temporellement l'observation d'un spécimen. Par exemple, concernant les embryons du poisson zèbre, les noyaux marqués avec la protéine GFP peuvent être imagés pendant un peu plus d'une dizaine d'heures en microscopie biphotonique. Une des avancées récentes les plus intéressantes en marqueurs fluorescents est les fluorophores *photo-activables*, ils agissent en tant que commutateurs *on-off* lors de la réception de l'excitation (Patterson & Lippincott-Schwartz (2002)). La disponibilité des techniques de marquage de protéines fluorescentes a changé considérablement la recherche en biologie. Dans ce projet, la possibilité d'effectuer des expériences avec divers marqueurs fluorescents fournit un outil extrêmement utile pour étudier *in vivo* les processus de l'embryogenèse.

1.4.3 Le microscope confocal à balayage laser : CLSM

Les images en microscopie à fluorescence classique ont une perte de résolution de l'image due à l'excitation des fluorochromes se situant hors du plan focal. En effet les fluorochromes sont excités par le laser sur toute l'épaisseur de la préparation, ce qui se traduit par une image contaminée par un bruit de fond. L'objectif de la microscopie confocale à balayage laser (CLSM pour *confocal laser scanning microscopy*) est d'éliminer la lumière provenant des plans défocalisés qui parasitent le plan focal (figure 1.16).

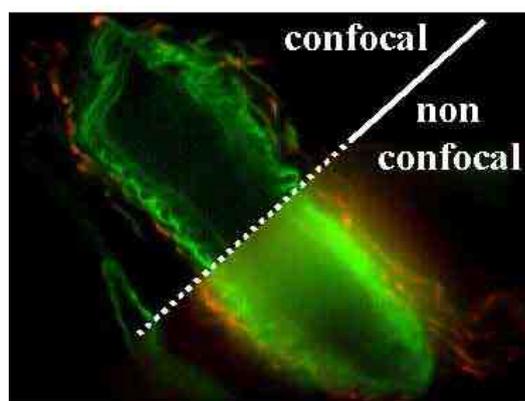


FIG. 1.16 – Différence d'acquisition d'images entre la microscopie à fluorescence classique et la microscopie confocale à balayage laser

Le concept du microscope confocal à balayage laser a été développé et breveté

par M Minsky en 1961 (Minsky (1961)). Cependant, son invention a été un peu précoce car le succès de cette technique est arrivée après l'expiration du brevet, avec les marqueurs fluorescents en biologie (Lichtman & Conchello (2005); Vonesch *et al.* (2006); Fricker *et al.* (2006)).

Le microscope confocal à balayage laser (figure 1.17) est un microscope de fluorescence où l'échantillon observé est éclairé point par point par un faisceau laser balayant le plan (X,Y) comme un scanner. Lors du passage du faisceau en un point, les molécules fluorescentes présentes à cette position dans toute l'épaisseur Z de l'échantillon sont excitées. La fluorescence émise est ensuite filtrée par un diaphragme situé au point de focalisation de la lumière avant le détecteur (photomultiplicateur). Par ce biais, on détecte uniquement la fluorescence provenant du plan de focalisation (X,Y). Les déplacements dans l'épaisseur de l'échantillon sont effectués à l'aide d'une platine motorisée. Ceci permet d'avoir des déplacements reproductibles, et par conséquent de faire des coupes optiques adjacentes. Les images faites en microscopie confocale étant numérisées, le traitement des coupes optiques adjacentes permet de visualiser et d'extraire des informations 2D et 3D. La résolution optique d'un microscope confocal est sensiblement plus élevée que celle d'un microscope conventionnel, tant dans le plan (X,Y) de visualisation (amélioration d'environ 30 %), que dans l'épaisseur Z de l'échantillon (amélioration d'environ 70 %). Les résolutions limites sont approximativement de 150 μm dans le plan (X,Y) et de 300 μm dans l'épaisseur Z de l'échantillon. La contribution des sondes fluorescentes localisées hors du plan de focalisation est nulle ou négligeable, ce qui améliore la netteté de l'image de l'échantillon observé. Le découpage optique de l'échantillon au lieu d'un découpage physique facilite la préparation de l'échantillon et réduit les artéfacts de préparation. Cependant comme en microscopie conventionnelle, il existe deux limites :

- la distance de travail de l'objectif utilisé impose une épaisseur maximale d'échantillon,
- la nature de la préparation vis à vis de la diffusion et de la réflexion de la lumière qui le traverse influence la qualité des images obtenues.

On peut obtenir des *time-lapse* (volume 3D + temps), en faisant l'acquisition temporelle de plusieurs volumes d'images. Le choix de la taille du volume en (x, y, z) impose une limite temporelle (t) d'imagerie, et cela est un point critique dans l'acquisition biologique *in-vivo* lorsque le phénomène à observer est plus long (Vermot *et al.* (2008)).

1.4.4 CLSM multi-photons : MLSM

La différence primordiale entre les microscopes confocaux et les microscopes bi-photons, est que le premier éclaire tout un plan et le deuxième permet de n'éclairer qu'une région précise. Les développements de l'optique non linéaire et les progrès réalisés dans la production de sources pulsées de lumière cohérente ont contribué à faire émerger une nouvelle technique de microscopie de fluorescence offrant des avantages multiples. Le concept de l'absorption de deux photons

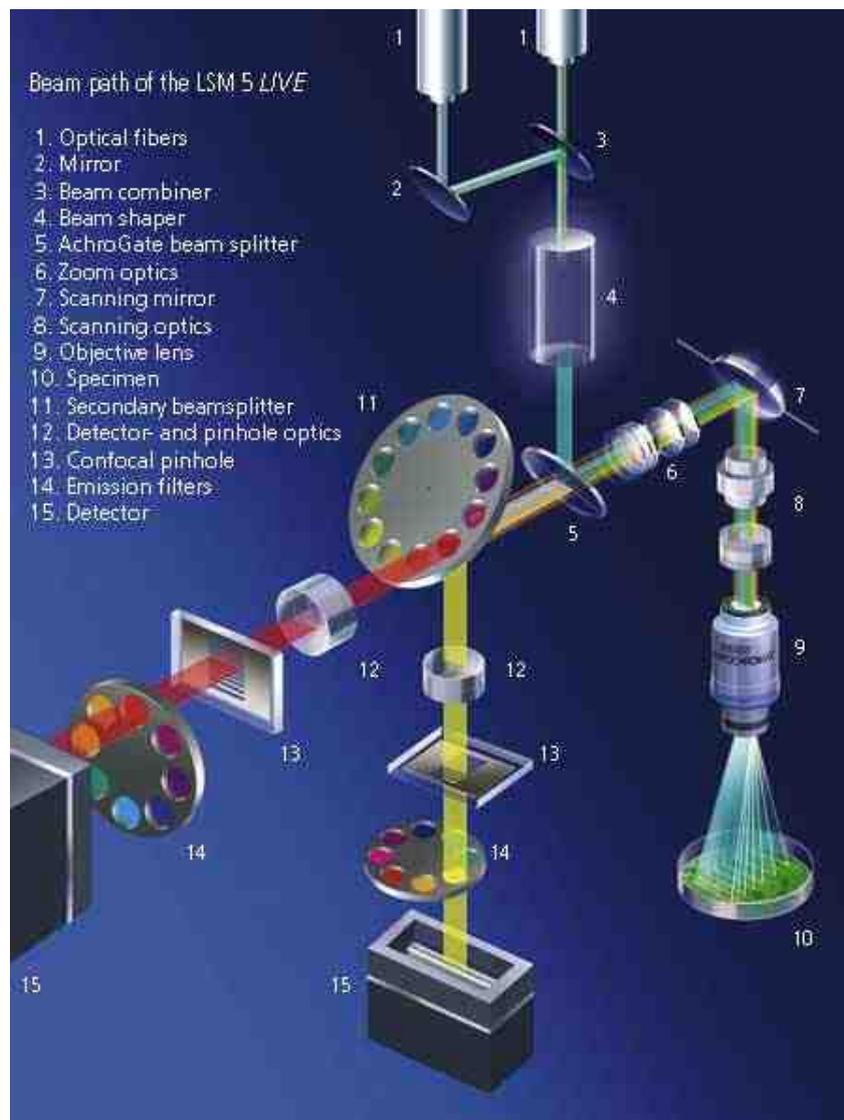


FIG. 1.17 – Représentation d'un microscope confocal à balayage laser. Pour fonctionner en fluorescence, un microscope comprend une source lumineuse adéquate et un bloc interchangeable situé sur le trajet lumineux. La source est constituée d'une lampe à vapeur de mercure dont le rayonnement est filtré en longueur d'onde par le filtre d'excitation qui conserve une bande étroite du spectre lumineux. Ce faisceau est dévié vers l'échantillon par un miroir dichroïque qui possède un fort coefficient de réflexion pour cette fenêtre de longueur d'onde et un fort coefficient de transmission pour la bande qui correspond au rayonnement fluorescent émis par l'échantillon. Ce rayonnement émis est sélectionné ensuite par un filtre barrière. Le filtre d'excitation, le miroir dichroïque et le filtre barrière, assemblés dans un seul bloc, peuvent être changés et choisis pour être spécifiques du fluorochrome utilisé. (D'après les caractéristiques techniques du LSM 5 LIVE de Zeiss)

a en premier été décrit en 1931 par M Goppert-Mayer (Goppert-Mayer (1931)) et consiste en la combinaison simultanée de l'énergie de deux photons pour la promotion du fluorophore à l'état excité. L'excitation moléculaire par absorption simultanée de deux photons a elle été introduite en microscopie confocale par W. Webb et son équipe (Denk *et al.* (1990)). Cette technique de fluorescence basée sur l'absorption non linéaire de la lumière est référencée sous le nom générique de microscopie à absorption multiphotonique (Xu *et al.* (1996)). A l'heure actuelle les systèmes disponibles permettent essentiellement de réaliser une absorption bi-photonique d'où leur appellation triviale de «microscope à deux photons» (Strickler & Webb (1991)). Bien que les mécanismes physiques impliqués soient complexes, le principe de l'absorption non-linéaire peut se comprendre de manière intuitive. Dans le cas de la fluorescence par absorption linéaire (ou monophotonique), l'énergie d'un photon de longueur d'onde caractéristique va permettre la transition de la molécule fluorescente de l'état de repos à l'état excité. Très rapidement la molécule va retourner à l'état de repos en restituant l'énergie accumulée en différentes étapes et sous différentes formes. L'une d'elles aboutit à l'émission d'un photon par la molécule. Le photon émis sera d'une longueur d'onde supérieure à celle du photon absorbé lors de l'excitation. La fluorescence est émise uniquement à partir d'une région localisée, créant les conditions d'une imagerie confocale inhérente à ce mode d'excitation et ne nécessitant pas de sténopé (*pinhole*) à l'émission comme en microscopie confocale un-photon. La profondeur de pénétration de l'excitation bi-photonique (proche infra-rouge pour les fluorophores absorbant dans le visible) dans les milieux diffusants est d'environ 400 nm et permet de travailler sur les embryons (Periasamy *et al.* (1999)), les tissus (Parasassi *et al.* (2000)) et même sur animaux vivants (Denk & Svoboda (1997)). Les processus photodynamiques tels que la destruction irréversible des molécules fluorescentes ou la production de radicaux libres ne se produisent que dans une région focale. Cette propriété minimise sensiblement l'invasivité de la microscopie bi-photonique comparativement à la microscopie confocale un photon.

1.4.5 Stratégie d'acquisition par parties

Compte tenu des dimensions de l'embryon du poisson zèbre (environ 700 μm) et des limites de l'acquisition (300 μm en profondeur) nous ne pouvons pas imager la totalité de l'embryon en une seule fois. De plus nous voulons reconstruire sur du plus long terme l'arbre de lignage cellulaire du poisson zèbre depuis la première cellule jusqu'à son âge adulte. Nous avons donc établi une stratégie d'acquisition de l'embryon par parties (appelée *multi-shot*). Cela consiste à faire l'acquisition de plusieurs positions d'organismes différents (figure 1.19) et à différents moments du développement. La reconstruction doit se faire ensuite en associant les différents volumes qui comportent des chevauchements spatiaux et temporels. Cela signifie que l'arbre de lignage cellulaire sera établi, au mieux, par une probabilité et non pas comme la reconstruction exacte d'un organisme unique. Nous avons développé dans le chapitre 4 une méthode automatique pour obtenir un référen-

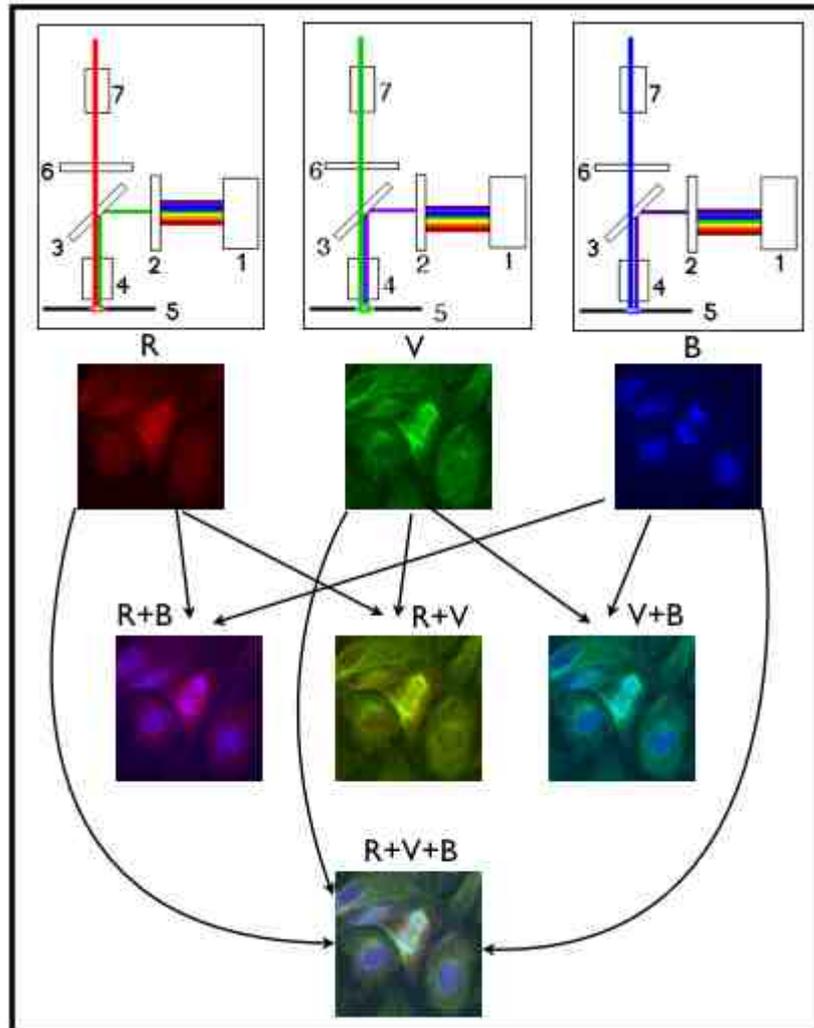


FIG. 1.18 – Exemple d'utilisation de microscopie 3 photons sur une préparation de cellules de Kangourou (PTK) en culture. Elles ont été marquées avec un anticorps anti-kinétochore couplé à un dérivé de la rhodamine (TRITC pour *Tetra methyl Rhodamine Iso Thio Cyanate*), avec un autre anticorps anti-tubuline marqué avec un dérivé de la fluorescéine (FITC pour *Fluoresceine Iso Thio Cyanate*) et avec le DAPI (*Di Aminido Phenyl Indol*), réactif spécifique de l'ADN. A gauche, utilisation de la fluorescence rouge pour visualiser les kinétochores aux pôles du fuseau. Au milieu, la fluorescence verte permet d'observer des microtubules du fuseau, et à droite la fluorescence bleue, l'ADN dans les noyaux en interphase et des chromosomes de l'anaphase. On peut ainsi combiner les différentes couleurs. (Images de Caroline Benlot, à l'UMPC)

tiel spatio-temporel global permettant d'effectuer la reconstruction des différents volumes. Ce référentiel est basé sur les axes principaux (animal-végétatif et médio-latéral) du développement de l'embryon et l'importance d'événements clés dans sa morphogenèse comme la création de l'*hypoblaste* (cf 1.2.2.4).

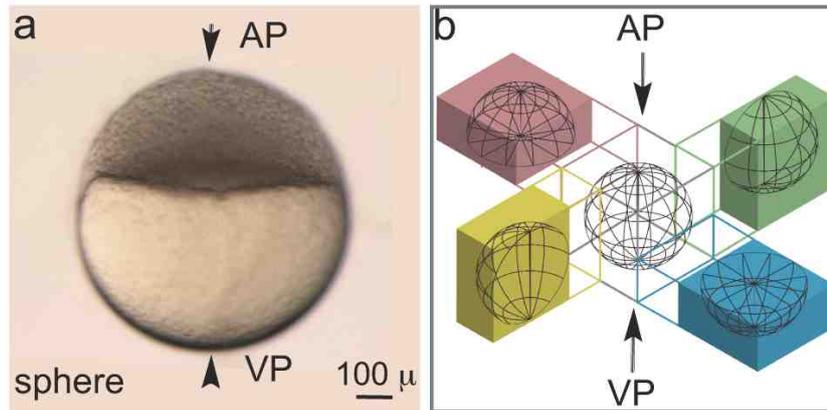


FIG. 1.19 – Stratégie *multi-shot* : l'embryon assimilé à une sphère (a) est imagé en 4 parties différentes (b).

1.5 L'approche des systèmes complexes

1.5.1 Les systèmes complexes sont

Ah qu'il serait beau de pouvoir définir les systèmes complexes (SC) en une belle phrase explicative et récapitulative. Or aujourd'hui, les experts de la communauté ne sont pas encore tout à fait d'accord sur une définition exacte. On peut simplement les introduire par : "Un système complexe est un système composé d'un grand nombre d'entités en interaction locale et simultanée, dont on ne peut prévoir le comportement par un simple raccourci de calcul".

Cela nous donne déjà une idée assez précise de ce que peut être les systèmes complexes. Mais alors on peut trouver de nombreux exemples ou contre-exemples qui ne correspondraient pas à cette définition. Il faut donc pour cela creuser un peu plus loin dans l'explication.

Un système complexe est composé de grandes quantités d'éléments - ou d'agents - qui interagissent simultanément. La modélisation d'un système passe par la conception d'un certain nombre de règles d'évolution, puis une phase de simulation du système en itérant ces règles jusqu'à obtenir un résultat structuré. Un système est dit complexe si le résultat final n'est pas prédictible directement en connaissant les règles.

Les SC se situent de façon ambiguë dans des environnements turbulents, instables et changeants. Ils évoluent et s'adaptent au travers des dynamiques d'interactions internes et externes. Ils existent à travers plusieurs échelles, représentent

des systèmes multi-composants de systèmes, et ils ne sont pas prévisibles au sens scientifique classique.

Ils résultent souvent de processus d'évolution et d'adaptation. Ils présentent des propriétés émergentes : le niveau microscopique sous-jacent fait émerger des formes organisées au niveau macroscopique, lequel influence en retour le niveau microscopique. Des interactions locales et globales peuvent ainsi se conjuguer dans la description de leurs dynamiques. Les caractéristiques d'auto-organisation, d'émergence de propriétés ou de structures cohérentes, sont souvent exigées pour qualifier un système de complexe.

Du fait de la diversité des systèmes complexes, leur étude est intrinsèquement interdisciplinaire. Deux approches complémentaires sont utilisées : certains scientifiques aux prises avec un système complexe particulier cherchent à le comprendre, d'autres cherchent des méthodes et définitions générales applicables à de nombreux systèmes différents.

La reconstruction de la dynamique multi-niveaux des systèmes complexes représente un défi majeur pour la science moderne, et cela devient de plus en plus accessible par l'intermédiaire de la montée en puissance des capacités informatiques.

La science des systèmes complexes cherche à fournir de nouvelles façons de comprendre les facteurs universels physiques, biologiques, écologiques et sociaux. Elle est le processus de reconstruction de la théorie à partir de données. Elle fait le pont entre l'individuel et le collectif : des gènes en passant par les organismes jusqu'aux écosystèmes ; des atomes aux produits à travers les matériaux ; des blocs-notes à Internet, en passant par les citoyens. Elle doit être présente de façon transversale à l'ensemble des disciplines, tout en faisant partie intégrante de chacune. Elle se veut de réduire le fossé entre les sciences pures et appliquées. La science des SC doit aussi permettre de créer des formalismes phénoménologiques pertinents adaptés aux données étudiées. Cela à la fois le point de vue théorique de la complexité algorithmique de Kolmogorov (Li & Vitányi (1997)), et du rasoir d'Occam en épistémologie.

Deux types de démarches interdisciplinaires sont principalement possibles. La première consiste à se donner un objet de recherche intrinsèquement multidisciplinaire comme dans le cas de la cognition : elle conduit à poser des questions sur un même objet à partir de points de vue qui peuvent être très différents. La seconde consiste à étudier une même question à propos d'objets de recherche très différents. C'est cette seconde démarche qui relève d'une science des systèmes complexes. Les deux démarches sont *duales* l'une de l'autre et se renforcent mutuellement.

1.5.2 L'embryon vu comme un système complexe

La multicellularité s'accompagne d'une diversification et d'une spécialisation des cellules de l'organisme. La morphogénèse d'un individu est une caractéristique propre à chaque espèce. A partir d'une cellule unique, la genèse de la

forme correspond à un processus complexe d'émergence que nous voudrions comprendre à tous les niveaux d'organisation et d'observation. Au XIX^{ème} siècle, les naturalistes ont exploré la notion de morphogenèse comme une description morphologique à au moins trois niveaux d'organisation : global, tissulaire et cellulaire. Cette approche structurale et morphologique est essentielle à l'embryologie classique comme à la biologie du développement contemporaine. Nous pouvons certes identifier une période de l'histoire où les approches à l'échelle moléculaire et génétique se contentaient de références morphologiques sommaires. Mais envisager la morphogenèse dans le champ des sciences de la complexité revient à intégrer tous les aspects de la morphodynamique cellulaire avec les descriptions moléculaires des vingt dernières années.

Le développement embryonnaire est le résultat d'une dynamique complexe qui fait intervenir différentes échelles d'intégrations. Nous pouvons décrire ce système complexe par la figure 1.20.

L'étude du système consiste ainsi à décrire deux informations :

- des faits stylisés caractéristiques de la dynamique du système à chaque niveau d'intégration (relatif aux fonctions D_i)
- des fonctions de *coarse graining* nécessaire pour passer d'une échelle de description basse à une échelle de description haute (les fonctions ϕ_i).

Le passage d'un niveau bas à un niveau haut implique une perte d'information. Par exemple au niveau tissulaire nous n'avons plus accès aux trajectoires des cellules mais seulement à des trajectoires moyennes. Par contre, ce passage vers un niveau de description plus grossier permet d'observer l'*émergence* de comportements nouveaux. Ils sont la conséquence d'interactions au niveau bas qui permettent à une population d'avoir un comportement collectif organisé. On peut expliquer par de tels phénomènes la cohésion d'un troupeau ou d'un banc de poisson sans que ceux-ci aient besoin d'un chef ou d'un champ extérieur pour se diriger (Gregoire (2001)). Par ailleurs, la présence d'un niveau d'intégration au dessus d'un autre n'est pas neutre pour les interactions du niveau bas. Le niveau haut peut imposer des contraintes sur le niveau bas, c'est l'*immersion*. La formation d'un épithélium (structure de haut niveau) qui contraint le comportement des cellules sont par exemple une illustration de cette caractéristique chez l'embryon du poisson zèbre.

Les données du projet nous donnent accès non seulement aux mouvements et aux déformations individuelles des cellules mais aussi aux événements morphogénétiques à l'échelle de l'embryon. Cela nous a permis d'envisager une description multi-échelles systématique des événements morphogénétiques : *l'embryome* (Annexe B). Nous avons donc mis en place des protocoles de mesures sur des données 4D du développement du poisson zèbre. Les mesures réalisées nous donnent accès à des paramètres de natures géométriques et cinématiques permettant de décrire les interactions cellulaires et la formation des structures au cours du développement. Elles se situent à différentes échelles, d'un petit groupe de cellules jusqu'à la formation d'importants organes. Ces études descriptives de l'embryogenèse per-

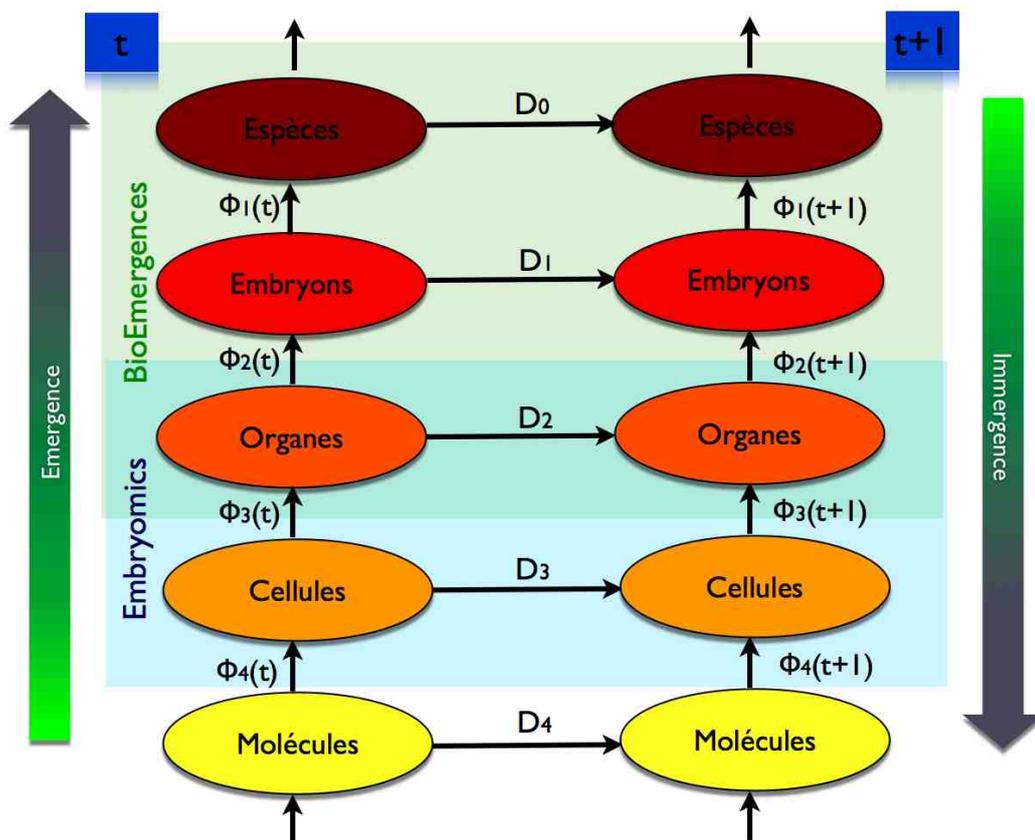


FIG. 1.20 – Schéma de principe représentant les échelles et les dynamiques en jeu dans le développement de l'embryon. Les données récoltées par le projet Embryomics permettent une étude concrète des niveaux cellulaire et tissulaire ainsi que de leurs interactions dans l'embryon vivant. Celles pour le projet BioEmergences se focalisent sur l'étude des différentes espèces et la comparaison des embryons à travers leurs organes. Les fonctions D_i représentent les dynamiques temporelles entre les niveaux. Les fonctions ϕ_i sont nécessaires pour passer d'une échelle à une autre.

mettent de s'interroger sur des notions de relations entre forces et comportements cellulaires, sur les liens entre interactions cellulaires et formation de structures (émergence et immergeance) ainsi qu' à une échelle plus élevée sur la description et la comparaison systématique des embryomes d'individus d'une même espèce.

Cet type d'approche permet d'expliquer l'embryogenèse des vertébrés en terme de mouvements collectifs de populations cohérentes de cellules qui correspondraient au concept biologique de champ morphogénétique. Nous faisons l'hypothèse que dans l'embryon, les mouvements collectifs de populations cellulaires émergent de mouvements locaux des cellules, de leurs interactions et de leurs fluctuations. La question de l'émergence du comportement cellulaire dans l'embryon de vertébré à partir des interactions à courte et longue portées relève d'un problème générique, transversal aux systèmes complexes, qui suppose la manipulation de larges séries de données spatio-temporelles et la détection de corrélations spatio-temporelles dans la dynamique des populations.

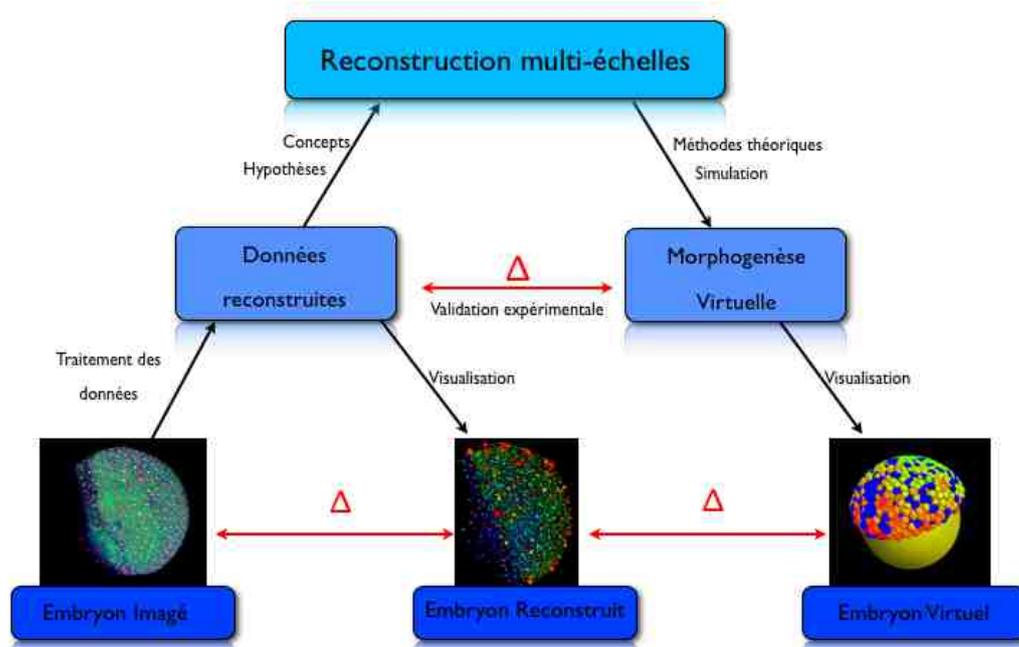


FIG. 1.21 – Représentation de la reconstruction multi-échelles de l'embryogenèse. A partir des données originales de l'embryon, nous effectuons un traitement des données permettant d'obtenir des données reconstruites. Grâce à des outils de visualisation, nous pouvons évaluer l'écart entre les données brutes et les données reconstruites. En utilisant les concepts et les hypothèses appropriés aux différentes échelles, nous pouvons ainsi reconstruire l'embryon par partie. De plus par des modèles théoriques nous pouvons simuler un embryon virtuel.

Ainsi, la notion de complexité intrinsèque au développement de l'embryon im-

plique de mettre en place une approche se situant de la même façon à différentes échelles (figure 1.21). Chacune de ces échelles comporte des interactions multi-niveaux entre les éléments, qui nous amène à devoir les décrire de façon rigoureuse par un meta langage, comme un système complexe. Nous avons créé un formalisme adapté aux différentes échelles du problème : l'identité cellulaire, les champs morphogénétiques, la carte des territoires présomptifs, le lignage cellulaire, etc.

Le développement de l'embryon au sein de la science des systèmes complexes peut être décrit comme

- une science expérimentale : la reconstruction multi-niveaux des dynamiques d'organisation.
- une science théorique : établissement de (nouveaux) concepts théoriques, des méthodes et des algorithmes pour traiter les questions théoriques en reconstruction.

1.5.3 Modèles mathématiques

Aujourd'hui, la science a identifié l'individu en le découpant en différentes parties : l'ADN et les gènes. Leurs règles d'interactions commencent simplement à être explorées. La prochaine étape pour comprendre les mécanismes liés à la vie est la compréhension des processus impliqués à différentes échelles. Cela consiste à avoir une vision globale du *où, quand et pourquoi* ces processus se produisent. De nombreux groupes ont récemment commencé à se concentrer sur ces problèmes à partir d'un champ de la biologie systémique (Varela (1979); Kitano (2002)). Ils essaient d'expliquer et de comprendre comment toutes les parties concernées (gènes, protéines, cellules ...) peuvent agir comme un système unique qui fonctionne correctement à des échelles différentes et qui, plus encore, est en mesure de produire un organisme complexe. La première approche théorique mathématique de la morphogenèse a été proposée par D'Arcy Wentworth Thompson en 1917 (Thompson *et al.* (1992)). Il a fait valoir que les formes géométriques euclidiennes offraient des solutions efficaces aux caractérisations morphologiques naturelles, comme la culture des mollusques ou des cornes en forme de spirale. En 1953, Alan Turing postule que des produits chimiques appelés morphogènes génèrent des organes quand ils sont présents à une densité suffisante (Turing *et al.* (1992)). Les organes, sous formes de structures spécifiques, seraient créés par mécanismes de réaction-diffusion. En 1976, René Thom en appliquant sa théorie des catastrophes donne une explication où les états biologiques plausibles sont des attracteurs stables du modèle de l'embryogenèse (Thom (1976)). D'autres tentatives de formalisation pour caractériser les processus de l'embryogenèse ont essayé d'introduire à différentes échelles des régulations internes et des pressions externes. En 1968, Aristid Lindenmayer a proposé les L-Systèmes (*L-Systems*), un formalisme mathématique pour modéliser la morphogenèse d'organismes multicellulaires (Lindenmayer (1968)). Les L-Systèmes permettent la reconstruction de structures d'arbre (les plantes, les artères, poumons, etc...) avec une règle de crois-

sance déterministe ou probabiliste. La forme récursive des règles des L-Systèmes conduit à l'auto-similarité et produit des structures fractales très semblables à des plantes naturelles.

Néanmoins, la reproduction de modèles d'organismes plus complexes, avec des comportements dynamiques, reste toujours un défi. Il est nécessaire d'inclure les facteurs environnementaux dans les modèles pour avoir une description plus réaliste. Par conséquent, des extensions des L-systèmes, sensibles à l'environnement ont été développés. Appliqués à la morphogenèse des vertébrés, les L-systèmes pourraient représenter un modèle abstrait des instructions codées par le génome. Ainsi, le modèle peut être complété par le rôle de la position de l'information, des comportements de cellules (adhérence, mobilité ...) et des pressions externes. Dans un cadre multi-échelles englobant une architecture modulaire, les L-systèmes pourraient permettre de découvrir de nouveaux phénomènes émergents.

En 1968, Lewis Wolpert a proposé le *French Flag Model* comme une illustration de l'importance de la position de l'information (généralement sous forme de gradient) dans la morphogenèse (Wolpert (1994)). Ce modèle conceptualise l'idée d'un *morphogène* précédemment définie par A. Turing, comme une substance chimique dont la distribution non homogène dans un tissu en développement provoque l'expression de gènes différents en fonction de la concentration de la substance. Dans ce modèle la concentration en morphogène dépend de la distance à la source et les différents degrés de concentrations conditionnent les différenciations cellulaires.

Les concepts de la position de l'information et du morphogène ont finalement été confirmés par les travaux de Christiane Nüsslein-Volhard (Nüsslein-Volhard & Wieschaus (1980), Ephrussi & Johnston (2004)) (Prix Nobel de médecine en 1995, avec Eric Wieschaus et Edward B. Lewis, pour leurs recherches sur le contrôle génétique du développement de l'embryon). Elle a identifié le premier morphogène (la protéine *bicoid*) agissant au début du développement de la drosophile (Driever & Nüsslein-Volhard (1988); Roth *et al.* (1989)). Sa traduction se fait après fertilisation, et sa production conserve une répartition sous la forme de gradient le long de l'axe antero-postérieur. Par conséquent, cette protéine morphogénétique détermine le développement de la structure antérieure de la larve et l'activation de la transcription de plusieurs gènes zygotiques dépendent de sa concentration le long de l'axe.

Bien que le le *French Flag Model* fût une représentation assez visionnaire, des travaux récents critiquent le phénomène de position statique, qui dépend d'un seuil de concentrations associées aux différentes couleurs (Jaeger & Reinitz (2006)). Le modèle *Drunken French Flag* propose de ne pas traiter la position de l'information seulement de façon spatiale mais en tant qu'information spatio-temporelle. Ce concept permettrait au modèle de mieux s'accorder avec les principes de dynamique de réseau.

Nous proposerons dans une perspective de simuler l'embryogenèse du zebra-

fish un modèle basé sur l'information spatiale et dérivé des L-Systems. Ce modèle, présenté dans l'Annexe C se base sur des ressorts comme en mécanique des fluides sur des ressorts, à partir de certaines caractéristiques d'information spatiale et héritage de l'information lors des divisions.

1.6 Reconstruction automatisée

1.6.1 Notion de *workflow*

L'organisation du travail est la discipline qui a pour but l'amélioration du rendement par une approche scientifique. Nombreux sont les domaines qui ont un apport à cette tâche, qu'il s'agisse des sciences humaines, économiques, ou encore informatiques. La réalisation d'un système de gestion de flux de travail (*workflow management system*) est l'objet d'un domaine de connaissances spécifiques.

De tels systèmes reposent sur l'utilisation des processus. Un processus étant un ensemble de tâches reliées entre elles pour en définir l'ordre. Lorsque nous ajoutons à cela la notion de ressource (l'entité capable de prendre en charge la réalisation d'une tâche), ainsi que sa gestion, nous obtenons un *workflow*. La *Workflow Management Coalition* (WFMC), organisme ayant pour but de standardiser les moyens mis en oeuvre pour l'implémentation des workflows en entreprise, définit cette notion de *workflow* de la manière suivante : « support ou automatisation informatisée d'un processus business, en tout ou en partie » (Coalition (1995)).

L'utilisation du terme « workflow » permet de délimiter d'une part la définition du processus, des ressources et de leurs attributions, et d'autre part l'utilisation dynamique de ces informations. Le « système de gestion de workflow », aura quant à lui pour rôle de gérer et exécuter l'utilisation du workflow.

La représentation d'un workflow consiste en une forme de cartographie des tâches. Plusieurs chemins peuvent y être empruntés pour l'objet de travail qui y cheminera.

Historiquement, si la place de l'informatique au sein des entreprises a connu un essor important au cours des dernières années, son utilisation dans le cadre spécifique de la gestion du travail aura été plus tardive. Auparavant, la règle d'or était « En premier lieu l'organisation, en second lieu l'informatisation. » (VanderAalst & VanHee (2000)). Désormais l'organisation et le suivi du travail sont des compétences déléguées à l'informatique, réduisant le gouffre entre les organisations et leurs systèmes informatiques.

Des outils tels que Staffware, Cosa, Windows Workflow Foundation, pour ne citer que les plus connus, sont des exemples de systèmes de gestion de workflow qui sont utilisés par un grand nombre d'organisations.

1.6.2 Un workflow pour la reconstruction de l'embryogenèse

L'automatisation de l'ensemble des traitements effectués sur une quantité massive de données permet d'obtenir une précision et une qualité qui ne seraient pas

possibles d'avoir manuellement. Ces dernières années, l'acquisition d'images et de nouvelles méthodes de traitements ont eu un grand impact dans la recherche bio-médicale, mais cela reste encore un domaine très ouvert (Megason & Fraser (2003)). L'état de l'art des méthodes de traitements d'images doit être redéfini et évolué afin d'aborder les nouveaux défis. De ce fait, au cours de la dernière année, il y a eu une énorme explosion de la quantité de documents consacrés à ce nouveau domaine de recherche "informatique bio-image" (Peng (2008)).

Nous avons conçu un *workflow* de traitement des données permettant d'obtenir la reconstruction d'embryons de façon rapide, simple et standard (Annexe A).

La reconstruction phénoménologique de la morphogenèse animale à partir d'imagerie *in toto* d'embryons de poisson zèbre a nécessité la mise en place de méthodes mathématiques et d'outils informatiques adaptés. Du filtrage à la reconstruction de l'arbre de lignage cellulaire, en passant par des outils statistiques, chacune des étapes du traitement des données embryonnaires est imbriquée pour arriver à la reconstruction totale. Afin de comparer différents embryons ou différents processus morphologiques, chaque étape du traitement est standardisée. Elles sont pour la plupart très gourmandes en ressources de calculs, nous avons donc dû utiliser plusieurs centres de calculs afin de pouvoir traiter les données dans des temps raisonnables. En sortie de chacun des traitements une très grosse quantité est produite, il a donc fallu mettre en place un système de gestion des données et de leur historique au sein du workflow. La mise en place d'une base de donnée a permis à certaines applications d'avoir un accès très rapide aux données et cela de manière transversale. Pour arriver à de tels objectifs, nous avons mis en place une standardisation des données d'imagerie, un typage des données, ainsi que des outils de visualisation permettant l'annotation d'éléments biologiques et leur évaluation.

1.7 L'apprentissage actif de l'embryogénèse

1.7.1 Qu'est-ce que l'apprentissage actif ?

L'objectif principal des systèmes d'apprentissages est de tirer profits des tendances générales à partir d'un nombre limité de données. Nous distinguons, dans la majorité des cas, deux principes : *l'apprentissage supervisé* et *l'apprentissage non supervisé*.

La tâche de *l'apprentissage supervisé* est de prédire un aspect supplémentaire à partir des objets données. Un point central de l'apprentissage supervisé est la tâche de classification où aux objets correspondent des valeurs discrètes appelées *labels*. L'objectif de la classification (grâce à un *classifieur*) est de donner un label à chaque objet inconnu à partir d'un certain nombre d'objets déjà manuellement étiquetés.

L'autre grand domaine des systèmes d'apprentissages est *l'apprentissage non supervisé*. La frontière entre l'apprentissage supervisé et non supervisé n'est pas entièrement nette, mais contrairement à la classification il n'y a pas de données manuellement étiquetées. L'apprentissage non supervisé fait un regroupement (*clus-*

tering) des données en mesurant leur similitudes et à partir de là permet d'établir un modèle sur leurs frontières.

Pour ces deux domaines d'apprentissages, nous devons en premier recueillir une quantité importante de données choisies au hasard parmi l'ensemble des données et ensuite attribuer un classifieur ou un modèle. Cette méthode est dite d'apprentissage passif (Figure 1.22a).

La première étape de collecte des données s'avère souvent être un travail long et gourmand en ressources. Il est, par conséquent, particulièrement intéressant de trouver des solutions adaptées nous permettant de minimiser ces ressources autant que possible. Nous supposons en général que nous pouvons piocher des données au hasard car elles sont indépendantes et identiquement distribuées. Cependant, dans beaucoup de situations, nous pouvons aiguiller le processus d'échantillonnage par l'observation de certains types de cas précis pris dans la base des données. Cela est appelé l'apprentissage actif. L'apprentissage actif rassemble des informations sur le monde en posant des questions et en recevant des réponses (Figure 1.22b).

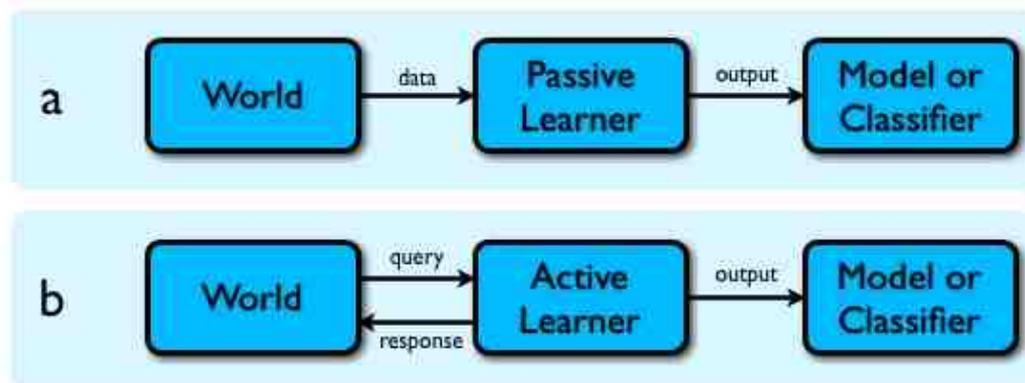


FIG. 1.22 – Schéma général pour le cas a : apprentissage passif et b : apprentissage actif

Nous distinguons deux types de requêtes possibles lors de l'apprentissage actif. Les requêtes par sélection où nous avons la possibilité de demander des données correspondant à un certain type spécifique. Ainsi, si par exemple chaque élément possède plusieurs attributs, nous pouvons demander seulement certains attributs sur l'ensemble des données. Les requêtes par sélection se prennent en général dans un réservoir de données (*pool-based*) (Lewis & Gale (1994)) où nous avons seulement un sous-groupe labélisé de l'ensemble des données.

L'autre possibilité est la requête par intervention, lorsque le système peut demander des éléments impliquant des interventions à effectuer. Dans ce cas, le système implique des dépendances probabilistes au modèle pouvant être remplacées par une intervention (Pearl (2000)). Par exemple, nous pouvons choisir quel type de nourriture donner à un rat de laboratoire, ce qui aura pour conséquence qu'il ne

mangera plus que le type que nous avons sélectionné et non plus ce qu'il voudrait manger en temps normal.

Nous pouvons maintenant introduire la notion d'évaluation du modèle (ou fonction de perte). Il est possible à chaque instant d'évaluer la qualité du modèle et donc d'en adapter la prochaine intervention. Il faut noter que cette approche est *myope* au sens où nous demandons une seule intervention spécifique basée sur cette mesure de qualité. Elle représente une approximation standard utilisée dans les problèmes de décision séquentielle (Horvitz & Rutledge (n.d.); Latombe (1991)).

1.7.2 Les impacts de l'apprentissage actif

Il y a eu plusieurs travaux portant sur l'apprentissage actif concernant la classification, la régression et les fonctions d'optimisations.

Dans le cadre de la classification, une méthode consiste à échantillonner un ensemble de classifieurs basés sur une distribution a priori d'hypothèses et de faire des requêtes d'exemples en fonction du degré de désaccord des différents classifieurs (Seung *et al.* (1992); Freund *et al.* (1997)). Cette méthode a été appliquée dans des domaines où la spécification à partir d'une distribution préalable est naturelle. Elles ont été utilisées avec des modèles probabilistes (Dagan & Engelson (1995)) et en particulier avec le *modèle naïf de Bayes* pour la classification de texte dans un cadre d'apprentissage bayésien (McCallum (1998)). Le classifieur naïf de Bayes fournit un modèle interprétable incorporant des valeurs manquantes en utilisant la connaissance a priori. Cependant, cela n'est pas toujours aussi performant que des méthodes *discriminatives* telles que les machines à vecteurs de support (SVM cf 1.7.3), en particulier dans le cadre de la classification de texte (Joachims *et al.* (1998); Dumais *et al.* (1998)).

Lewis et Gale (1994) introduisent la notion d'incertitude sur l'échantillonnage où le choix de la prochaine instance à classifier est peu certaine. Ils l'ont appliqué en utilisant la régression logistique (Lewis & Gale (1994)) et des arbres de décision (Lewis & Catlett (1994)).

Dans le cas de classification binaire, l'apprentissage actif appliqué aux SVM est similaire à la notion d'incertitude, mais fournit cependant beaucoup moins de justifications sur l'efficacité de l'algorithme.

Dans le cas de la régression, l'apprentissage actif a été introduit par Cohn *et al.*. Ils se servent de l'erreur quadratique de perte du modèle en tant que mesure de sa qualité et utilisent cette fonction pour choisir les requêtes qui réduisent la variance du système (Cohn *et al.* (1996)).

L'apprentissage actif a également été utilisé dans le cadre d'optimisation de fonction (Box (1987)). En 1998 Moore *et al.* ont introduit l'algorithme Q2 qui approche la fonction par une surface quadratique et choisit les requêtes sur les points les plus éloignés de ceux déjà demandés (Moore *et al.* (1998)).

Un autre domaine majeur est l'apprentissage par renforcement où les actions renforcent un état par observation de l'état actuel du modèle (Kaelbling *et al.*

(1996); *TD-learning* : Sutton (1988); *Q-learning* : Watkins & Dayan (1992)). L'état peut-être stochastique et ne garantit donc pas d'aboutir au même résultat. Contrairement à l'apprentissage supervisé, nous n'avons généralement pas de retour sur la conséquence des actions. Cependant, contrairement à l'apprentissage non supervisé, nous pouvons connaître l'impact d'une série d'actions en donnant une *récompense*. L'objectif est de trouver un système de gestion des actions de manière à maximiser la récompense. Si nous avons trouvé un moyen d'agir sur l'état en recevant une récompense raisonnable, alors devrions-nous continuer à exploiter ce que nous savons, en continuant à agir de la même façon, ou devrions-nous essayer d'explorer une autre possibilité du domaine de manière à agir dans l'espoir d'améliorer notre récompense? (*compromis exploration/exploitation* (Kearns & Koller (1999))).

1.7.3 SVM : Machine à Vecteurs de Support ou Séparateurs à Vaste Marge

Les machines à vecteurs de support (SVM) (Vapnik (1982); Boser *et al.* (1992)) ont de solides fondements théoriques et un succès empirique excellent. Elles ont été appliquées à de nombreuses tâches comme la reconnaissance de chiffres manuscrits (LeCun *et al.* (1995)), la reconnaissance d'objets (Nakajima *et al.* (2000)), et la classification de textes (Joachims (1999); Dumais *et al.* (1998)). La justification intuitive de cette méthode d'apprentissage est la suivante : si l'ensemble d'apprentissage est linéairement séparable, il semble naturel de le séparer de telle sorte que ses éléments soient le plus loin possible de la séparation choisie. Les SVM ont rapidement été adoptées pour leurs capacités à travailler avec des données de grandes dimensions.

Les séparateurs à vastes marges reposent sur deux idées clés : la notion de marge maximale et la notion de fonction noyau. La figure 1.23 donne un aperçu schématique de ces concepts. Ces deux notions existaient depuis plusieurs années avant qu'elles ne soient mises en commun pour construire les SVM. Les SVM sont une généralisation des classifieurs linéaires. Ils sont définis à partir de modèles linéaires dont l'apprentissage se fait en minimisant l'erreur empirique et dans le même temps, en essayant de maximiser une marge géométrique. Cette marge géométrique est la plus petite distance entre un point de l'ensemble d'apprentissage et la séparation réalisée par le modèle. Les échantillons les plus proches de la marge sont appelés *vecteurs supports*. Ce choix est justifié par la théorie de Vapnik-Chervonenkis (ou théorie statistique de l'apprentissage), qui montre que la frontière de séparation de marge maximale possède la plus petite capacité (Vapnik (2000)).

L'utilité des fonctions noyaux dans le contexte de l'apprentissage artificiel a été montré dès 1964. Afin de pouvoir traiter des cas où les données ne sont pas linéairement séparables, la deuxième idée clé des SVM est de transformer l'espace de représentation des données d'entrées en un espace de plus grande dimension (possiblement de dimension infinie), dans lequel il est probable qu'il existe une fonction séparatrice linéaire. Cela est réalisé grâce à une fonction noyau, qui doit

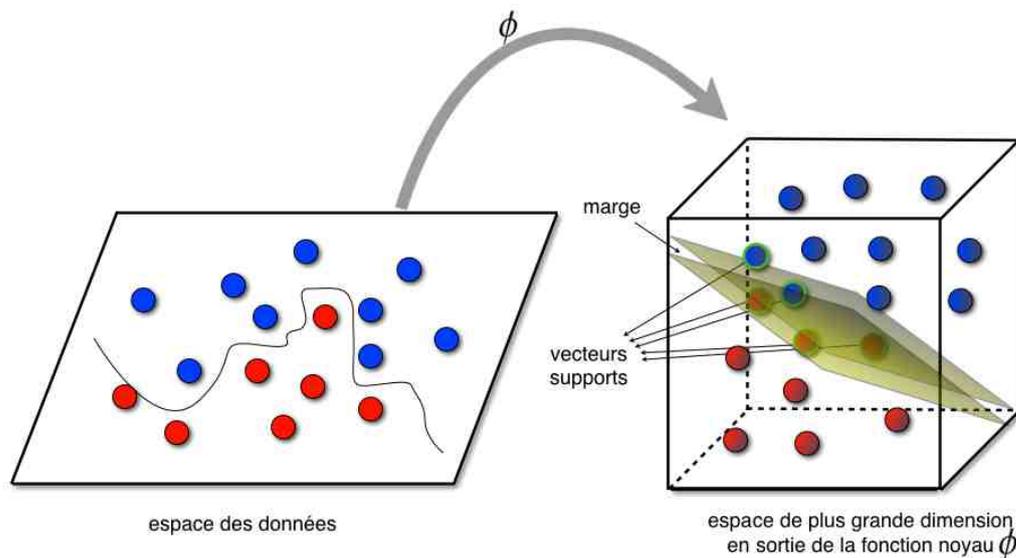


FIG. 1.23 – Représentation Schématique des deux idées clés des SVM : la notion de marge maximale et la notion de fonction noyau Φ .

respecter certaines conditions, et qui a l'avantage de ne pas nécessiter la connaissance explicite de la transformation à appliquer pour le changement d'espace. Les fonctions noyaux permettent de transformer un produit scalaire dans un espace de grande dimension, ce qui est coûteux, en une simple évaluation ponctuelle d'une fonction. Cette technique est connue sous le nom de *kernel trick* (Aizerman *et al.* (1964)).

Tong et al ont effectué de l'apprentissage actif en faisant les requêtes dans les données classifiées par les SVM, cela afin de réduire la complexité algorithmique (Tong & Koller (2002) ; Tong & Chang (2001)) .

Le lecteur pourra se référer pour plus de détails techniques aux articles de V. Vapnik : Vapnik (1982) ; Boser *et al.* (1992) et à son livre très complet : "The Nature of Statistical Learning Theory" (Vapnik (2000)).

1.7.4 Apprentissage actif de l'embryon

Il est important maintenant de comprendre pourquoi la reconstruction de l'embryon nécessite des outils basés sur un apprentissage actif.

Comme nous l'avons vu en 1.5.2, l'embryon peut être assimilé à un système complexe, et sa reconstruction à la science des systèmes complexes. La notion de structures émergentes des niveaux inférieurs est partie intégrante de la morphogenèse. Jusqu'à aujourd'hui les biologistes ont annoté à la main l'ensemble des cellules spécifiques caractérisant les champs morphogénétiques, cela afin de créer une carte des territoires présomptifs. Or la caractérisation d'un champ mor-

phogénétique est encore un concept assez peu connu. Quels sont les facteurs qui rentrent en jeu dans cette caractérisation ? Les agencements cellulaires, le mouvement des cellules, l'orientation des divisions, la forme des membranes etc., personne ne le sait encore vraiment même si des prémisses émergent dans certains travaux spécifiques. De plus les biologistes savent très bien détecter les champs morphogénétiques lorsque le développement de l'embryon est déjà bien évolué, mais lorsqu'il est encore précoce, les frontières sont difficiles à déterminer. Un système d'apprentissage actif peut alors apprendre à partir des annotations précédemment effectuées par les biologistes et les interroger lorsque le système n'est pas sûr de pouvoir bien caractériser une frontière. Cela afin d'obtenir un modèle et ainsi l'ensemble des caractéristiques principales qui constituent un champ morphogénétique. A travers cet exemple, nous pouvons extrapoler ce besoin de mettre en place des systèmes d'apprentissages actifs à toute notion de structures émergentes et cela à des niveaux multi-échelles et multi-temporels.

Les différents éléments de l'embryon sont considérés comme uniques au sens où ils ont des comportements individuels spécifiques. Or nous voulons lors de la reconstruction de l'embryon être capables de faire des *catégories* d'éléments. Des techniques de regroupement des éléments semblables doivent donc être mis en place. Nous pouvons pour cela utiliser des techniques d'apprentissage *non supervisé* qui peuvent permettre de faire émerger des *clusters* d'éléments distincts. Or les éléments de l'embryon sont soumis à beaucoup de variabilité dues en premier à leurs fonctions biologiques intrinsèques mais due aussi aux erreurs de post-traitement possibles. Il faut donc que le biologiste vienne en aide à ce type de regroupement en *labélisant* certains éléments. De plus comme nous le décrivons dans l'Annexe A le processus de reconstruction nécessite de traiter un très grand nombre de données. Il devient donc impossible pour le biologiste d'annoter suffisamment d'éléments afin d'avoir une partition de données connues suffisante pour effectuer un apprentissage *supervisé*. Il est alors évident que des outils d'apprentissage actif doivent être mis en place afin d'être capable d'aider la catégorisation des éléments .

1.7.5 Reconnaissance automatique des cellules en divisions

La reconstruction du lignage cellulaire passe par la construction d'un arbre binaire ou les noeuds de jonctions de l'arbre représentent les divisions cellulaires (Figure 1.25). La détection des mitoses est donc un point primordial dans la reconstruction de l'embryogénèse. En outre, l'identification des mitoses est elle-même un outil essentiel dans la mesure de l'effet anti-prolifératif des drogues testées. Mais lors du suivi des cellules, elle reste cependant une étape difficile. En effet, la vitesse de déplacement des cellules après la cytokinèse et leurs changements de voisinage sont une source d'erreurs importantes. L'identification des mitoses indépendamment du lignage doit permettre la conception d'un nouveau type de suivi cellulaire de façon plus précise.

L'acquisition d'imagerie par microscopie bi-photonique faite à l'échelle de

l'embryon ne permet pas d'observer au niveau intra-cellulaire les différentes phases de la division cellulaire tel que nous l'avons présenté dans 1.3.1. Mais elles sont reconnaissables dès la prophase grâce à la condensation de la chromatine. Cette condensation génère une hausse de l'intensité dans l'image et une déformation de la forme du noyau (figure 1.24). Dès lors, il est possible d'utiliser ces caractéristiques pour différencier la cellule prête à se diviser des autres.

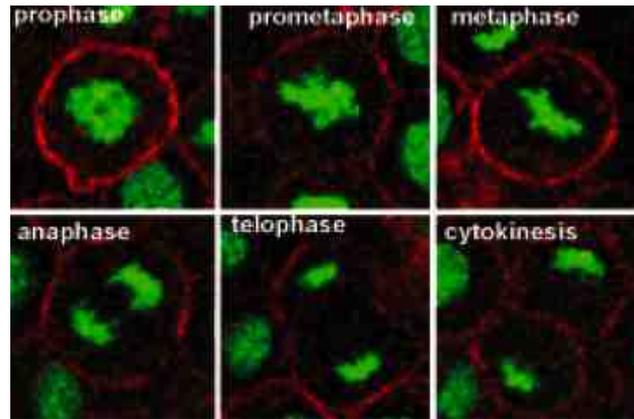


FIG. 1.24 – Les différentes phases de la cellule en division observées dans le contexte d'acquisition par imagerie bi-photonique à l'échelle de l'embryon.

Dues aux avancées récentes en microscopie, la détection et la classification des phases de la division cellulaire constitue un champ de recherche relativement récent.

Murphy et al. ont utilisé des réseaux de neurones pour classifier et détecter les différentes caractéristiques des structures sub-cellulaires sur un type de cellules cancéreuses en culture en microscopie à fluorescence (Boland & Murphy (2001); Boland *et al.* (1998); Chen & Murphy (2006); Murphy *et al.* (2003)). En 2004, Gallardo et al en utilisant des *modèles de Markov cachés* (HMM) permettent d'améliorer de quelques pourcents la reconnaissance des cellules en divisions (Gallardo *et al.* (2004)). Un modèle utilisant les informations de contexte a été proposé, toutefois, sa performance est fortement tributaire de la robustesse de l'algorithme de suivi des cellules en fonction du nombre de cellules. (Wang *et al.* (2007))

Un système automatisé, basé sur la classification par des SVM a été proposé pour le classement des différentes phases de la division cellulaire en utilisant une base de connaissance précédemment labélisée (Tao *et al.* (2007); Harder *et al.* (2006)). Wang et al en 2008 ont proposé d'utiliser des SVM sur la base d'un apprentissage actif permettant d'effectuer une classification en ligne des différentes phases de la cellule en division (Wang *et al.* (2008)). Ils proposent un modèle pour décrire la forme de la segmentation du noyau des cellules. La qualité des images sur lesquelles ils travaillent est cependant très bonne et leur algorithme en est fortement dépendant.

Bien que certaines méthodes ont été proposées pour la reconnaissance des

phases du cycle cellulaire, leur classification automatique reste encore un problème difficile à résoudre pour la biologie cellulaire utilisant la microscopie à fluorescence (Zhou & Wong (2006)).

En raison de l'objectif principal du projet qui est de reconstruire l'arbre de lignage sur le maximum de cellules possibles, nous faisons l'acquisition d'un très grand volume imagé. Cela entraîne une baisse de la qualité des images en profondeur et une grande variation dans la caractérisation des cellules. De plus, le poisson zèbre au cours de son développement présente rapidement un très grand nombre de cellules et pourtant un faible pourcentage d'entre elles est en division. Il devient donc trop fastidieux d'annoter à la main une base de donnée d'apprentissage suffisante à la détection. Dans ce contexte, les différentes méthodes proposées jusque-là ne s'avèrent pas assez robustes pour permettre une bonne reconnaissance des mitoses.

Dans le chapitre 2), nous avons tenté de faire le maximum d'économie d'une trop longue annotation d'un grand nombre de cas en élaborant un système d'apprentissage actif basé sur la catégorisation par les SVM et utilisant l'information de contexte temporel et spatial pour extrapoler les annotations. Nous utilisons différents types d'apprentissages permettant de prendre en compte toutes les informations connues et de minimiser le nombre d'annotations. Les taux d'erreurs de cette reconnaissance automatique sont mesurés par comparaison à un standard de référence (*gold standard*) correspondant aux annotations d'un expert. Cela nous a permis de catégoriser les caractéristiques de la division cellulaire (cf chapitre 4).

1.8 Reconstruction phénoménologique de l'arbre de lignage cellulaire

L'une des approches les plus puissantes à la compréhension du développement des organismes multicellulaires est la reconstruction des lignées de cellules. Au cours des dernières décennies, de nombreuses méthodes ont été élaborées pour suivre les cellules et leurs descendantes dans l'embryon vivant (Stern & Fraser (2001)).

L'étude des lignées de cellules a été et demeure d'une importance cruciale dans biologie du développement. La reconstruction totale de l'arbre de lignage cellulaire consiste à suivre depuis la première cellule, l'ensemble des cellules et leur descendance au cours du développement embryonnaire (appelé le *tracking* des cellules). Cette reconstruction paraît être un candidat naturel à la représentation des dynamiques spatio-temporelles.

Jusqu'à présent, la construction de l'arbre du lignage cellulaire était un travail fastidieux réalisé à la main qui ne pouvait être réalisé de manière exhaustive que pour des organismes simples comme le nématode *Caenorhabditis elegans* qui comprend environ un millier de cellules (Sulston *et al.* (1983)) (Figure 1.25).

Les systèmes d'imageries utilisés pour l'acquisition des données de notre projet permet d'obtenir simultanément deux canaux de fluorescence avec les noyaux

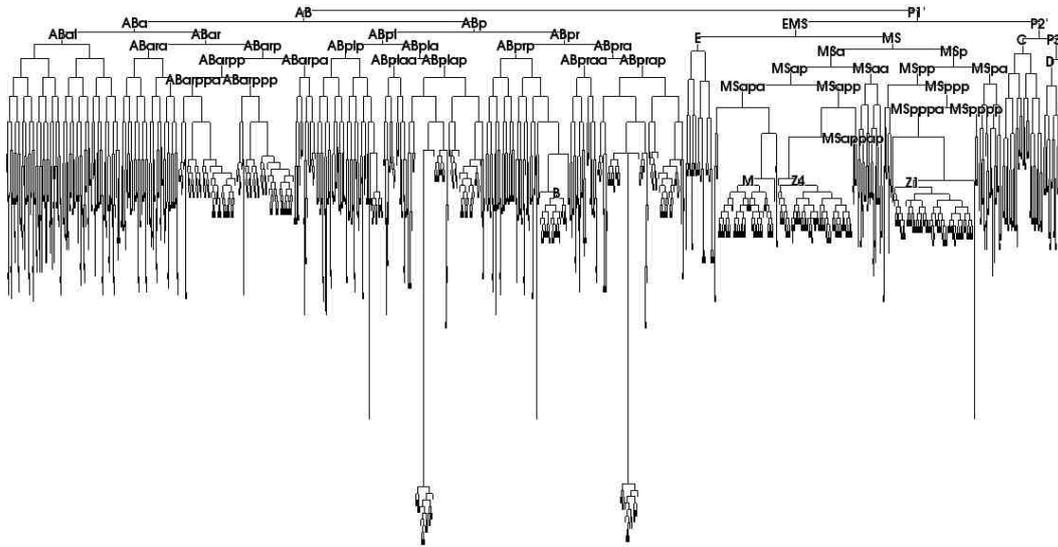


FIG. 1.25 – Arbre de lignage cellulaire complet du nématode *Caenorhabditis elegans* (de [Cedilnik et al. \(2007\)](#))

et les membranes des cellules durant le développement embryonnaire. Presque toutes les méthodes existantes pour effectuer le suivi des cellules vivantes ou autres composés moléculaires reposent sur la même stratégie universelle ([Meijering et al. \(2006\)](#)). Un élément n'existe pas de façon continue, mais est discrétisé dans le temps. Il faut donc être capable de reconnaître et de lier l'ensemble des éléments à travers les volumes imagés ([Zimmer et al. \(2006\)](#)). La majorité des travaux effectués pour suivre des éléments dans le temps sont effectués sur des cultures de cellules. Nous pouvons classer la littérature en deux sous groupes *suivi des particules par détection* ou *suivi par modèle évolutif*.

1.8.1 Suivi des particules par détection

Dans cette approche, les cellules sont d'abord détectées à chaque pas de temps, sur la base de l'information contenue dans les images (intensité, texture, gradient de caractéristiques, etc ..) ([Al-Kofahi et al. \(2006\)](#)), puis sont associées entre elles entre les différents pas de temps consécutifs. Cette approche est efficace et robuste en terme de calcul lorsque la densité cellulaire est relativement faible mais la détection des divisions cellulaires et leur suivies est souvent problématique ([Kirubakaran et al. \(2001\)](#)).

Par exemple, En 2005 Bhanson et al. ont établi un système automatisé de mesure de les déplacements cellulaires et de la prolifération dans le temps. Mais ce système est incapable de faire la distinction entre des cellules qui ne sont pas bien séparées ([Bahnon et al. \(2005\)](#)). Al-Kofahi et al. ont utilisé une méthode *seed-watershed* ([Vincent & Soille \(1991\)](#)) permettant de faire la distinction entre des

cellules très proches. Ils font ensuite à partir des caractéristiques des cellules un *matching* entre 2 pas de temps consécutif afin de déterminer les trajectoires cellulaires (Al-Kofahi *et al.* (2006)). Ils reconnaissent que le suivi devient difficile lors de forte concentration cellulaire et n'ont pas de correspondance pour les cellules qui entrent dans l'image.

Dans une autre étude, Yang *et al.* applique la méthode *watershed* (Cheng (1995)) au noyaux marqués par fluorescence afin de suivre la progression du cycle cellulaire (Zhou & Wong (2005)), mais ne font pas la reconstruction du lignage cellulaire..

D'autres méthodes assez populaires (Smal *et al.* (2006) ; Smal *et al.* (2007) ; Godinez *et al.* (2009)) se basent sur le filtrage des différentes particules (Doucet & Ristic (2002)), et utilisent leur détection au travers des probabilités *Bayésiennes*. Bien que ces techniques soit bien adaptées pour le suivi des particules, leur extension au suivi des cellules n'est pas très adapté.

1.8.2 Suivi par modèle évolutif

L'approche des modèles évolutifs tient compte de l'évolution de la forme des cellules entre chaque pas de temps (Debeir *et al.* (2005) ; Zimmer *et al.* (2002b) ; Zimmer & Olivo-Marin (2005) ; Mukherjee *et al.* (2004)) ou dans des volumes spatio-temporels (Padfield *et al.* (2006) ; Padfield *et al.* (2008) ; Padfield *et al.* (2009)). Les techniques basées sur des modèles de contours actifs permettent d'estimer de façon précise la morphologie des cellules, mais dépendent trop des paramètres pour s'adapter aux contacts entre les cellules et détecter les mitoses.

Par exemple, Zimmer *et al.* ont adapté le modèle classique du *serpent* pour permettre le suivi des cellules. Ils ajoutent des forces de répulsion entre les différents *serpents* pour permettre de gérer des contacts trop proches de cellules ainsi que l'intégration d'*opérateurs topologiques* pour accéder à la division cellulaire (Zimmer *et al.* (2002b) ; Zimmer & Olivo-Marin (2005)). Toutefois, les coûts en terme de calculs deviennent prohibitif pour le suivi d'un grand nombre de cellules.

Debeir *et al.* ont simplifié le problème de suivi en ne considérant que le barycentre des positions, sans tenir compte des contacts cellulaires (Debeir *et al.* (2005)). Cette approche semblable aux modèles des *serpents* ne permet pas toutefois de gérer divisions cellulaires. Les auteurs ont proposé comme solution de suivre les cellules en remontant le temps à partir de la dernière image (*backtracking*) et d'autorisé la fusion de cellules. Ce qui permet de simplifier considérablement le problème mais ne représente pas une solution adaptée pour le suivi des cellules en temps réel pendant l'acquisition des images. En outre, cette méthode exige une initialisation par l'identification manuelle du centre des cellules, et ne permet pas intégrer les nouvelles cellules qui arrivent dans l'images.

Des modèles de géométrie de contours actifs, grâce aux méthodes des *level set* (Sethian (1999)) ; ont récemment été mise en oeuvre pour permettre le suivi des cellules (Mukherjee *et al.* (2004) ; Yang *et al.* (2005) ; Dufour *et al.* (2005) ; Padfield *et al.* (2006) ; Padfield *et al.* (2008)). Ces modèles géométriques sont générale-

ment réputés pour être plus performant que les modèles paramétriques. Cependant, l'utilisation des *level set* dans sa forme classique n'est pas appropriée car elle permet la fusion de cellules trop proches et de plus elle se révèle très coûteux en terme de calculs. Pour remédier partiellement à ce problème de *fusion cellulaire*, Zhang et al. ont rajouté une contrainte au modèle empêchant le recouvrement des contours (Zhang et al. (2004)). Dufour et al. ont étendu cette approche en 3D pour le suivi des cellules fluorescentes (Dufour et al. (2005)) mais elle reste très limitée en terme de mémoire et de puissance de calcul, ce qui la rend inapte à gérer un grand nombre de cellules.

La plupart des études antérieures sur le suivi des cellules ne tiennent pas en compte les contacts cellulaires (Mukherjee et al. (2004)), ou ont recours à des post-traitement *off-line* pour corriger les fusions de cellules (Yang et al. (2005); Bunyak et al. (2006)). De plus ces méthodes utilisent peu les informations contextuelles. Padfield et al. ont abordé le problème du suivi par une tâche de segmentation spatio-temporelle (Padfield et al. (2006); Padfield et al. (2008)). Cette méthode peut être plus précise qu'un *matching* entre deux pas de temps, cependant elle nécessite la mise en place de méthodes additionnelles de post-traitement pour séparer les groupes de cellules (Padfield et al. (2008)). Elle est aussi beaucoup plus gourmande en terme de calculs que le traitement séquentiel.

Li et al en 2008 (Li et al. (2008)) ont proposé une méthode entièrement automatisée de suivi des cellules *in vitro*. Ils combinent l'information *bottom-up* et *top-down* de l'analyse d'image par l'intégration de plusieurs modules collaboratifs. de géométrie de contour actif (Osher & Sethian (1988)) Leur méthode a été testé sur un grande une variété de populations de cellules et ils obtiennent une précision entre 86.9 et 92.5%.

1.8.3 Suivi des cellules du poisson zèbre

Chez le poisson zèbre de nombreuses études ont été menées en marquant au cours du développement une cellule répondant à une protéine spécifique. Cela permettait de la suivre au cours du temps et de caractériser sa descendance, mais pas de reconstruire totalement l'arbre de lignage cellulaire (Hatta et al. (2006); Goldstein & Fishman (1998)).

Des travaux récents ont proposé la première reconstruction de l'arbre de lignage pendant les 24 premières heures du développement du poisson zèbre (figure 1.26) (Keller et al. (2008)). Les images des noyaux cellulaires de l'embryon ont été réalisées avec un nouveau type de microscope par sélection du plan illuminé (SPIM) (Huisken et al. (2004)). Elles donnent accès à l'ensemble de l'embryon au cours de cette longue période de développement. Ils ont effectué un simple mais efficace enchaînement du traitement des données, ce qui leur a permis de décrire de façon qualitative de nouveaux comportements cellulaires au cours des premiers cycles de divisions et d'en extraire les mesures statistiques. Néanmoins, leur processus de validation est incomplet, le suivi des cellules dans le temps s'avère ne pas être de bonne qualité, et les données brutes, ainsi que le l'ensemble du traitement

restent inexploitable pour reconstruire réellement l'arbre de lignage cellulaire.

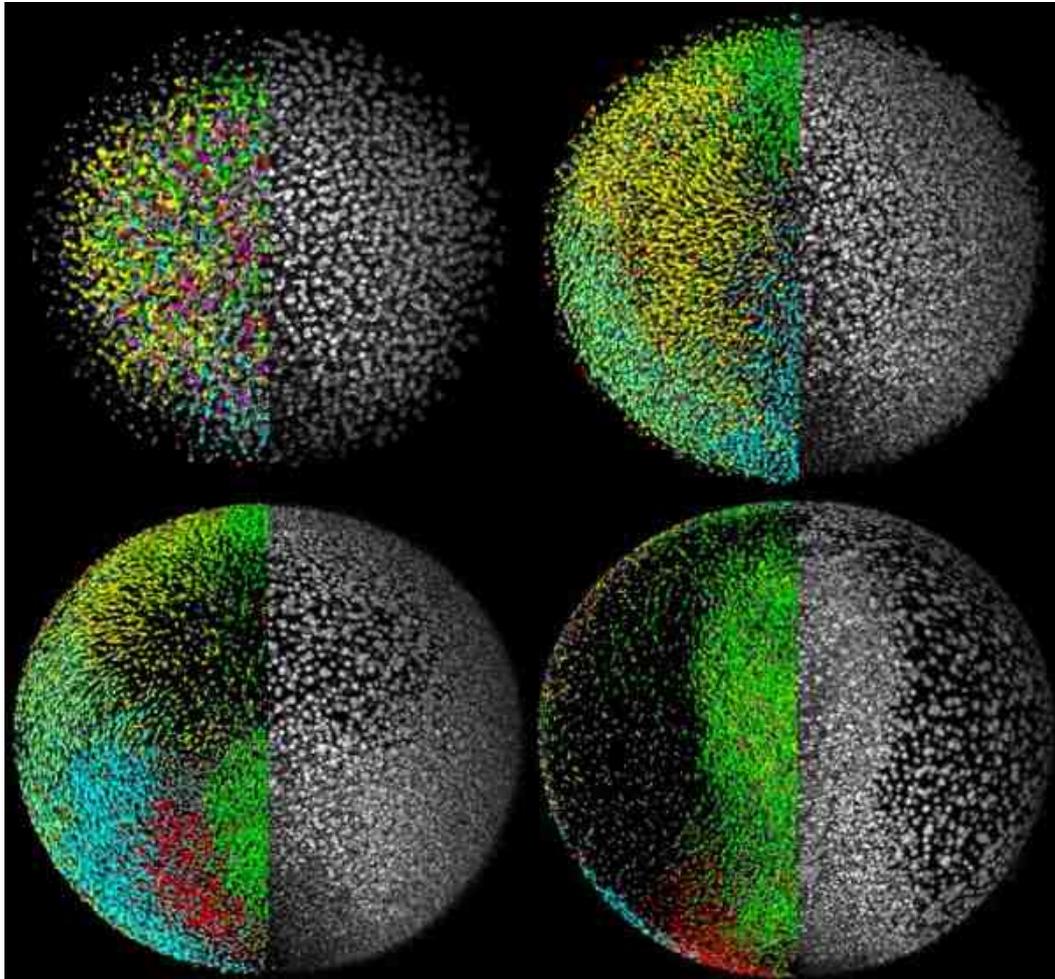


FIG. 1.26 – Suivi des cellules au cours de différentes phases du développement du poisson zèbre La coloration correspond à la vitesse de déplacement (de Keller *et al.* (2008)).

Des partenaires de notre consortium ont proposé (Melani *et al.* (2007)) à partir des bases du workflow (Annexe A) la reconstruction de l'arbre chez le poisson zèbre en utilisant des stratégies de minimisation locale et globale des informations. Cette reconstruction du lignage a été suffisante pour analyser certains phénomènes descriptifs qui ne nécessitent pas d'avoir un suivi des cellules complètement parfait. Lorsque nous avons été confrontés aux études à l'échelle cellulaire, comme l'analyse des clones, où aucune erreur n'est admissible puisque sinon elle changera la descendance de la cellule, nous avons amélioré nos techniques afin de ne pouvoir montrer au biologiste que la partie reconstruite dont nous étions sûrs. Nous avons donc mis au point dans le chapitre 3 une stratégie basée sur la recons-

truction de l'arbre de façon probabiliste. Un des problèmes majeur est que parfois l'algorithme de détection du centre des noyaux des cellules n'arrive pas à en trouver certaines, ou en détecte qui n'existent pas. Nous proposons alors de compléter les informations manquantes (Shalizi & Shalizi (2004)) en utilisant la connaissance de l'histoire passée et future des cellules (trajectoires, déformations, forme, etc.). Nous avons maximiser la vraisemblance des informations en faisant des itérations sur la connaissance préalable (Dempster *et al.* (1977)). Il s'agit d'utiliser les caractéristiques de voisinages cellulaires correspondant au lignage établi dans un premier temps pour en corriger les erreurs. Ces stratégies permettent d'améliorer de façon significative le suivi des cellules dans l'espace et dans le temps et de reconstruire ainsi l'histoire clonale des cellules. Ce que nous allons gagner en précision est essentiel à l'analyse de la morphogenèse et aussi à la validité et à la robustesse des protocoles que nous mettons au point pour le test des drogues anti-cancéreuses.

Active Machine Learning for Embryogenesis apply on cell division detection

Dans le chapitre 0, nous introduisons les notions clés à la compréhension de l'apprentissage actif (1.7.1), et y présentons l'état de l'art (1.7.2). Nous décrivons ensuite les Machines à Vecteurs Supports (SVM 1.7.3), et présentons les différents travaux sur la reconnaissance automatique des cellules en divisions (1.7.5).

Résumé

La reconstruction du lignage cellulaire passe par la construction d'un arbre binaire où les noeuds de jonctions de l'arbre représentent les divisions cellulaires. Lors du suivi des cellules, compte tenu de la vitesse de déplacement des cellules après la cytokinèse et leurs changements de voisinage, la détection des mitoses reste une étape difficile. L'identification des mitoses indépendamment du lignage doit permettre la conception d'un nouveau type de suivi cellulaire de façon plus précise.

L'acquisition d'imagerie par microscopie bi-photonique faite à l'échelle de l'embryon permet de reconnaître une cellule en division par sa condensation de la chromatine qui génère une hausse d'intensité dans l'image. La reconstruction automatique de la plus grande partie possible de l'arbre de lignage du poisson zèbre, nous confronte à beaucoup de variabilité dans la caractérisation de ces éléments rares. De plus, le poisson zèbre au cours de son développement présente rapidement un très grand nombre de cellules entraînant une impossibilité de tous les annoter.

Nous présentons dans ce papier une stratégie générale d'apprentissage actif avec une extrapolation temporelle et spatiale de l'information annotée. Le système utilise les vecteurs supports d'une catégorisation par les séparateurs à vastes marges. Nous utilisons différents types d'apprentissages permettant de prendre en compte toutes les informations connues et de minimiser le nombre d'annotations.

Cette stratégie peut être adaptée à l'observation et l'analyse des systèmes complexes, tels que les organismes multicellulaires. Nous nous focalisons sur la reconnaissance des formes afin de caractériser des patterns spécifiques à la morphogénèse : la détection des phases des cellules en division.

Les taux d'erreurs de cette reconnaissance automatique sont mesurés par comparaison à un standard de référence correspondant aux annotations d'un expert sur des oursins.

Sommaire

2.1 Introduction	54
2.1.1 Cell phase detection	55
2.1.2 Active machine learning apply on complex system	56
2.2 Active machine learning in the Embryomics project	57
2.2.1 Architecture of the automated Embryomics workflow	57
2.2.2 Assisting the Embryomics workflow with active machine learning	57
2.2.3 Annotation tool for supervised training	58
2.3 Active machine learning applied to cell division detection	59
2.3.1 Labels	59
2.3.2 Feature space of individual cells	59
2.3.3 The AML mitotic cell classifier	59
2.4 Results And Analysis	63
2.4.1 Measures	64
2.4.2 Parameters MDB and STM	64
2.4.3 Result depending on each step of the AML	65
2.5 Perspective and Discussion	66

2.1 Introduction

The goal of the European project *Embryomics* is to fully reconstruct in space and time the cell lineage of different species of deuterostome embryos, from the one-cell stage throughout embryogenesis. This task is crucial in understanding how organs are shaped during development and how they are clonally related. The genetic regulatory and biomechanical processes underlying spatial ontogenesis are still largely unknown. An automated identification of cell lineages and memberships to morphogenetic fields would prove extremely useful to major medical applications such as stem cell and cancer research. In particular, it could greatly contribute to the promising field of tissue and organ culture. Embryonic development in live animals can be viewed by time-lapse laser scanning microscopy of live embryos, which have been engineered with markers to highlight their cellular structures (Megason & Fraser (2003)). The type of images collected through this technology reveals the morphodynamics of cell division and pattern formation. We use various advanced machine vision methods and tools (filtering, detection, segmentation, tracking, matching, etc.) to process these images, linking them together as modules of a "workflow". At every step of this workflow, difficult perceptual and cognitive comparison tasks that would usually involve human learning must be performed. The main goal of the Embryomics project is to automate this process with the use of machine learning methods. In this paper we present global strategies for designing *active* machine learning methods suited for the observation and

analysis of complex systems, such as multicellular organisms. We pay particular attention to pattern recognition applied to the morphological features of embryogenesis, and give an example of cell boundary mitosis phase detection.

2.1.1 Cell phase detection

Quantifying the changes in cell cycle timing before and after drug treatment is useful for effective drug discovery research. Knowledge of the cell cycle progression, e.g., interphase, prophase, metaphase, and anaphase, is important in improving our understanding of the effects of various drugs on cancer cells (Yan *et al.* (2006), Zhou & Wong (2006)). Cell cycle progress can be identified by measuring changes in the nucleus and membranes as a function of time. Automated time-lapse fluorescence microscopy imaging provides an effective method of observing and studying nuclei dynamically and is an important quantitative technique in the fields of cell biology and systems biology (Zhou *et al.* (2006)).

Nevertheless, the vast amount and complexity of image data acquired from automated microscopy renders manual analysis unreasonably time-consuming. Accurate automatic classification of cell nuclei into interphase, prophase, metaphase, or anaphase, is an unresolved issue in cell biology studies using fluorescence microscopy. Murphy *et al.* (Chen & Murphy (2006), Boland *et al.* (1998)) have proposed different feature extraction, feature reduction, and classification algorithms for a similar problem of classification of subcellular location patterns in fluorescence microscope images. Methods have also been proposed to identify the cell cycle phase recognition. Gallardo *et al.* (Gallardo *et al.* (2004)) used Hidden Markov Models (HMMs) to classify the feature vector sequences that are extracted from the segmented potential mitotic cells. Chen *et al.* (Chen *et al.* (2006)) proposed an automated system to segment, classify and track individuals in live cell population, in which the KNN classifier with a set of seven features was used. A novel hybrid fragments merging method based on watershed segmentation and HMMs has also been proposed for cell phase identification (Yan *et al.* (2006), Zhou *et al.* (2006)).

A context-based model (Wang *et al.* (2007)) has been proposed to deal with this issue. However, its performance is highly dependent on the robustness of the tracking algorithm. When the cells are over-populated, the tracking algorithm becomes less reliable and the context information becomes less informative.

Wang and al propose an online SVM for cell cycle phase identification in cell culture (Wang *et al.* (2008)). Their method online learning algorithms acquire and analyze cell cycle behaviors of a population of cells. Here, online refers to the fact that the classifier will be continually updated by the misclassified samples to accommodate the ever-changing experimental conditions.

All of the above methods are efficient enough in their context of image acquisition (cell cultures, small volume in 3D, short time between two volumes, etc). Due to the aim of our project we acquire images in 4D (3D + times), over a long time period. We spend shortest time possible on each volume to minimize the bleaching effect, and used the largest volume size possible. Consequently the images

are noisy and the quality of each slice depend heavily on the position of the volume in z . Taking in to account all these parameters, we can access a large number of cells during 12 hours of development although the quality is very variable.

We create a online algorithm to detect the cells in division by classifying them in each different phase of the mitosis. Here, online refers to the fact that the classifier will be continually updated by the misclassified samples to accommodate the ever-changing experimental conditions. We apply a global strategy using the context information and extrapolate each manual annotation.

2.1.2 Active machine learning apply on complex system

Active machine learning (AML) algorithms combine automated classification with human labeling. They are used when a great number of examples are available but obtaining labels for them is costly, i.e., typically requires consulting a human expert (Bowring *et al.* (2004)) and/or sophisticated preprocessing methods. This is often the case with biological and social systems, whose intrinsic complexity can create large quantities of unlabeled data. Exploiting this kind of data can be approached by AML, during which a high-level symbol-naming expert is asked to define categories and trace best-guess boundaries based on as little data as possible. The automated procedure then consists of using the latest available knowledge (i.e., the state of the AML system) and correcting hypotheses in an iterative fashion when new data is added.

In this paper, the AML methodology that we propose involves four algorithmic steps. These steps and their application to biological cell classification are explained in greater detail in section 2.3.

1. Preliminary classification using expert knowledge
2. Data smoothing and clustering based on step 1
3. First active classification of a fixed data set :
 - a) AML system asks human supervisor to annotate data at class boundaries
 - b) AML extend label information in space and timee
 - c) AML system refines boundaries using new knowledge and loops back to 3(a) until it converges
4. Successive classification of new data points :
 - a) AML system classifies new data based on its latest state
 - b) AML system corrects class boundaries through step 3

In the remainder of this article, section 2.2.2 gives an overview of the algorithmic workflow used in the Embryomics project and the place occupied by AML in this context. Section 2.3 describes the above four AML steps as applied to cell classification. We then present and discuss our results in section 2.4.

2.2 Active machine learning in the Embryomics project

2.2.1 Architecture of the automated Embryomics workflow

The image processing workflow starts with 4D image acquisition (time-lapse series of 2D section stacks) and ends with the reconstruction of the cell lineage tree, i.e., the genealogy of all cell divisions (mitosis) and deaths (apoptosis) that occurred during the development of the embryo. Cell lineage can be represented as a forward branching process by tracking the 3D trajectory and fate of each cell. In this paper, we report experiments on the sea urchin embryo (Davidson (1993)). Due to the limited number of cells, it is possible to labelize all of them to easily validate our method. However the highest goal of the project is to reconstruct the lineage tree of the zebrafish. This is a fast developing vertebrate model easily accessible in its earliest stages, while its relatively large number of cells in later stages (50,000 after 24h) presents more challenges (Kimmel *et al.* (1995)). Acquired images are first passed through denoising and structure enhancement filters (Perona & Malik (1990), Drblikova (2005)), before the proper detection (Duda & Hart (1972)) and segmentation (Sethian (1999), Sarti *et al.* (2000)) of cell nuclei and membranes can take place. The figure 2.1 represent the rawdata, segmentation of both nuclei and membranes of the sea urchin. These preprocessing stages lead to the reconstruction of the embryo in 4D, and are followed by the temporal tracking (Horn & Schunck (1981), Melani *et al.* (2007)) of cell divisions, movement and death, which produce the final lineage tree. At this point, various interesting biological features and events at different scales, from individual cells (e.g., mitosis) to tissues and organs (e.g., deformations), can be classified and recognized. Ultimately, the workflow makes it possible to categorize cells into *morphogenetic fields*.

To implement these rules, we used a cell tracking algorithm. Our aim was to match each cell at time t to one (if it did not divide) or two (if it was dividing) cells at time $t + 1$. Our tracking method consisted of minimizing an intercellular difference measure, which combined physical distance between spatial coordinates and feature distance in the 42 features space described above. A temporal analysis of this tracking was then integrated in the clustering by applying the phase rules : previously labeled cells that created inconsistencies were removed from their clusters. This tracking-based correction was repeatedly applied until convergence.

2.2.2 Assisting the Embryomics workflow with active machine learning

One crucial step in the reconstruction of the cell lineage tree is the identification of dividing or "mitotic" cells. Classifier techniques can prove useful to correct or support machine vision methods. Apriori knowledge about the relevant features of cell division can be included via machine learning into the design of an artificial assistant, integrating detection and classification in a feedback loop.

This would allow for more precise predictions of the fate of mitotic cells and better tracking of their trajectories through successive division phases. Machine

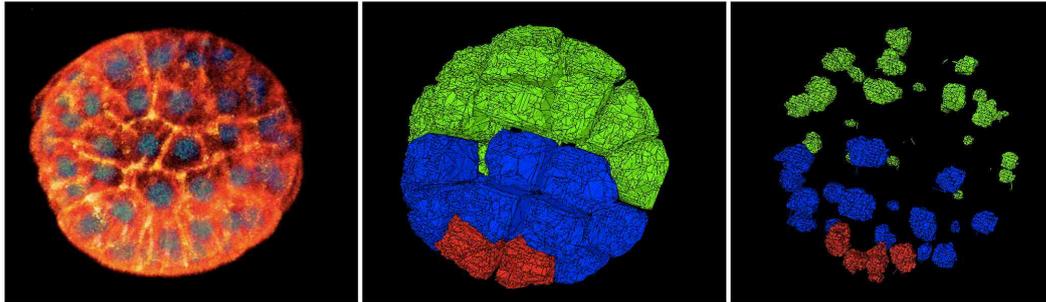


FIG. 2.1 – The whole embryo of sea urchin is filmed stage at 3h45 post fertilization (16cells stage) stage at 6h45 post fertilization (early blastula). At the beginning, we can only detect the membrane green staining, the red and green signals increase with the time. (mRNA injection at the one stage cell : H2B-mCherry, GFP-RAs (100 ng/ul each mRNA)). From left to right : raw data with both membrane and nuclei (cf video [Sea Urchin Rawdata](#)); membrane segmentation(cf video [Sea Urchin Membrane Segmentation](#)), nuclei segmentation (cf video [Sea Urchin Nuclei Segmentation](#)). The color in the segmentation pictures represent : red : micromere, blue :mesomere, green :macromere

learning integration constitutes a challenging task, partly because of the great variability of the segmented objects. However, a major benefit is the automation of the processes of hand tracking, visual inspection and cell counting that are extremely time consuming for biologists. We expect that machine learning and computation methods should soon rival the gold standard currently provided by the human experts.

2.2.3 Annotation tool for supervised training

For biologists to intervene during the AML process and visualize the ongoing 4D reconstruction at all times, a graphical platform was designed and implemented using open-source software (VTK, ITK, FLTK Kitware). Its interface was customized for the Embryomics project (cf video [Visualization and annotation of cells](#)). At the beginning the video depicts a red square for all centers detection of nuclei. Yellow lines then appear, which represent cells trajectories. Volume rendering of the original image is added in blue, and nuclei surface of selected cell is in red. We follow in time one cell in division, then volume rendering is changed for the ortho-slices to visualize the cell more accurately, which allows us to annotate their phases), to allow experts to annotate the segmented cells corresponding to a sub-region of the cell lineage tree, and interpret higher-level embryogenetic processes (morphogenetic fields, organs, macroscopic deformations, etc.).

2.3 Active machine learning applied to cell division detection

With the goal of producing the best lineage tree possible, the classification task consisted of automatically determining the mitotic phase of an unlabeled cell among seven possibilities (interphase, prophase, telophase, etc. ; see list below).

2.3.1 Labels

We distinguish two kinds of labeling, one which is manual, done by a human expert, these are assumed to be completely accurate, and one other labeling called automatic labeling, done by the extrapolation part of the AML program. This annotation are represented in the legend 2.3.

2.3.2 Feature space of individual cells

Each cell x was represented by a d -dimensional feature vector $x = \{p_1, \dots, p_d\} \in \mathbb{R}^d$. The list of d features was calculated using both the segmented data and the raw data, based on the cell's membrane and nucleus. The feature space typically contained the following dimensions : *volume, center of gravity, detected center, surface area, normalized shape index, axis ratios of included ellipse, including rectangular box, voxel intensity variables (max, min, sum, mean, standard deviation) within box and cell boundaries*. Features pertaining to image acquisition were also added, such as *cell age, microscope zoom, laser intensity, size and position of imaged volume*. The complete feature space was of dimension $d = 42$. All these features were used to compare any two cells at various distances, whether across the embryo or within local neighborhoods.

2.3.3 The AML mitotic cell classifier

The present task can be described as estimating a classification function $f : \mathbb{R}^d \rightarrow \{1, \dots, N\}$ where N is the number of classes, here $N = 7$ different mitotic phases (Figure 2.2) including *Interphase*. Only cells that were correctly segmented could be classified. The four-step AML methodology introduced in section 2 was implemented here as follows :

1. Preliminary classification using expert knowledge :

- a) A human supervisor using prior biological and physical knowledge initialized f by preclassifying cells in their corresponding classes. For this, the supervisor specified a set of constraints $K = \{k_1, \dots, k_N\}$ (one per class), calculated in feature space. Constraint k_i represented the boundaries of a specific partition in feature space, such that a cell x was put in class i (i.e., $f(x) = i$) if it satisfied constraint k_i . Biologists distinguish among six different mitotic phases - outside of *Interphase*, during

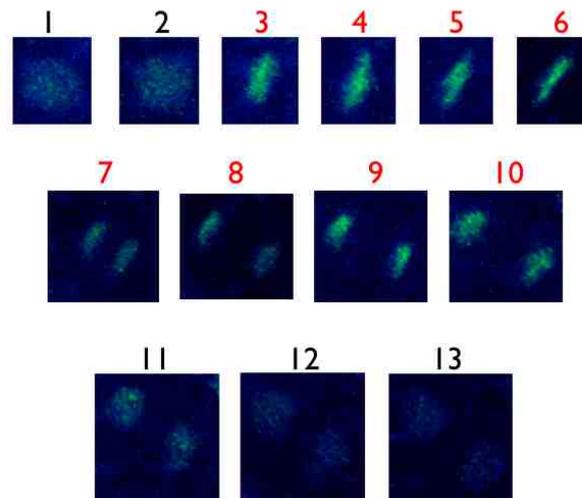


FIG. 2.2 – The evolution of a cell during phases of mitosis. In red is the cell respecting the principal first constraints of pre-characterisation. Image strategy : H2B/eGFP fusion protein staining the nuclei. $\Delta T = 2'35''$

which the cell is not dividing. Each phase can be translated into such constraints on the basis of their observed characteristics :

- *Prophase* : Fragmentation of the nucleus envelope and condensation of the chromosomes, along with changes in nucleus shape and increased brightness. There is also a deformation of the cell membrane from elliptic to spheric.
- *Prometaphase* : Changes in volume and convexity of nucleus shape.
- *Metaphase* : Planar arrangement of the chromosomes in the future cell division plane.
- *Anaphase* : Separation of the two sets of chromosomes yielding two nuclei within a single cell membrane.
- *Telophase* : Membrane constriction in the cell division plane.
- *Cytokinesis* : Separation of the daughter cells.

- b) In parallel to manual labeling, automated methods were also used to preclassify cells. For example, cells undergoing cytokinesis could be detected by an image processing algorithm based on surface evolution governed by nonlinear PDEs, or generalized "subjective surfaces" equations. This kind of numerical method is based on semi-implicit time discretization and finite volume space approximation Mikula *et al.* (2008a).
- c) We evaluate the time duration of mitosis phases. It appears on the observation of a large various of cell to be around 20 mn. We add the temporal information about mitosis evolution. It took the form of two competing constraints :
- **Phase conservation** : A cell in phase i at time t is likely to remain in

the same phase i at time $t + 1$.

– **Phase transition**

- A cell in *Interphase* at t is likely to transfer to the first phase of mitosis (*Prophase*) at $t + 1$.
- A cell in phase i at t is likely to transfer to phase $i + 1$ at $t + 1$.
- A cell in the last phase of mitosis (*Cytokinesis*) at t is likely to transfer to *Interphase* at $t + 1$.

2. **Data smoothing** : For each one of the seven subsets of points that belong to the same class, as labeled by step 1, a more homogeneous cluster was formed by normalizing the data and eliminating outliers. This was achieved by using a density clustering method based on support vector machines (SVM, Vapnik (1982)), in particular One-Class SVM clustering (Scholkopf *et al.* (2001)). All the data of one class was put into a hypersphere, which represented one of the seven clusters against all other cells.

3. **First active classification** : The quality of the reconstruction obtained thus far was increased by manually annotating specific cells near the class boundaries and extending this information automatically.

a) Through the visualization tool, the AML system asked the experts to annotate certain cells. It began with the points farthest from the boundaries. These represented the cells that were the most likely to change the boundaries of the cluster. At each annotation, clusters were recalculated in parallel. It finished with the cells closest to the boundaries. This had the effect of moving the boundaries to an optimal position. These cells were represented by the support vectors of a Multi-Class SVM classifier (Bennett & Demiriz (1999)). Practically, they corresponded to the points closest to the hyperplans that separated the classes. The new boundaries were computed after each reevaluation of a support vector that was incorrectly classified, .

b) We extend the manual labelization in space and time

– *space* : Similar cells have similar characteristics. In the features space, they need to be close. We define a variable **Maximim Distance Value** (MDV) correspond at an euclidian distance between two elements in the features space. After each manual annotation we extended this label to the nearest cells in the features space by an hypersphere until MDV. If during the expansion we met a cell previously manually labeled as being in another class, we stopped the expansion (figure 2.4 with the legend 2.3).

– *time* : The different phases of the cell division creates links between class in temporal evolution. Depending on the Δt between two volumes the phase transition rules are not completely inflexible. It is possible for a cell in phase i at t to transit to a cell in phase $i, i+1$ or $i+2$ at $t + 1$. After each manual annotation, a time expansion can be done in the other class though the future and the past time of the mitosis.

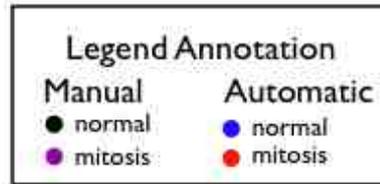


FIG. 2.3 – Legend of manual and automatic labélisation

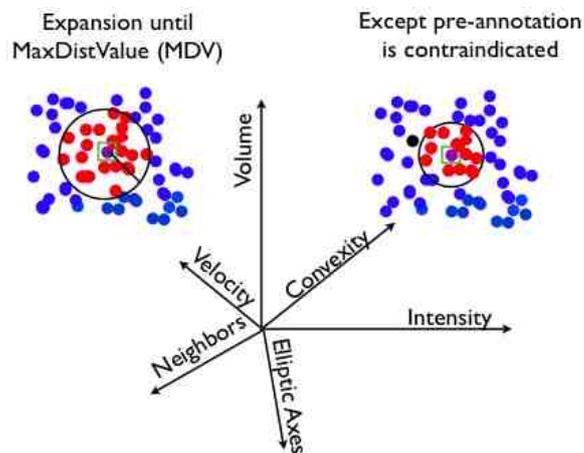


FIG. 2.4 – Example of expansion in the features space until Maximim Distance Value or until pre-annotation contraindicate the label.

We used the first tracking to extend the manual labeling in time, knowing that the two links created by the division are not done, but only at best one of these. We created a variable called " Segment Time of Mitosis " (STM) until the expansion could be done. Depending on the phases which is manual annotate and the difficulty to visually recognize the exact phases in our images, we can extend the information more in the futur or in the past. If during the future or past expansion time, we have a cell manual-labelling in the interphase class,we can extend the time-segment in the other direction. On the contrary any confirmation of previous manual labeling in the time-segment allowed us to extend the information in this direction (figure 2.5 with the legend 2.3).

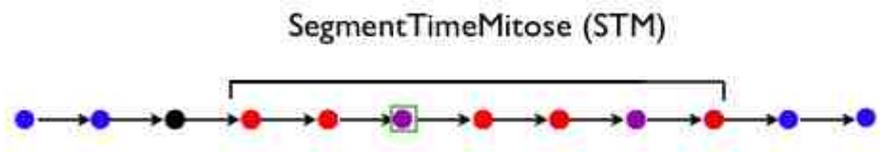


FIG. 2.5 – Example of expansion in time until Segment Time of Mitosis or until pre-annotation contraindicate the label.

- c) The different classes were recomputed until the supervisor agreed with the classification. The system eventually converged to the best margin between classes.
4. Successive classification :
 - a) A new 3D image was added and categorized using the knowledge of the previous Multi-Class SVM classification.
 - b) Classification was readjusted in step 3.

2.4 Results And Analysis

The goal of the AML methodology described in this paper is to automatically detect some pattern when it would be a high human cost to have a sufficient pre-label pool considering the huge number of elements and the difficulty to characterize them. We also use the AML to detect mitosis in the early embryogenesis of the zebra-fish where we are working with 10 millions of discretized elements. But it is impossible in this case to verify the results of our methodology. So to prove the efficiency of each part of the algorithm we use a completely labeled data set. In the sea-urchin embryogenesis we can access to some data-set where we are working with some thousands of elements.

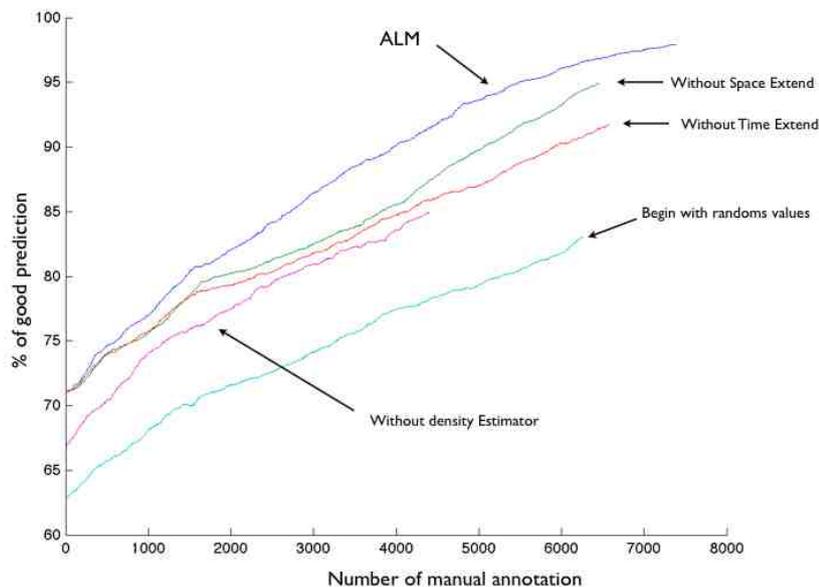


FIG. 2.6 – The evolution of the percentage of good prediction depending of the number on manual annotation. We compare the global AML methodology without each step of the procedure.

2.4.1 Measures

We need to create a measure of the efficiency of the methodology at each step of the learning procedure. The result function we use is defined by a percentage of the similarity divided by the differences. The advantage to have all labeled data is that at each manual labelization we are able to measure the increasing of the learning algorithm. The blue curve on figure 2.6 represent the result of the AML methodology depending on the number of manual annotation.

2.4.2 Parameters MDV and STM

We saw that in our methodology we have two parameters to define, corresponding on an expansion from a manual labeling to an automatic-labeling though a maximal distance in space (MDV) and a maximal segment in time (STM). As a first step we try different values of this parameters to obtain the best cluster. The video [MDV and STM 3D Evolution](#) shows the global result of the AML algorithms depending on these two parameters after each new manual annotation. The better values correspond to :

1. **MDV = 0.2.** This measure of similarities by an euclidian distance in the features space, normalized between 0 and 1, allow the algorithm to automatically annotate closest values.
2. **STM = 20 mn** The expansion in time correspond exactly to the real time duration of a cell in mitosis.

2.4.3 Result depending on each step of the AML

We measure the influence of each steps of the AML Methodology. We compare all the AML results without specific steps and analyze the impact on the global procedure.

1. Supervised Learning

To measure the impact of the first step of the AML algorithms, we apply all the AML methodology without using the first expert annotation. We create a random first labels data base using only the expected percentage of division. The result is represented in figure 2.6 by the cyan curve. Due to this respect of theoretical percentage of cell division, the algorithm starts with 63% good results corresponding to the cells which are not in division. At the beginning of the classification, it represents 8% less than the global strategy which gives the measure of this supervised step.

2. Unsupervised Learning

We measure the impact of the density estimator by reproducing all the AML algorithms without this step. The purple curve on figure 2.6 represents this measure. It begins at 67% good results, which represents 4% less than the global AML methodology.

3. Expansion effect

One important point in the AML methodology is the automatic expansion of each manual labeling. We have a look on the impact effect of both expansion with deleting each of these parts on the global AML methodology.

- **space** The green curve on figure 2.6 represents the result of the AML methodology without the extended space effect. The blue curve represents the AML complete with the best parameter of MDV. Due to the noise in the data, the different cells don't look very similar in the features space, but the extended space effect allows the AML algorithm to improve the results by a few percent, depending on the number of manual annotations.

- **time** The red curve on figure 2.6 represents the result of the AML methodology without the extended time effect. After a few annotations the time expansion effect improves the result by around 5%. Due to the time process of division, the temporal expansion is very important.

4. Self-supervised Learning

We had a new data set of sea urchin, which corresponded to a later part of the development. The whole embryo is filmed from 5h45 post fertilization until 9h post fertilization (corresponding to pre hatching blastula). The video [Second Sea Urchin](#) represents the raw data of this experiment. In figure 2.7, we compare the AML procedure starting with knowing the previous data (the green curve) and apply to this new sea urchin, to the AML procedure starting without any knowledge (the blue curve).

The self-supervised Learning part gave 83% good result without any annotation and increased the prediction during the first 3000 annotations. The stan-

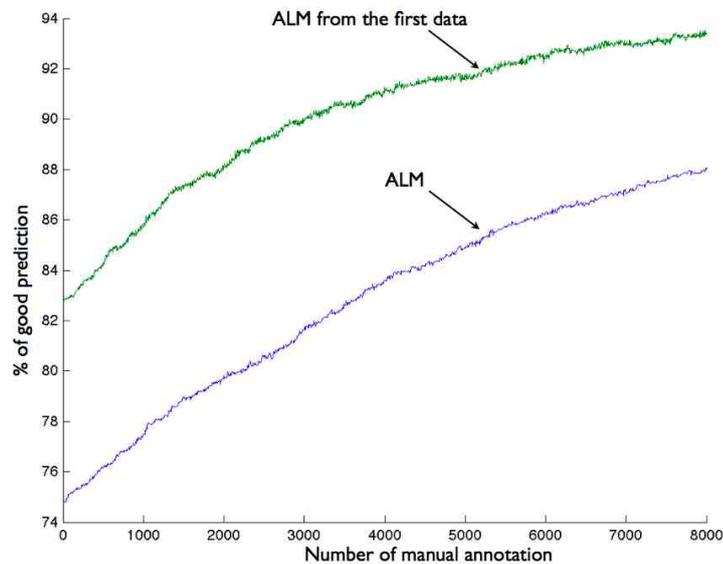


FIG. 2.7 – A new data set is add in the AML algorithm.

standard AML procedure started with 8% less good predictions. It proves that the algorithms can learn from a previous data even if it is not corresponding to the same period of development.

2.5 Perspective and Discussion

With the success on this test for validation, we applied the AML methodology to different zebrafish embryos. For each embryo, it involved over 4,000 cells followed during 10 hours (corresponding to 150 time steps). It allowed us to complete the Embryomics workflow and provide the first reliable lineage tree from a fully automated procedure. This had not been achieved in living embryos before. Our methods will now be systematically used for clustering morphodynamical patterns at several scales and measuring differences between individuals.

In future work we plan to add a last step to the AML process, a feature extraction algorithm that would distinguish and select the most relevant parameters for each clusters. It would increase the quality of prediction and reduce computation time. We also expect from this approach to be able to extract the most important biological features during the cellular division.

In addition, due to the difficulty in distinguishing the different mitosis phases by eye, ambiguous cells that could not be classified at step 3 should also be annotated by several experts in order to compare biological expertise.

Our ultimate goal is to design a European platform assisting biologists in the characterization of emergent phenomena. Such a goal involves designing an ex-

cellent artificial assistant that can combine various types of learning methods :

- bottom-up sensorimotor learning method capable of autonomously learning categories, prototypes and the morphodynamics of emergent phenomena through self-supervision
- top-down symbolic learning, including supervised learning of biological meaning of the emergent phenomena, and reinforcement learning for an efficient exploration behavior

MorphoTrack

Sommaire

3.1	Concept	70
3.2	Description	70
3.2.1	Problématique	70
3.2.2	Mesure de l'algorithme de <i>tracking</i>	72
3.2.3	Création d'un premier <i>tracking</i> par registration	75
3.3	Définition de <i>MorphoTrack</i>	76
3.3.1	L'algorithme EM	76
3.3.2	Mécanique des fluides complexes	79
3.3.3	Mouvement Brownien et processus stochastique	80
3.3.4	Ensemble statistiquement homogène	81
3.4	M- Maximisation	81
3.4.1	Définition des cellules et de leur <i>features</i>	81
3.4.2	Notion de voisinage cellulaire	82
3.4.3	Définition de la matrice de variance - covariance du mouvement cellulaire	83
3.4.4	Définition de la matrice de déformation du voisinage associée à chaque cellule	83
3.4.5	Détection des frontières	85
3.4.6	Cellules en division	86
3.4.7	Statistiques sur les mesures	88
3.5	E - Estimation	89
3.5.1	f , fonction de <i>tracking</i>	89
3.5.2	Fenêtre spatio-temporelle autour d'une cellule	89
3.5.3	Mesure de probabilités associées à f	90
3.5.4	Probabilité associée à une branche de l'arbre	91
3.5.5	Définition de la loi de probabilité associée	92
3.5.6	Probabilités associées au <i>tracking forward</i>	93
3.5.7	Probabilités associées au <i>tracking backward</i>	97
3.6	Résultats	97
3.7	Le langage <i>MorphoTrack</i>	103
3.8	Interprétation et Conclusion	103
3.9	Démonstration	105
3.10	Symboles	105

3.1 Concept

Tout processus de morphogenèse multi-cellulaire est géométriquement et morphodynamiquement parfaitement défini par le lignage cellulaire spatio-temporel. Il s'agit ici d'une affirmation simple : si à chaque instant, on connaît la position de chaque cellule, chaque division, chaque apoptose, il est clair que nous avons la géométrie de l'organisme multi-cellulaire à chaque instant et donc aussi sa morphodynamique. Les choses se compliquent lorsque l'on fait une reconstruction du lignage cellulaire spatio-temporel, pour des données bruitées, espacées dans le temps. Cette reconstruction se fait alors seulement en probabilité, plusieurs lignages cellulaires étant compatibles avec les données brutes.

3.2 Description

La technique d'acquisition consiste à prendre des volumes 3D d'images de l'embryon et de réitérer quelques instants plus tard. Suite à cette méthodologie, chaque embryon est discretisé temporellement. Il faut dans un premier temps pour chaque volume d'imagerie retrouver le centre des cellules. Mais à ce stade du workflow une cellule n'existe qu'à un pas de temps donné et n'a donc aucun passé ou futur. La définition d'une cellule au sens biologique avec sa propre histoire est donc à recréer. Il faut ensuite trouver les liens dans le temps entre ces cellules a-temporelles. On parlera, la plupart du temps, dans ce chapitre, d'une cellule comme le centre d'un noyau à un instant précis dans le temps et dans l'espace. Nous appelons, par anglicisme, le *tracking des cellules* la méthodologie qui consiste pour l'ensemble des centres à différents pas de temps, à recréer les liaisons temporelles entre ces cellules (à 'tracker des cellules').

3.2.1 Problématique

Lors de la conservation du nombre d'éléments entre deux pas de temps différents, le problème du *tracking* des éléments consiste à trouver en conservant certaines règles de topologie (maillage, voisinage, ec..) la correspondance d'*élément à élément*. Dans le cas où un élément manquerait à la correspondance (suite par exemple à une mauvaise détection des éléments à ce pas de temps), il *suffirait* de le rajouter pour avoir le suivi complet (passé et futur) de cet élément. Donc une petite variation dans la détection des centres ou dans le filtrage des images pourrait être *simplement* complétée en sachant que l'information est manquante. Et, de la même façon, une information rajoutée pourrait être effacée par l'analyse de la conservation du nombre d'éléments.

Les éléments dans notre cas précis sont des cellules. Ces dernières n'ont pas un comportement unimodal le long de leur vie (et donc de leur trajectoire) puisqu'elles peuvent se diviser (mitose) et même mourir (apoptose). Ces deux phénomènes n'ajoutent pas seulement deux contraintes supplémentaires à intégrer, mais font perdre la notion de conservation vue précédemment. Dans le cas où une cellule

serait en supplément à un pas de temps, alors est-ce une fausse cellule détectée par l'algorithme des centres ou bien une division cellulaire ? Nous avons ici plusieurs critères à prendre en compte pour obtenir le meilleur arbre de lignage cellulaire de l'embryon. Il faut tenir compte des performances et de la robustesse de chacun des algorithmes du workflow, ainsi que de l'altération possible du comportement cellulaire lors de l'imagerie.

3.2.1.1 Compromis de l'acquisition de l'imagerie

Les biologistes, lors de l'acquisition, doivent faire une multitude de compromis. La première année du projet Embryomics a consisté, en fonction du retour des performances de chacun des algorithmes, à trouver les meilleurs paramètres possibles. Les contraintes d'acquisition peuvent se résumer en quatre points :

- **Moyennage spatial** : chaque volume d'imagerie 3D est donc composé de plusieurs coupes en 2D. Le microscope a la possibilité de faire sur chacune des coupes plusieurs passages, ce qui permet d'éliminer le bruit dû à l'émission de fausses informations. Cela améliore considérablement la qualité de l'image mais prend une fois plus de temps à chaque passage.
- **Zoom et taille des voxels** : un embryon de poisson zèbre fait environ $700\mu m$ de diamètre et chaque cellule fait en moyenne entre 10 et $20\mu m$. Dans le cadre de la stratégie *multi-shot* (cf 1.4.5), nous voulons faire l'acquisition de la plus grande partie de l'animal possible. Il faut donc faire un compromis au niveau du zoom de l'objectif du microscope. De plus chaque pixel de l'image obtenue définit en réalité une taille en microns de l'embryon. Ainsi l'espace entre deux coupes en z peut-il différer selon les besoins. Par exemple, un noyau de cellule fait entre 2 et 5μ , donc si nous faisons une coupe tous les 5μ , il devient très difficile d'obtenir la forme du noyau et même de le détecter.
- **Durée totale d'acquisition** : le laser permettant d'illuminer une coupe chauffe considérablement l'embryon, aussi, plus on expose l'animal au laser, plus on détériore les cellules rapidement. Or, nous voulons faire une acquisition le plus longtemps possible pour pouvoir étudier pendant un maximum de temps le développement de l'embryon.
- **Délai entre chaque volume** : chaque coupe de 512×512 pixels prend environ 1s, sans faire de moyennage. Un volume imagé fait souvent plusieurs centaines de coupes, et donc son temps d'acquisition est loin d'être négligeable. Ainsi, entre deux acquisitions consécutives, selon les différents paramètres précédents, les déplacements cellulaires peuvent devenir important. De plus si nous ne voulons pas que les cellules subissent de façon trop importante la chaleur émise par le laser, nous sommes obligés entre deux acquisitions de volume de laisser l'animal se reposer.

3.2.1.2 Détection des centres variables

La détection des centres est une étape cruciale dans le *tracking* des cellules. Nous utilisons ici des algorithmes adaptés à la morphologie du noyau de la cellule et configurés à ses caractéristiques. Mais plus on descend dans le volume de l'embryon, plus le capteur du microscope doit traverser de couches cellulaires pour faire l'acquisition de la coupe. Cela diminue considérablement (et même linéairement) l'intensité dans l'image. Quand le signal est faible, le bruit augmente, et s'il devient trop important l'algorithme de détection des centres peut créer des centres qui n'existent pas en réalité. On trouve aussi des cellules avec des noyaux relativement grands (cellules épithéliales) et une variation de l'intensité à l'intérieur de noyau. Cela peut engendrer une détection de deux centres pour la même cellule. Nous appelons les **faux positifs** des centres qui ont été détectés mais qui ne correspondent à aucun noyau de cellule.

Il peut s'avérer aussi que, si l'espace entre deux coupes est trop grand, on ne puisse pas bien séparer les noyaux les uns des autres et ainsi ne pas être capable de détecter le noyau des cellules. De plus le niveau d'intensité du noyau n'est parfois pas assez fort et donc l'algorithme de détection n'arrive pas à le détecter. Nous appelons les **faux négatifs**, les noyaux de cellules qui n'auraient pas de centres détectés correspondants.

3.2.1.3 Modification du nombre de cellules au cours du temps

Le problème des faux positifs et des faux négatifs ne peut pas être résolu par l'étude de la conservation du nombre d'éléments au cours du temps. Deux facteurs affectent cette modification du nombre de cellules dans l'image entre deux pas de temps :

- **Bord de l'image** : l'acquisition ne permet pas d'imager en totalité le zebrafish. Nous sommes obligés de ne prendre qu'une partie représentant une calotte de la sphère assimilée à l'embryon. Selon les différents mouvements cellulaires au cours de l'embryogenèse, des cellules peuvent rentrer ou sortir du volume imagé entre deux pas de temps. Il est donc possible de ne pas trouver de correspondance temporelle passé ou future pour les cellules situées au bord de l'image.
- **Division et mort cellulaire** : ces deux facteurs biologiques importants font qu'une cellule peut naître d'une division ou mourir d'une apoptose. Ce qui a pour conséquence dans le premier cas de devoir créer un double lien temporel entre les cellules filles et la cellule mère au pas de temps précédent, et, dans le deuxième, de ne pas créer de lien avec le pas de temps suivant.

3.2.2 Mesure de l'algorithme de *tracking*

La mesure de la validité d'un algorithme de *tracking* est elle-même un facteur relativement compliqué. Il ne faut pas en sous-estimer la problématique car, comme on le verra dans le paragraphe suivant, il est nécessaire pour la cohésion

biologique de la reconstruction de l'embryogenèse d'avoir une performance optimale. Nous avons dans un premier temps eu recours à des critères statistiques sur la performance d'un *tracking* en fonction d'hypothèses connues par les biologistes. Par exemple le taux de mort et division cellulaire nous ont permis d'avoir de premières estimations. Mais sur des événements rares comme ceux-ci, ces statistiques ne suffisent pas et ne permettent en aucun cas de savoir, même si on a le pourcentage désiré, si nous n'avons pas que des faux positifs (de division ou d'apoptose) et aucune détection des réelles informations.

Nous avons ensuite, à l'aide de mesures standards, estimé le déplacement moyen d'une cellule entre deux pas de temps, puis mesuré les performances du *tracking* en fonction de la déviation à la conservation de cette propriété. Mais là encore, ce n'était qu'une mesure estimative et en aucun cas une valeur précise sur l'exactitude des liens temporels. Car une cellule peut, contrairement à son voisinage, avoir un comportement très particulier et se déplacer rapidement. Si l'algorithme de *tracking* l'a bien suivie au cours du temps, la déviation de déplacement de cette cellule par rapport à la norme est élevée : on aurait donc pu statistiquement la considérer comme faussement trackée.

Ces premières notions de mesures sont par contre très utiles lorsqu'il s'agit de comparer des *trackings* entre eux. Car un *tracking* est plus performant qu'un autre s'il s'avère plus proche statistiquement des attentes biologiques.

Nous avons fini, après quelques mois de réflexion sur la finalité d'un *tracking*, par créer une mesure plus traditionnelle qui consiste à annoter un certain nombre de liens manuellement, et à vérifier les performances de l'algorithme par rapport à cette annotation. Il n'a pas été si simple de se rabattre sur cette mesure, pourtant apparemment si évidente, car, pour un seul embryon, nous avons des millions de liens à annoter si nous voulons le valider entièrement. Cela représente par conséquent une quantité de travail astronomique, qui plus est laborieux. De plus, un panel de quelques centaines de liens était amplement insuffisant car il ne permettait en aucun cas de prendre en considération des éléments rares comme les mitoses ou les apoptoses, ainsi que les variations spatiales ou temporelles.

De plus il s'est avéré que la plupart des algorithmes du workflow avaient besoin de retour pour mesurer leur validité et qu'un système global devait être mis en place. Nous avons donc développé l'interface de visualisation de la reconstruction de l'embryon et validé (ou invalidé) un grand nombre de liens sur plusieurs embryons qui possédaient des caractéristiques d'imageries différentes (cf A.2.4.2).

A partir de l'annotation manuelle d'un très grand nombre de lien sur certains embryons, il nous a donc été possible à chaque phase de l'élaboration de l'algorithme de *tracking* de valider et de mesurer les différentes méthodes proposées.

3.2.2.1 Performance à observer

Après que la procédure de la mesure du *tracking* a été mise en place, il a fallu, en vue d'optimiser les différents facteurs de l'acquisition biologique, savoir ce que l'on attendait des performances de l'algorithme de *tracking*. A la question "Est-ce

que 95% de liens valides suffisent?", on aurait dans un premier temps répondu intuitivement que "oui". Or, il s'avère, au vu des résultats de ce tracking, que cela dépend énormément de ce que l'on veut étudier au sens biologique.

Par exemple, si on veut suivre une cellule pendant plusieurs heures et être capable d'étudier la moindre variation dans son déplacement, il est très important d'avoir un *tracking* complètement juste (à 100%). Car même un "99%" entraîne que, sur 100 liens à la suite, il est fort probable d'avoir un mauvais lien, et donc, à un moment donné de l'histoire de la cellule, d'avoir interverti une cellule avec une cellule voisine et par conséquent de ne plus avoir aucune notion réelle de la définition des champs morphogénétiques. De plus, si on veut étudier la notion de temps de vie de cette cellule entre deux divisions, nous ne pouvons pas manquer une vraie division cellulaire ou ajouter une mitose qui n'existe pas en réalité.

Par contre, si on veut par exemple, *simplement* avoir des notions de mouvements cellulaires globaux, un *tracking* parfait n'est pas obligatoire.

Grâce à l'interface de visualisation, nous nous sommes rendus compte des principaux problèmes liés au *tracking*, mais nous nous sommes aussi aperçu que même un expert en biologie, après des heures passées à apprendre et à reconnaître le comportement cellulaire est parfois dans l'indécision pour annoter certains éléments spécifiques. Il nous a donc paru évident qu'une intervention humaine serait toujours nécessaire, au moins tant que les qualités d'images resteraient ce qu'elles sont aujourd'hui.

Au vue de ces exigences, et compte tenu qu'il est pour l'instant impossible d'avoir un *tracking* parfait en prenant en compte les techniques actuelles d'acquisition nous avons opté pour une évaluation probabilistique de la performance de l'algorithme de tracking.

Nous associerons donc une probabilité à chaque lien temporel créé. Cela permet d'apposer une validité automatique quasi certaine sur un grand nombre de liens et de faire ressortir les liens sur lesquels l'algorithme a des doutes.

Avec cette technique, nous restons dans la continuité de la méthode d'*Active Machine Learning*, vue au chapitre 2. Car chaque lien avec une probabilité faible peut être annoté par le biologiste et servir à la procédure de *tracking* elle-même pour améliorer ses performances dans une optique d'apprentissage.

3.2.2.2 Embryon utilisé pour mesurer l'algorithme

Pour valider chacun des points méthodologiques de cet algorithme de tracking, nous allons utiliser un embryon qui contient environ 3,4 millions de centres détectés. Nous avons sur cet embryon environ 80000 liens validés dont 1200 mitoses. De plus nous avons quelques sous-volumes complètement annotés manuellement, ce qui permet de vérifier aisément les faux positifs. Cet embryon contient 480 pas de temps espacés de $\Delta t = 1mn07s$. Le volume imagé fait $700\mu m \times 700\mu m \times 141\mu m$. Dans le cadre des acquisitions effectuées en général, cela représente un volume imagé standard avec un Δt entre deux images assez petit. L'avantage de faire les

tests de mesure sur cet embryon, est que nous pouvons augmenter la valeur de Δt en enlevant une image sur deux par exemple, ce qui nous permettra aussi de mesurer notre algorithme pour un $\Delta t = 2mn14s$.

3.2.3 Création d'un premier *tracking* par registration

Nous avons dans un premier temps créé un *tracking* par *registration* des cellules. Vu que, biologiquement, une cellule est contrainte à n'avoir qu'une et une seule mère, alors qu'elle peut avoir une ou deux filles, il est beaucoup plus facile dans cette partie de fonctionner avec le temps à reculons (*backtracing*), c'est-à-dire de partir du dernier volume temporel, et de remonter le temps jusqu'au premier volume.

3.2.3.1 Le centre le plus proche

Nous utilisons pour cela la méthodologie qui consiste à rechercher le point le plus proche dans le passé. Cela consiste pour un centre i à un instant t ayant pour coordonnées $\{x_i, y_i, z_i\}$ de chercher dans la listes des cellules au temps $t - 1$ le centre j ayant pour coordonnées $\{x_j, y_j, z_j\}$ qui minimise la fonction de calcul de distance euclidienne entre les deux points i et j dans un espace 3D.

Nous n'imposons pour l'instant aucune contrainte biologique ou physique, si bien qu'un centre a t peut-être lié, à $t + 1$, à un, deux centres ou plus (donc une mère peut avoir plusieurs filles). De même, nous n'avons pas imposé de contrainte de *backtracking*, donc une cellule peut avoir plusieurs mères.

Nous obtenons lors de ce première *tracking* :

- 98% de liens corrects,
- énormément de fausse mitoses,
- beaucoup de liens triples (ou plus).

Nous avons donc montré en premier qu'en utilisant une technique *simple*, comment il est possible d'obtenir un *tracking* relativement performant sans respecter aucune contrainte biologique. Nous voyons ici que certaines cellules peuvent se diviser en plus de deux cellules, et que le nombre de fausses divisions est extrêmement élevé. Effectuer un *tracking* biologiquement correct devient alors bien plus complexe.

3.2.3.2 Le centre le plus proche sous contrainte d'un seul et unique lien

Pour conserver les liens valides biologiquement, nous avons ajouté une procédure récursive qui choisi au pas de temps précédent le centre le plus proche sous réserve qu'il ne soit déjà affecté à un centre dont la distance serait inférieure à celle en cours.

Nous obtenons dans ce cas :

- 96% de liens correct.

– plus aucun double lien.

Nous avons donc baissé de quelques pourcent la performance du *tracking* mais respecté la cohérence biologique (bien que nous n’ayons encore détecté aucune mitose). Ce *tracking* a l’avantage d’avoir d’être très correct et de ne comporter aucune incohérence. Nous l’utilisons dans la suite comme point de départ, et allons essayer d’en améliorer ses performances. Il apparaît évidemment beaucoup plus facile et logique de travailler avec un tracking d’ores et déjà cohérent biologiquement.

3.3 Définition de *MorphoTrack*

Ce premier *tracking* cohérent biologiquement et valide à 96% est le point de départ de notre algorithme *MorphoTrack*. L’algorithme prend en entrée un *tracking* donné et fournit en sortie un *tracking* amélioré, c’est-à-dire qui comporte plus de liens valides. Nous itérons cette procédure, jusqu’à ne plus observer aucune amélioration.

A chaque itération nous définissons des groupes de cellules similaires. Sur ces groupes, nous effectuons un ensemble de mesures caractéristiques du comportement cellulaire et en dégageons des comportements types. Nous corrigeons les cellules ayant une trop forte déviation à ces standards de comportements, en modifiant leurs liens *mal* prédits par le *tracking* précédent.

Cette correction se fait à l’aide d’un ensemble de mesures probabilistes. Nous associons à chaque cellule une mesure de probabilité en fonction de sa déviation au comportement standard.

L’itération de l’algorithme est définie sur le modèle de l’algorithme d’Estimation-Maximisation (EM).

Tout algorithme EM met en scène des variables cachées. Dans *MorphoTrack*, les variables cachées sont précisément contenues dans la fonction de tracking f puisque les transitions à l’instant t de la cellule i à la cellule $f(i)$ à l’instant $t + \Delta t$ ne sont *pas observées*. Ces transitions sont seulement une probabilité de se produire. La fonction de tracking globale a une probabilité $p(f)$ de se produire.

La procédure d’estimation-maximisation est améliorante à chaque étape (McLachlan *et al.* (1997)). Elle converge donc, mais pas nécessairement vers le maximum de vraisemblance exact.

3.3.1 L’algorithme EM

Nous allons rappeler ici la logique de l’algorithme EM, qui est très proche de *MorphoTrack*, ce que nous montrerons en conclusion. L’algorithme estimation-maximisation (en anglais *Expectation-maximisation algorithm*, abrégé EM) a été proposé par Dempster *et al.* en 1977 (Dempster *et al.* (1977)). L’algorithme EM est une procédure itérative efficace pour calculer le maximum de vraisemblance (ML pour *Maximum Likelihood*) en présence de données manquantes ou cachées. Le ML estime les paramètres du modèle pour lesquels les données observées sont les plus

justes. L'estimation du maximum de vraisemblance est une méthode statistique courante utilisée pour inférer les paramètres de la distribution de probabilité d'un échantillon donné. Elle a été développée par le statisticien et généticien Ronald Fisher entre 1912 et 1922.

Chaque itération de l'algorithme EM se fait en deux étapes :

- **E - Estimation** : les probabilités des données manquantes sont *estimées* à partir des données observées.
- **M - Maximisation** : on cherche les lois de probabilité qui maximisent la vraisemblance des données observées.

La convergence de l'algorithme est assurée sous garantie que le ML augmente à chaque itération.

3.3.1.1 Définition de l'EM

Nous allons donner les éléments de bases permettant de comprendre l'algorithme EM. Nous associerons, en *italique*, à chaque définition, la correspondance des paramètres dans notre étude du calcul de l'arbre de lignage cellulaire le plus probable.

Soit X l'ensemble des données (*les caractéristiques sur l'ensemble des cellules*). Nous devons trouver θ (*les paramètres des lois de probabilité*) qui maximise la probabilité $P(X|\theta)$. Cela revient à estimer le maximum de vraisemblance (*le meilleur arbre de lignage cellulaire*) avec θ comme paramètre et X comme variable donnée. Afin d'estimer θ , il est commun d'introduire la fonction du logarithme népérien du maximum de vraisemblance, définie par :

$$L(\theta) = \ln P(X|\theta) \quad (3.1)$$

Comme $\ln(x)$ est une fonction strictement croissante, la valeur de θ qui maximise $P(X|\theta)$ maximise aussi $L(\theta)$. L'algorithme EM est une méthode itérative pour maximiser $L(\theta)$. L'estimation de θ à la $n^{\text{ème}}$ itération est notée θ_n . Nous notons z la partie cachée ou manquante des données (*les liens temporels*). On a alors, par la formule de Bayes,

$$P(X|\theta) = \sum_z P(X|z, \theta)P(z|\theta) \quad (3.2)$$

On démontre ensuite, grâce à l'inégalité de Jensen pour la fonction logarithme, que

$$\begin{aligned} L(\theta) &= \ln(P(X|\theta)) \\ &\geq \sum_z P(z|X, \theta_n) \ln\left(\frac{P(X|z, \theta)P(z|\theta)}{P(z|X, \theta_n)}\right) \\ \text{qu'on note} \quad & l(\theta|\theta_n) \end{aligned}$$

On montre ensuite facilement que $l(\theta_n|\theta_n) = L(\theta_n)$ et on fait le raisonnement suivant : si on trouve un θ' tel que $l(\theta'|\theta_n) \geq l(\theta_n|\theta_n)$, alors il est une meilleure estimation du paramètre θ , puisque $L(\theta') \geq L(\theta_n)$. Pour trouver un tel θ' , on choisit de maximiser $l(\theta|\theta_n)$ par rapport à θ . On met donc à jour ainsi :

$$\theta_{n+1} = \arg \max_{\theta} l(\theta|\theta_n) \quad (3.3)$$

$$= \arg \max_{\theta} \left\{ \sum_z P(z|X, \theta_n) \ln \left(\frac{P(X|z, \theta)P(z|\theta)}{P(z|X, \theta_n)} \right) \right\} \quad (3.4)$$

$$= \arg \max_{\theta} \left\{ \sum_z P(z|X, \theta_n) \ln(P(X|z, \theta)P(z|\theta)) \right\} \quad (3.5)$$

$$= \arg \max_{\theta} \left\{ \sum_z P(z|X, \theta_n) \ln P(X, z|\theta) \right\} \quad (3.6)$$

$$= \arg \max_{\theta} \{ E_{z|X, \theta_n} \{ \ln P(X, z|\theta) \} \} \quad (3.7)$$

A partir de cette équation, l'algorithme EM consiste à effectuer des itérations sur :

- **Etape - E** : Déterminer l'estimation conditionnelle $E_{z|X, \theta_n} \{ \ln P(X, z|\theta) \}$.
- **Etape - M** : Maximiser l'expression en fonction de θ .

3.3.1.2 Définition de l'algorithme ε -EM

Nous avons étendu l'algorithme EM à un algorithme ε -EM. La procédure EM remplace la maximisation de $L(\theta)$ par celle de " $l(\theta)$ ", de manière itérative :

$$\theta_{n+1} = \arg \max_{\theta} l(\theta|\theta_n) \quad (3.8)$$

$$= \arg \max_{\theta} \left\{ \sum_z P(z|X, \theta_n) \ln P(X, z|\theta) \right\} \quad (3.9)$$

Or, certains trackings z sont hautement improbables, comme tous ceux qui attribuent à une cellule donnée i un successeur trop éloigné, par exemple à l'autre bout de l'embryon. Nos calculs ont ainsi fait apparaître des probabilités $P(z'|X, \theta_n) \approx 10^{-200}$, que nous avons retirés de la somme, une procédure que nous souhaitons légitimer ici. Soit un tel z' , alors :

$$l(\theta|\theta_n) = \sum_{z \neq z'} P(z|X, \theta_n) \ln P(X, z|\theta) + P(z'|X, \theta_n) \ln P(X, z'|\theta) \quad (3.10)$$

$$= \lambda(\theta|\theta_n) + \varepsilon(\theta|\theta_n) \quad (3.11)$$

Imaginons qu'on ait minoré pour tout θ , $\varepsilon(\theta|\theta_n)$ par un ε fixé, de telle sorte que :

$$-\varepsilon < \varepsilon(\theta|\theta_n) < 0$$

On utilise le raisonnement suivant. Soit $\hat{\theta}_{n+1} = \arg \max_{\theta} \lambda(\theta|\theta_n)$.

Par 3.11 on a $0 \leq \lambda(\theta|\theta_n) - l(\theta|\theta_n) \leq \varepsilon$, puis :

$$l(\hat{\theta}_{n+1}|\theta_n) \geq \lambda(\hat{\theta}_{n+1}|\theta_n) - \varepsilon \quad (3.12)$$

$$\lambda(\theta_n|\theta_n) \geq l(\theta_n|\theta_n) \quad (3.13)$$

Il suffit donc que

$$\lambda(\hat{\theta}_{n+1}|\theta_n) \geq \lambda(\theta_n|\theta_n) + \varepsilon \quad (3.14)$$

pour que $l(\hat{\theta}_{n+1}|\theta_n) \geq l(\theta_n|\theta_n)$ et donc $L(\hat{\theta}_{n+1}) \geq L(\theta_n)$, ce qui nous ramène au cadre de l'algorithme précédent. On pourra donc utiliser $\hat{\theta}_{n+1}$ à la place de θ_{n+1} à chaque étape. Inversement, 3.14 nous fournit une condition d'arrêt pour l'algorithme que nous appliquons : si l'incrément obtenu dans la maximisation de $\lambda(\theta_n|\theta)$ est inférieur à ε , nous nous arrêtons.

Nous avons réussi à minorer $\varepsilon(\theta|\theta_n)$ par un ε fixé dans un compact arbitrairement grand.

3.3.1.3 Définition de l'algorithme E'M'

L'algorithme précédent peut encore être simplifié en ne conservant de tous les termes de la somme que le terme "maximal", ce qui revient à éliminer progressivement tous les autres, considérés comme des " ε ".

L'algorithme *MorphoTrack* que nous avons implémenté est ainsi défini ci-dessus comme un algorithme E'M' :

- Etape E' : $z_n = \arg \max_z P(z|X, \theta_n)$
- Etape M' : $\theta_{n+1} = \arg \max_\theta \ln P(X, z_n|\theta)$

L'algorithme E'M' est presque un algorithme EM (cf. 3.3.1.1) au vu de ce qui précède. En effet,

- Etape E' : $|l(\theta|\theta_n) - P(z_n|X, \theta_n) \ln P(X, z_n|\theta)| \leq \varepsilon$ car z_n concentre l'essentiel des probabilités.
- Etape M' : $\hat{\theta}_{n+1} = \arg \max_\theta \ln P(X, z_n|\theta) = \arg \max_\theta P(z_n|X, \theta_n) \ln P(X, z_n|\theta)$ réalise donc le maximum de $l(\theta|\theta_n)$ à ε près.

3.3.2 Mécanique des fluides complexes

La base de la mécanique des milieux continus est l'étude des déformations et des phénomènes associés à une transformation d'un milieu. La notion de déformation sert à quantifier de quelle manière les longueurs ont été dilatées et les angles ont changé dans le milieu. Une branche de la mécanique des milieux continus est la mécanique des fluides. Elle consiste à modéliser la matière à l'aide de particules assez petites pour relever de l'analyse mathématique mais assez grandes par rapport aux molécules pour être décrites par des fonctions continues. Les fluides non newtoniens (comme le sang, les gels, les boues, les pâtes, les suspensions, les

émulsions,...) peuvent avoir des comportements très variés et généralement plus complexes. La mécanique des fluides complexes s'applique lorsque la déformation n'est pas directement proportionnelle à la force qu'on lui applique. Le meilleur exemple est celui du sable mouillé en bord de mer : quand on frappe le sable, il a la viscosité élevée d'un solide, alors que lorsqu'on appuie doucement dessus, il se comporte comme une pâte. L'étude de la mécanique des fluides non newtoniens est ainsi l'approche la plus appropriée à l'étude du milieu cellulaire au niveau de la cellule.

La déformation élastique est une déformation réversible : le milieu retourne à son état initial lorsque l'on supprime les sollicitations. Un choc élastique est un choc entre deux corps qui produit un rebond entièrement régi par l'élasticité des zones d'impact. Le système composé des corps qui se heurtent conserve :

- son énergie cinétique, du fait de l'absence de dissipation,
- sa quantité de mouvement, comme tout système isolé.

3.3.3 Mouvement Brownien et processus stochastique

Un grand nombre de phénomènes causaux à une échelle inférieure à la cellule, ne peuvent pas être mesurés sur nos volumes imagés. Ainsi, à l'échelle de l'observation cellulaire, les mouvements déterministes des phénomènes sous-jacents apparaissent comme du bruit.

A partir de l'étude des caractéristiques mesurées sur les données, nous faisons l'hypothèse que le mouvement cellulaire est Brownien.

Sur un vecteur d'état $X_i(t)$ de la cellule i à l'instant t et sa vitesse $V_i(t)$, l'équation du mouvement Brownien d'une cellule isolée est défini par :

$$dX_i(t) = V_i(t)dt + \sigma_i(t)dw_i(t) \quad (3.15)$$

où le vecteur dw_i représente le bruit Brownien et la matrice σ_i représente l'impact du bruit sur chacun des états X_i de la cellule i .

L'écart entre la prédiction déterministe et la position réelle est défini par : $\varepsilon_i(t) = dX_i(t) - V_i(t)dt = \sigma_i(t)dw_i(t)$

Afin de déterminer σ , la partie déterministe qui reflète la dépendance aux causalités sous-jacentes, nous calculons la matrice de variance-covariance sur i définie par :

$$\sum_i(t) = \langle \varepsilon_i(t)^t \varepsilon_i(t) \rangle = \sigma_i(t)^t \sigma_i(t) \quad (3.16)$$

Nous modélisons la trajectoire d'une cellule comme un processus stochastique avec des bruits indépendants. Les causes sont analogues mais tout se passe comme si elles se produisaient de façon indépendante. Ainsi nous dégagons trois hypothèses fondamentales :

- **Hypothèse 1** : Les bruits sont indépendants les uns des autres.
- **Hypothèse 2** : Les cellules ont des dynamiques stochastiques indépendantes.
- **Hypothèse 3** : Les déformations obéissent également à des dynamiques stochastiques indépendantes.

3.3.4 Ensemble statistiquement homogène

On définit un ensemble statistiquement homogène comme un sous-groupe de cellules similaires suivant certains critères statistiques. Les premiers ensembles statistiquement homogènes dépendent des vraisemblances biologiques, par exemple nous séparons les cellules en division des cellules en interphase. Les deuxièmes sont liés à des critères de respect de ces contraintes, ainsi les cellules qui meurent ou naissent subitement dans l'image ont des correspondance en terme de probabilité d'existence spécifique. De façon plus générale, nous définissons un ensemble statistiquement homogène comme un sous-groupe de cellules ayant une ressemblance *significantive* sur une liste des variables associées aux cellules.

La mesure de la vraisemblance dépend du type de la variable. Par exemple, un critère de respect temporel de l'information est un aspect fondamental et sa vraisemblance se mesure par une distance temporelle dans l'espace-temps à une dimension. Un autre exemple est la position spatiale des cellules dans l'embryon où, par une mesure de distance dans un espace 3D, on peut créer des zones statistiquement homogènes en respectant un critère de distance maximale.

Ainsi, selon la définition de la variable, on associe une mesure vraisemblance sous forme de groupement de l'information (*cluster*).

Les variables associées aux cellules sont dans notre étude :

- les variables d'états (position, temps, features, etc...)
- les variables associées au déplacement (vitesse, accélération)
- les variables de changement de trajectoire (courbure, chocs, etc...)
- les variables de déformation.

Ces ensembles statistiquement homogènes sont définis en 3.4.5, après que les notations auront été mises en place.

3.4 M- Maximisation

Nous allons définir une série de mesures permettant de maximiser la vraisemblance des informations. Elles seront utilisées dans la partie Maximisation de l'algorithme.

3.4.1 Définition des cellules et de leur *features*

On définit $c_i(t)$, la cellule i , qui a pour coordonnées spatiales le vecteur (x, y, z) , à un temps t par :

- Soit $X_i(t) = \{x_i^1, x_i^2, \dots, x_i^n\}$ le vecteur des n features associés de la cellule i à l'instant t
- Soit $\vec{V}_i(t) = \frac{(X_i(t) - X_i(t-1))}{\Delta t}$ la vitesse de variation des features de la cellule i à l'instant t

Remarque Cela n'est valable que sur l'ensemble des cellules $i(t)$ qui ont un prédécesseur à l'instant $t - 1$.

On peut en voir la représentation sur la figure 3.1.

Bien que chaque calcul se fasse à un instant précis dans le temps, par commodité de lecture pour la suite du document nous rendrons (t) invisible. Nous n'ajouterons la notion temporelle que lorsqu'il y aura une ambiguïté possible.

3.4.2 Notion de voisinage cellulaire

Il nous faut introduire pour la suite du document des notions de voisinage cellulaire. L'ensemble des voisins peut être calculé selon plusieurs procédures. Nous pouvons simplement prendre toutes les cellules j dans une boîte carrée de dimension d_{box} centrée aux coordonnées de la cellule i , ou de même dans une sphère de rayon r_{sphere} centré en i . Une définition plus élaborée serait de prendre les cellules j qui auraient un contact membranaires direct avec la cellule i . Ce contact peut être calculé par segmentation des membranes des cellules ou par approximation en utilisant un diagramme de Voronoï (Voronoi (1904)).

Nous allons introduire la notion de niveau d de voisinage cellulaire (figure 3.2). Il correspond aux cellules voisines du niveau $d - 1$ n'appartenant pas à l'ensemble des cellules des niveaux inférieurs. Par récursivité on pose $d = 1$ comme définissant le voisinage immédiat. Ainsi, dans le cas simple, cela représente un facteur multiplicatif de la dimension de la boîte ($d \times r_{box}$) ou du rayon de la sphère ($d \times r_{sphere}$). Dans le cas plus élaboré, cela correspond aux cellules qui auraient un contact membranaire avec les cellules du niveau inférieur.

Dans la suite de ce chapitre, le calcul de l'ensemble des voisins est un paramètre important à étudier mais n'entre pas en compte dans les définitions générales.

Nous introduisons donc les définitions de voisinage suivantes (à un instant t précis dans le temps) :

- Soit ϑ_i l'ensemble des n cellules constituant le voisinage de la cellule i , i compris.
- Soit $\vartheta_j \setminus i$ l'ensemble des cellules constituant le voisinage immédiat de la cellule j , i étant exclue.
- Et donc $\vartheta_i \setminus i = \vartheta_i^-$ l'ensemble des cellules constituant le voisinage immédiat de la cellule i , i exclue.
- Soit ϑ_i^d l'ensemble des cellules constituant le voisinage au niveau d de la cellule i , i incluse, avec $\vartheta_i^1 = \vartheta_i$
- De même ϑ_i^{d-} l'ensemble des cellules constituant le voisinage au niveau d de la cellule i , i exclue, avec $\vartheta_i^- = \vartheta_i^{1-}$

Nous ajoutons une hypothèse : si une cellule i est voisine de la cellule j , alors la cellule j est voisine de i . Soit une définition d'un voisinage réciproque tel que $i \in \vartheta_j \iff j \in \vartheta_i$.

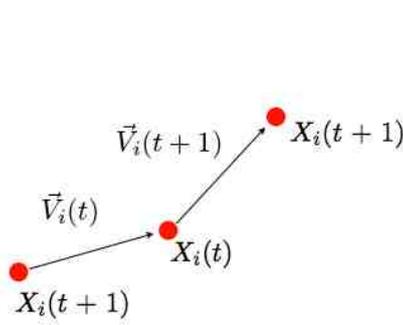


FIG. 3.1 – Pour 2 features, représentation du vecteur des features $X_i(t)$ et de la vitesse de variation des features $\vec{V}_i(t)$

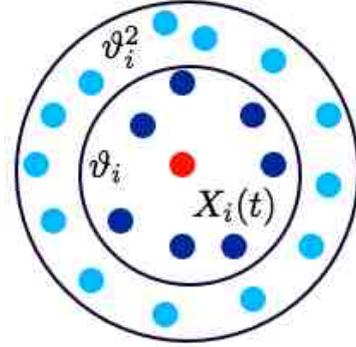


FIG. 3.2 – Représentation du voisinage d'une cellule au degré 1 (ϑ_i) et 2 (ϑ_i^2)

3.4.3 Définition de la matrice de variance - covariance du mouvement cellulaire

Soit $\bar{\vec{V}}_i = \frac{\sum_{j \in \vartheta_i} \vec{V}_j}{|\vartheta_i|}$ la vitesse de variation des features, moyenne calculée sur le voisinage ϑ_i .

La matrice de variance covariance sur l'ensemble des variations des features de la cellule i est définie par :

$$\sum_{\vec{V}_i} = \frac{\sum_{j \in \vartheta_i} (\vec{V}_j - \bar{\vec{V}}_i) \otimes (\vec{V}_j - \bar{\vec{V}}_i)}{|\vartheta_i|} \quad (3.17)$$

Les support de variation des features de la cellule i sont définis par :

$$S_{\vec{V}_i} = (\sup_{j \in \vartheta_i} \vec{V}_j, \inf_{j \in \vartheta_i} \vec{V}_j) \quad (3.18)$$

3.4.4 Définition de la matrice de déformation du voisinage associée à chaque cellule

Un point essentiel dans le *tracking* des cellules est de conserver au maximum le maillage cellulaire global, et pour cela nous allons introduire la notion de déformation du maillage.

La notion de déformation du maillage cellulaire appliquée à une cellule doit être étudiée dans le cas général sur l'ensemble des voisins constituant la cellule i mais i étant exclue. Cela afin que, quelle que soit la mesure que nous devons faire sur le contexte de la cellule i , les interférences (comme une mauvaise segmentation ou

un mauvais tracking) de cette cellule n'aient pas d'impact sur le calcul de la déformation du voisinage.

On associe à chaque cellule i une matrice de déformation M_i sur son voisinage ϑ_i^- respectant la minimisation :

$$M_i = \arg \min_M \sum_{j \in \vartheta_i^-} \|(X_j(t+1) - X_i(t+1)) - M(X_j(t) - X_i(t))\|^2 \quad (3.19)$$

Pour résoudre cette équation, il nous faut définir quelques notions supplémentaires :

- Soit $\delta_{X_{ij}} = X_j - X_i$ l'écartement des features entre deux cellules i et j
- Soit $\delta_{\vec{V}_{ij}} = \vec{V}_j - \vec{V}_i$ l'écartement des vitesses de variation des features entre deux cellules i et j

On définit A_i la matrice telle que (démonstration en Annexe)

$$A_i = \frac{\sum_{j \in \vartheta_i^-} \delta_{X_{ij}} \otimes \delta_{X_{ij}}}{|\vartheta_i^-|} = \sum_{M_i} + \overline{\delta_{X_{ij}}} \otimes \overline{\delta_{X_{ij}}} \quad (3.20)$$

où $\sum_{M_i} = (\delta_{X_{ij}} - \overline{\delta_{X_{ij}}}) \otimes (\delta_{X_{ij}} - \overline{\delta_{X_{ij}}})$ représente la matrice de variance-covariance des déformations de la cellule, et $\overline{\delta_{X_{ij}}}$ représente le biais tel que le centre de la cellule i soit considéré comme le barycentre de son voisinage ϑ_i^- .

Ainsi que B_i la matrice telle que

$$B_i = \frac{\sum_{j \in \vartheta_i^-} \delta_{\vec{V}_{ij}} \otimes \delta_{X_{ij}}}{|\vartheta_i^-|} = \overline{B_i} + \overline{\delta_{\vec{V}_{ij}}} \otimes \overline{\delta_{X_{ij}}} \quad (3.21)$$

où $\overline{B_i} = (\delta_{\vec{V}_{ij}} - \overline{\delta_{\vec{V}_{ij}}}) \otimes (\delta_{X_{ij}} - \overline{\delta_{X_{ij}}})$ représente la matrice de corrélation entre \vec{V}_{ij} et X_{ij} sur son voisinage ϑ_i^- .

La matrice de déformation du voisinage de la cellule i est définie par

$$\hat{M}_i = B_i A_i^{-1} \quad (3.22)$$

avec

$$\hat{M}_i = \arg \min_M \sum_{j \in \vartheta_i^-} \|(\vec{V}_i - \vec{V}_j) \Delta t - M(X_j - X_i)\|^2$$

Ainsi, la matrice de déformation moyenne sur l'ensemble du voisinage de la cellule i par :

$$\overline{\hat{M}_i} = \frac{\sum_{j \in \vartheta_i^-} \hat{M}_j}{|\vartheta_i^-|} \quad (3.23)$$

Le support de distribution de la matrice de déformation est défini par :

$$S_{M_i} = \left(\inf_{j \in \vartheta_i^-} \hat{M}_j, \sup_{j \in \vartheta_i^-} \hat{M}_j \right) \quad (3.24)$$

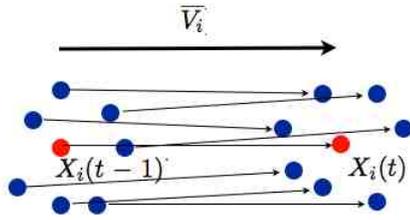


FIG. 3.3 – Représentation de la vitesse moyenne de variation entre deux pas de temps

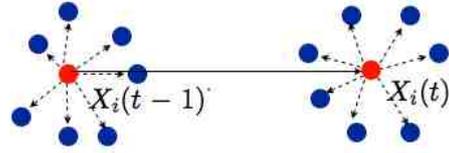


FIG. 3.4 – Déformation du maillage du voisinage entre deux pas de temps

3.4.5 Détection des frontières

Ce paragraphe explicite la méthode utilisée pour construire des zones statistiquement homogènes. Elles sont essentielles pour l'estimation des lois de probabilités. Les voisinages des cellules seront toujours construits au sein des champs statistiquement homogènes. On va chercher par la suite à trouver des discontinuités de mouvement entre champs morphogénétiques. Il nous faut pour cela introduire des notions de frontières entre ensembles de cellules. Ainsi, une frontière ne sera pas nécessairement simplement spatiale mais pourra aussi s'appliquer à chaque feature, s'il permet de distinguer des groupes d'éléments (par exemple les cellules épithéliales qui sont bien plus grosses que les autres cellules des couches internes).

On définit deux sous-groupes ϑ_{i_1} et ϑ_{i_2} dans l'ensemble du voisinage de la cellule i , tel que $\vartheta_{i_1} \cup \vartheta_{i_2} = \vartheta_i$ et $\vartheta_{i_1} \cap \vartheta_{i_2} = \emptyset$ tel que $i \notin \vartheta_{i_1}$ et $i \notin \vartheta_{i_2}$ (figure 3.5).

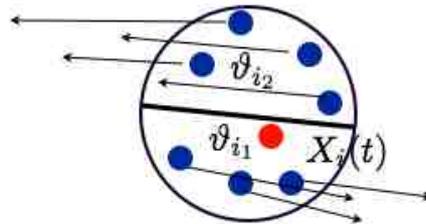


FIG. 3.5 – Représentation de deux sous groupes de voisins lors d'une discontinuité de la vitesse de variation des features

On a pour chacun des sous-ensembles :

- Leur moyenne des vitesses de variation \bar{V}_{i_1} et \bar{V}_{i_2}
- La variance définie telle que $\sigma_{i_1}^2 = \frac{\sum_{j \in \vartheta_{i_1}} \|\vec{V}_j - \bar{V}_{i_1}\|^2}{|\vartheta_{i_1}|}$ et $\sigma_{i_2}^2 = \frac{\sum_{j \in \vartheta_{i_2}} \|\vec{V}_j - \bar{V}_{i_2}\|^2}{|\vartheta_{i_2}|}$ (où σ_{i_1} et σ_{i_2} sont les écarts-types respectifs)

D'où, sur le voisinage ϑ_i de la cellule i , on a

$$\begin{aligned} |\vartheta_i| \sigma_i^2 &= \sum_{j \in \vartheta_i} \|\vec{V}_j - \vec{V}_i\|^2 \\ &= \sum_{j \in \vartheta_{i_1}} (\|\vec{V}_j - \vec{V}_{i_1}\|^2 + 2(\vec{V}_j - \vec{V}_{i_1}) \cdot (\vec{V}_{i_1} - \vec{V}_i)) + |\vartheta_{i_1}| \|\vec{V}_{i_1} - \vec{V}_i\|^2 \\ &+ \sum_{j \in \vartheta_{i_2}} (\|\vec{V}_j - \vec{V}_{i_2}\|^2 + 2(\vec{V}_j - \vec{V}_{i_2}) \cdot (\vec{V}_{i_2} - \vec{V}_i)) + |\vartheta_{i_2}| \|\vec{V}_{i_2} - \vec{V}_i\|^2 \end{aligned}$$

Soit

$$(|\vartheta_{i_1}| + |\vartheta_{i_2}|) \sigma_i^2 = |\vartheta_{i_1}| (\sigma_{i_1}^2 + \|\vec{V}_{i_1} - \vec{V}_i\|^2) + |\vartheta_{i_2}| (\sigma_{i_2}^2 + \|\vec{V}_{i_2} - \vec{V}_i\|^2)$$

Il faut donc trouver les ϑ_{i_1} et ϑ_{i_2} qui minimisent la variance intra-classe :

$$|\vartheta_{i_1}| \sigma_{i_1}^2 + |\vartheta_{i_2}| \sigma_{i_2}^2$$

ou qui maximisent la variance inter-classe :

$$\begin{aligned} (|\vartheta_{i_1}| + |\vartheta_{i_2}|) \sigma_i^2 - |\vartheta_{i_1}| \sigma_{i_1}^2 - |\vartheta_{i_2}| \sigma_{i_2}^2 &= |\vartheta_{i_1}| \|\vec{V}_{i_1} - \vec{V}_i\|^2 + |\vartheta_{i_2}| \|\vec{V}_{i_2} - \vec{V}_i\|^2 \\ \text{après simplification} &= |\vartheta_{i_1}| |\vartheta_{i_2}| \|\vec{V}_{i_1} - \vec{V}_{i_2}\|^2 \end{aligned}$$

qui consiste à bien séparer deux groupes.

Si on détecte des discontinuités dans le champ de vitesse, cela permet aussi de réduire les supports et la matrice de variance-covariance.

Il est aussi intéressant de faire le même calcul sur la matrice de déformation. Elle n'est autre que $\frac{\partial X}{\partial x}$. Une estimation de sa valeur est définie au 3.4.4. La norme de cette matrice fournit une autre manière d'estimer les "discontinuités" dans le champ des vitesses. La matrice permet même de voir quelles sont les composantes de la vitesse qui participent davantage séparément ou ensemble à la discontinuité.

Ces différentes mesures permettent notamment d'obtenir des patterns spatio-temporels homogènes statistiquement.

3.4.6 Cellules en division

Soit une cellule mère i au temps t et ses deux filles i_1^+ et i_2^+ au temps $t + 1$. La vitesse de variation de features dans le cas d'une division cellulaire revient simplement à calculer $\vec{V}_i^d = \frac{\vec{V}_{i_1^+} + \vec{V}_{i_2^+}}{2}$. De cette formule et de l'équation vue en 3.17 nous pouvons déduire la matrice de variance-covariance associée à ce *branching process*. Et de la même façon nous pouvons déterminer la matrice de déformation du voisinage autour de cette cellule en division.

Soit \vec{V}_i^- le mouvement moyen de la cellule i au voisinage immédiat, i exclue.

La loi de Stokes permet de mesurer les forces de frottements et les viscosités des liquides. Elle nous permet de définir la force de résistance du milieu cellulaire

\vec{F} où η est le coefficient de viscosité du fluide et k le coefficient caractéristique de la géométrie du solide, par l'hypothèse suivante :

$$\begin{aligned}\vec{F} &= -k\eta\vec{V} \\ &= -\alpha\vec{V} \text{ avec } \alpha = k\eta\end{aligned}$$

où $\vec{V} = \vec{V}_i - \overline{\vec{V}_i^-}$ représente la vitesse relative de la cellule dans son milieu. On observe ainsi le mouvement dans le champ morphogénétique (ou le voisinage cellulaire) en annulant la vitesse moyenne.

L'accélération étant colinéaire avec la vitesse et la force de résistance colinéaire à la trajectoire, d'après la relation fondamentale de la dynamique on a : $\vec{F} = \frac{m d\vec{V}}{dt}$. On obtient donc :

$$\begin{aligned}m_i \frac{d\vec{V}}{dt} &= -\alpha\vec{V} \\ m_i \frac{d(\vec{V}_i - \overline{\vec{V}_i^-})}{dt} &= -\alpha(\vec{V}_i - \overline{\vec{V}_i^-})\end{aligned}$$

Donc, en passant au module, nous obtenons en coordonnées curvilignes (système non fixe mais dépendant de la position du point dans l'espace) :

$$m_i \frac{d\|\vec{V}_i - \overline{\vec{V}_i^-}\|}{dt} = -\alpha\|\vec{V}_i - \overline{\vec{V}_i^-}\|$$

$\vec{V}_i(0)$ représente la vitesse initiale de la cellule à l'instant de la division et $\overline{\vec{V}_i^-}(0)$ sa vitesse moyenne (i exclue), on obtient :

$$\|\vec{V}_i(t) - \overline{\vec{V}_i^-}(t)\| = \|\vec{V}_i(0) - \overline{\vec{V}_i^-}(0)\| e^{-\frac{\alpha}{m_i}t} \quad (3.25)$$

D'après le théorème de l'énergie cinétique, $E = \frac{1}{2}\vec{V}^2$, on en déduit l'énergie associée à la cellule au moment de la division :

$$\begin{aligned}E_i(t) - E_i(0) &= \int F \cdot \vec{V} dt \\ &= \frac{1}{2}m_i\|\vec{V}_i(t) - \overline{\vec{V}_i^-}(t)\|^2 - \frac{1}{2}m_i\|\vec{V}_i(0) - \overline{\vec{V}_i^-}(0)\|^2\end{aligned}$$

Nous nous le verrons le chapitre 4 qu'après division les cellules filles se font rapatrier dans le mouvement moyen cellulaire. Le temps $\tau_i^{\vec{\sigma}}$ à partir duquel la trajectoire de la cellule i reprend la vitesse du champ moyen cellulaire est défini par :

$$\tau_i^{\vec{\sigma}} = \min\{t : \|\vec{V}_i(t) - \overline{\vec{V}_i^-}(t)\| \leq 2\sigma_i(t)\}$$

où $\sigma_i(t)$ symbolise l'impact du mouvement Brownien défini en 3.15.

La longueur de la trajectoire de la cellule fille jusqu'au temps $\tau_i^{\vec{\alpha}}$ est définie par :

$$\begin{aligned} L_i(t) &= \int_0^{\tau_i^{\vec{\alpha}}} \|\vec{V}_i(t) - \overline{\vec{V}_i}(t)\| dt \\ &= \frac{m_i}{\alpha} (1 - e^{-\frac{\alpha}{m_i} \tau_i^{\vec{\alpha}}}) \|\vec{V}_i(0) - \overline{\vec{V}_i}(0)\| \text{ obtenu par 3.25} \end{aligned}$$

Soit $\vec{V}_{i_1^+}$ la vitesse d'éjection de la fille i_1 au pas de temps $t + 1$, et $\vec{V}_{i_1^n}$ au pas de temps $t + n$ (idem pour i_2).

Dans le milieu local de la cellule, nous observons une vitesse moyenne $\overline{\vec{V}_{i_1^n}}$. La vitesse réelle d'éjection des filles doit être calculée en soustrayant cette vitesse moyenne du milieu cellulaire pour permettre d'en établir sa valeur réelle.

Soit $v_{i_1^n} = \vec{V}_{i_1^n} - \overline{\vec{V}_{i_1^n}}$ la vitesse réelle d'éjection de la fille i_1^n au pas de temps $t + n$.

En se situant dans une même zone statistiquement homogène, nous pouvons déterminer les valeurs moyennes et leur écart type sur chacune des mesures faites sur l'ensemble des mitoses. Ces valeurs moyennes seront utilisées dans la partie Maximisation du *tracking* pour calculer les probabilités associées aux événements.

Nous allons aussi effectuer cet ensemble de mesures sur le groupe des *voisines des cellules en division*.

3.4.7 Statistiques sur les mesures

De l'ensemble de ces mesures nous pouvons sortir un grand nombre des statistiques à plusieurs niveaux :

- Sur une cellule en division, nous pouvons donc extraire son énergie E_i , sa masse m_i , son temps de vie $\tau_i^{\vec{\alpha}}$ durant les différentes phases de la mitose, ainsi que sur sa longueur de la trajectoire, L_i , son nombre de chocs k_i avec son voisinage associés durant la division, avec ses angles corrélés (θ_k, ϕ_k) , et bien sûr le temps de vie global de la cellule T_i .
- Sur les paires de cellules i_1 et i_2 issues d'une division, nous pouvons mesurer un ensemble de corrélations mutuelles, leur masse (m_{i_1}, m_{i_2}) , comparer leurs énergies (E_{i_1}, E_{i_2}) , et corrélérer leur temps de vie (T_{i_1}, T_{i_2})
- Sur l'histoire de la cellule, nous pouvons interpréter le temps de vie de la cellule T_{i_1} en fonction du temps de vie de sa cellule mère T_{i_0} , ainsi que les corrélations avec la cellule s $\frac{1}{2}$ ur de façon probabiliste, $P((T_{i_1}, T_{i_2})/T_{i_0})$

Enfin, sur cet ensemble de statistique, il est possible d'en extraire les matrices de variance-covariance, de corrélation ainsi que leur supports.

Nous pouvons faire ces mesures sur un voisinage immédiat ($d = 1$) ou sur un voisinage plus large ($d = 2, 3..$).

3.5 E - Estimation

A partir des différentes mesures effectuées dans la partie Maximisation de l'algorithme, nous pouvons maintenant estimer la probabilité de la fonction de *tracking*. L'équation du maximum de vraisemblance définie en 3.1 nécessite des mesures de probabilités sur chaque élément en fonction du *tracking* courant. Nous allons donc dans la partie Estimation de l'algorithme, définir puis calculer un ensemble de *probabilités associées*. Nous parlerons de la *probabilité* d'un élément, *associée* à une notion, comme la mesure probabiliste de cet élément dans le contexte de cette notion. Ainsi nous définissons une *probabilité associée* au mouvement de la cellule, à sa déformation, à son existence dans temps, de se diviser, etc.

3.5.1 f , fonction de *tracking*

Soit f la fonction de *tracking forward* (vers l'avant) qui pour la cellule i au temps t associé au temps $t + 1$

$$i \xrightarrow{f} \begin{cases} i^+ & \text{si la cellule a un et un seul successeur} \\ (i_1^+, i_2^+) & \text{si la cellule a deux successeurs} \\ \emptyset & \text{si la cellule n'a pas de successeur} \end{cases} \quad (3.26)$$

Soit f^{-1} la fonction de *tracking backward* (vers l'arrière) telle que

$$i \xrightarrow{f^{-1}} \begin{cases} i^- & \text{si la cellule a un et un seul prédécesseur} \\ \emptyset & \text{si la cellule n'a pas de prédécesseur} \end{cases} \quad (3.27)$$

Par cette définition, nous respectons l'ensemble des contraintes biologiques, les plus fondamentales étant qu'une cellule ne peut avoir plus de 2 filles, et qu'une cellule a au maximum une et une seule mère.

De (3.26) et (3.27) nous définissons une fonction de *tracking global* définie par

$$\left. \begin{matrix} i^- \\ \emptyset \end{matrix} \right\} \xleftarrow{f^{-1}} i \xrightarrow{f} \begin{cases} i^+ \\ (i_1^+, i_2^+) \\ \emptyset \end{cases} \quad (3.28)$$

Par extrapolation on a la fonction f^τ qui associe le successeur de la cellule i à un temps τ , soit $i \xrightarrow{f^\tau} i^\tau$.

3.5.2 Fenêtre spatio-temporelle autour d'une cellule

Le départ de l'algorithme MorphoTrack se fait à partir d'un *tracking* déjà effectué, qui permet d'estimer les paramètres θ des lois de probabilités définies dans la partie Estimation ci-dessus. Connaissant ces lois de probabilités, nous cherchons maintenant à maximiser la probabilité du lignage cellulaire, ce qui suppose d'abord de la calculer. Son calcul fait intervenir une probabilité forward (à partir des états passés, 3.5.6) et une probabilité backward (à partir des états futurs, 3.5.7).

Après chaque itération nous obtenons en sortie de l'algorithme un *tracking* un peu meilleur.

Soit $i \xrightarrow{\tau} t$ la fonction temporelle τ qui à chaque cellule i associe son temps t .

On définit $\mathbb{F}_i = \{f^\tau(i) : \tau \in [-\eta, \eta]\}$ comme l'ensemble représentant les différents successeurs et prédécesseurs de la cellule i entre le temps $-\eta$ et η .

On peut donc représenter les fonctions f et τ par le schéma suivant :

$$\begin{array}{cccccccccccccccc}
 i^{-\eta} & \xleftarrow{f^{-\eta}} & \dots & i^{-2} & \xleftarrow{f^{-2}} & i^{-1} & \xleftarrow{f^{-1}} & i & \xrightarrow{f} & i^+ & \dots & \xrightarrow{f^\eta} & i^\eta \\
 \tau \downarrow & & & \tau \downarrow & & \tau \downarrow & & \tau \downarrow & & \downarrow \tau & & & \downarrow \tau \\
 t - \eta & \xleftarrow{pred} & \dots & t - 2 & \xleftarrow{pred} & t - 1 & \xleftarrow{pred} & t & \xrightarrow{succ} & t + 1 & \dots & \xrightarrow{succ} & t + \eta
 \end{array}$$

3.5.3 Mesure de probabilités associées à f

$$\theta_{n+1}^i = \sum p(z_i | \theta_n) (X_i(t + \eta) - X_i(t - \eta))$$

L'estimation des mesures nous permet de construire un modèle stochastique probabiliste. Nous allons donc définir un ensemble de probabilités permettant de mesurer leur distribution. Nous allons donner une définition d'un *tracking* probabiliste, mais auparavant, il nous faut introduire quelques notions :

- \mathbb{T} représente l'espace temporel des cellules à tracker.
- \mathbb{I}_t représente l'ensemble des cellules à tracker au temps t .

Soit $p(f)$ la probabilité p associée au *tracking* f . Afin d'obtenir le meilleur *tracking* possible à chaque itération, nous allons maximiser sa probabilité.

Sous l'hypothèse que les trajectoires des cellules peuvent être modélisées comme un processus stochastique (cf 3.3.3), nous obtenons une indépendance des probabilités qui nous permet de les multiplier entre elles.

Soit

$$\hat{f} = \arg \max_f p(f) \quad (3.29)$$

avec

$$\begin{aligned}
 p(f) &= \prod_{t \in \mathbb{T}} p_t(f) \text{ (sur l'ensemble de pas de temps)} \\
 &= \prod_{t \in \mathbb{T}} \prod_{i \in \mathbb{I}_t} p_i(f) \text{ (sur chacune des cellules)}
 \end{aligned}$$

La probabilité associée à chaque cellule i au temps t est définie par :

$$p_i(f) = p_i^{\vec{V}}(f) \times \sqrt{\prod_{j \in \vartheta_i^-} p_{ij}^{\vec{M}}(f)} \quad (3.30)$$

où

- ϑ_i^- représente le voisinage spatial de la cellule i à l'instant t sans i (voir paragraphe 3.4.2),
- $p_i^{\vec{V}}(f)$ représente la probabilité de i associée au mouvement,
- $p_{ij}^{\vec{M}}(f)$ représente la probabilité de i associée à la déformation de son voisinage.

La multiplication des probabilités associées à la déformation du voisinage est sous la racine pour ne pas multiplier deux fois la même valeur, car on a pour tout i et j dans l'ensemble des cellules $i \in \vartheta_j \iff j \in \vartheta_i$.

On définit chacune des probabilités comme la multiplication des probabilités passée et future, soit :

pour la probabilité de i associé au mouvement :

$$p_i^{\vec{V}}(f) = \hat{p}_i^{\vec{V}}(f) \times \hat{p}_i^{\vec{V}}(f) \quad (3.31)$$

et pour la probabilité de i associé au déformation de son voisinage :

$$p_{ij}^{\vec{M}}(f) = \hat{p}_{ij}^{\vec{M}}(f) \times \hat{p}_{ij}^{\vec{M}}(f) \quad (3.32)$$

Les différents $\hat{p}(f)$ représentent les meilleurs probabilités associées au *tracking* f . Dans le cas du *tracking forward* on a

$$\hat{p}_i^{\vec{V}}(f) = \max\{p_i^{\vec{V}}(f), p_i^{\vec{V}}(f), p_i^{\vec{V}}(f)\} \quad (3.33)$$

(où $p^{\vec{V}}$ représente la probabilité d'avoir un successeur, $p^{\vec{V}}$ celle de ne pas en avoir et $p^{\vec{V}}$ celle de se diviser (cf 3.5.6))

Et dans le cas du *tracking backward*, on a

$$\hat{p}_i^{\leftarrow}(f) = \max\{p_i^{\leftarrow}(f), p_i^{\leftarrow}(f)\} \quad (3.34)$$

(où p^{\leftarrow} représente la probabilité d'avoir un prédécesseur, p^{\leftarrow} celle de ne pas en avoir (cf 3.5.7))

3.5.4 Probabilité associée à une branche de l'arbre

Il est parfois utile de calculer localement la probabilité associée à une cellule sur plusieurs pas de temps, ce qui nous donne une probabilité associée à la branche d'un tracking.

Nous notons ϑ_i^η le voisinage spatio-temporel de la cellule i pour les temps $t \in [-\eta, \eta]$.

On définit la notion de probabilité de la cellule i sur son voisinage spatio-temporel ϑ_i^η par :

$$p_{i\eta}(f, \vartheta_i^\eta) = \prod_{t \in [-\eta, +\eta]} p_{f^t(i)}(f, \vartheta_{f^t(i)}(t)) \quad (3.35)$$

$$= \prod_{t \in [-\eta, +\eta]} \prod_{j \in \vartheta_i(t)} p_{f^t(j)}(f) \quad (3.36)$$

Remarque : L'ensemble des notations dans les paragraphes qui vont suivre sont représentées sur la figure 3.6. Dans cet exemple, on calcule la probabilité d'une cellule i à l'instant t sur son voisinage $j \in \{j_1, j_2, j_3\}$ avec $\eta = 3$. La probabilité associée à cette cellule est donc la multiplication de l'ensemble de ces probabilités.

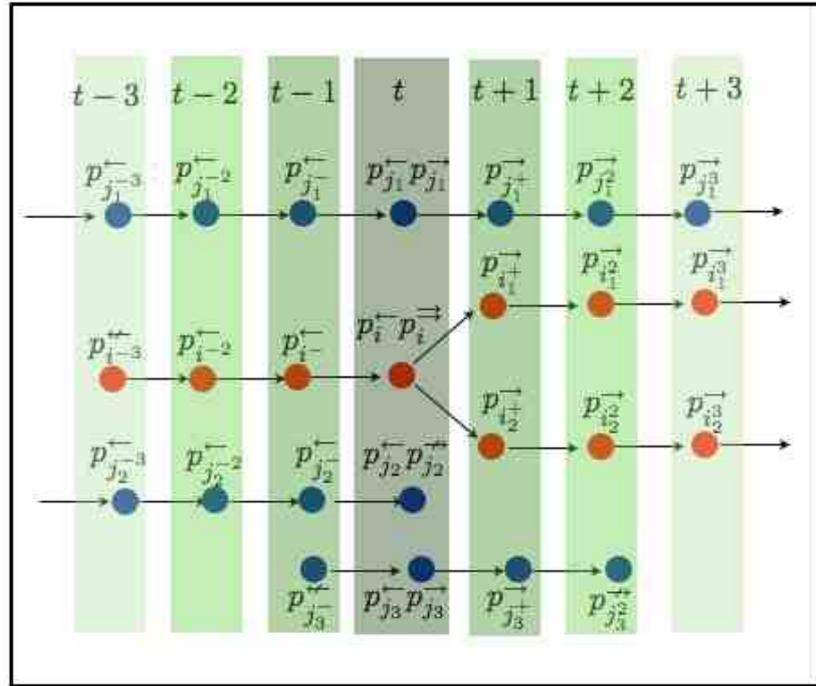


FIG. 3.6 – Représentation schématique des différents probabilités associées à la cellule i et son voisinage pour $t \in [-3, 3]$

3.5.5 Définition de la loi de probabilité associée

Nous sommes ici dans un cas à plusieurs dimensions (cf le nombre de *features*). Nous utilisons une loi normale multidimensionnelle représentée par une densité Gaussienne multivariée.

Cette loi admet une densité de probabilité qui, par le théorème de la limite centrale, est définie par :

$$p(x) = \frac{e^{-\frac{1}{2}(x-\bar{x})\Sigma^{-1}(x-\bar{x})}}{(2\pi)^{\frac{n}{2}}\sqrt{\det(\Sigma)}} \quad (3.37)$$

Remarque : le dénominateur est une simple normalisation qui assure que la densité intègre à 1, mais ne change rien à sa forme.

Dans certains calculs de probabilités dans les paragraphes suivants, nous devrons aussi calculer la loi normale pour la dimension 1. Elle est dans ce cas définie

par :

$$p(x) = \frac{e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}} \quad (3.38)$$

3.5.6 Probabilités associées au *tracking forward*

Dans le cas de l'étude du *tracking forward* f^+ , nous distinguerons trois cas :

- p^{\rightarrow} : la probabilité pour qu'une cellule ait un et un seul successeur.
- $p^{\vec{\rightarrow}}$: la probabilité pour qu'une cellule se divise.
- p^{\nrightarrow} : la probabilité pour qu'une cellule n'ait pas de successeur (apoptose ou disparition).

Soit

$$\left\{ \begin{array}{l} \text{à } i \xrightarrow{f} i^+ \text{ la probabilité associée } p_i^{\rightarrow}(f) \\ \text{à } i \xrightarrow{f} (i_1^+, i_2^+) \text{ la probabilité associée } p_i^{\vec{\rightarrow}}(f) \\ \text{à } i \xrightarrow{f} \emptyset \text{ la probabilité associée } p_i^{\nrightarrow}(f) \end{array} \right. \quad (3.39)$$

3.5.6.1 Probabilité d'avoir un successeur p_i^{\rightarrow}

Sous l'hypothèse que les probabilités soient indépendantes les unes des autres, nous définissons la probabilité qu'une cellule ait un et un seul successeur par la multiplication des probabilités du mouvement de cette cellule et de la déformation de son voisinage.

Soit

$$p_i^{\rightarrow}(f, \vartheta_i) = p_i^{\vec{\rightarrow}}(f) \times \prod_{j \in \vartheta_i} p_{ij}^{\frac{M_i}{\rightarrow}}(f) \quad (3.40)$$

Remarque : dans le cas où une cellule j est en mitose ou n'a pas de successeur, elle est ôtée de ce calcul.

Nous définissons le mouvement Brownien discrétisé par :

$$\Delta_i(t) = X_{i+}(t) - \bar{X}_i(t) - \overline{\vec{V}_i}(t)\Delta t \quad (3.41)$$

Nous définissons la probabilité associée au mouvement d'une cellule par :

$$p_i^{\vec{\rightarrow}}(f) = \frac{e^{-\frac{1}{2}\Delta_i \sum_{\vec{V}_i}^{-1} \Delta_i}}{(2\pi)^{\frac{|X_i|}{2}} \times \sqrt{\det(\sum_{\vec{V}_i})}} \quad (3.42)$$

où $\sum_{\vec{V}_i}$ est la matrice de variance-covariance des *features* calculée dans la partie Estimation de l'algorithme à l'équation 3.17.

De la même façon, nous définissons la *déformation Brownienne* par : $\Delta_{ij} = (X_{j+} - X_{i+}) - (X_j - X_i) - \hat{M}_i(X_j - X_i)$ (où \hat{M}_i représente la matrice de déformation calculée en (3.21)).

Et on définit la probabilité de déformation de la cellule j voisine de i par :

$$p_{ij}^{\frac{M_i}{\rightarrow}}(f) = \frac{e^{-\frac{1}{2}\Delta_{ij} \sum_{M_i}^{-1} \Delta_{ij}}}{(2\pi)^{\frac{|X_i|}{2}} \times \sqrt{\det(\sum_{M_i})}} \quad (3.43)$$

où \sum_{M_i} est la matrice de variance-covariance des déformations calculée en (3.20).

Cela sous contrainte que $j \in S_{M_i}$ (définition des supports en 3.24)

La probabilité de l'ensemble des successeurs possibles de la cellule i est représentée par .

$$p_i^{\rightarrow}(f, \vartheta_i) = \{f(i) | p_{f(i)} \geq \varepsilon\} = \{(i^+, p_i) | p_i \geq \varepsilon\}$$

Nous faisons une normalisation des probabilités telle que $\sum_{k \in [1, K]} \pi_k^{\rightarrow}(f, \vartheta_k) = 1$ avec

$$\pi_i^{\rightarrow}(f, \vartheta_i) = \frac{p_i^{\rightarrow}(f, \vartheta_i)}{\sum_{i \in \mathbb{I}^+} p_i^{\rightarrow}(f, \vartheta_i)} \quad (3.44)$$

Remarque : si la probabilité maximale est nulle ($\hat{p}_1 = 0$) alors nous sommes dans le cas que la cellule i est un faux positif (un centre ajouté par l'algorithme de détection de centres alors qu'il ne devrait pas exister) ou dans le cas que le successeur de i est un faux négatif (l'algorithme de détection de centres a oublié au pas de temps $t + 1$ un centre à cet endroit). Il nous est alors possible après plusieurs itérations de l'algorithme EM de *tracking* de le rajouter ou de l'enlever.

3.5.6.2 Probabilité de se diviser p_i^{\rightarrow}

La probabilité pour une cellule de se diviser est plus difficile à calculer car elle est un événement rare et qui a une très grande importance. Il faut prendre en compte l'ensemble des statistiques calculées dans la phase d'estimation de l'algorithme.

Soit i la cellule mère et j_1^+, j_2^+ ses deux filles au pas de temps suivant. Selon 3.28, on a $f(i) = (j_1^+, j_2^+)$, et (j_1^n, j_2^n) les filles au pas de temps $t + n$.

***i* - Statistiques sur la division cellulaire**

Nous pouvons sur l'ensemble des statistiques calculées en 3.4.7 définir pour chacune sa probabilité associée. A la fin de la partie Estimation nous calculons sur l'ensemble de mitoses, dans une zone statistiquement homogène (par exemple un espace temporel centré en t) la valeur moyenne de chacune des statistiques. Ainsi nous avons un ensemble \mathbb{C}_{j^n} de mesures associées à chacune des filles $j^n \in \{j_1^n, j_2^n\}$ tel que $\mathbb{C}_{j^n} = \{E_{j^n}, m_{j^n}, v_{j^n}, k_{j^n}, L_{j^n}, T_{j^n}, \tau_{j^n}^{\rightarrow}\}$.

On peut calculer sur cet ensemble à chaque pas de temps n après division la valeur moyenne de chacune des mesures, soit $\overline{\mathbb{C}_n} = \{\overline{E_{j^n}}, \overline{m_{j^n}}, \overline{v_{j^n}}, \overline{k_{j^n}}, \overline{L_{j^n}}, \overline{T_{j^n}}, \overline{\tau_{j^n}^{\rightarrow}}\}$ et leur l'écart-type associé $\sigma_{\mathbb{C}_n} = \{\sigma_{E_{j^n}}, \sigma_{m_{j^n}}, \sigma_{v_{j^n}}, \sigma_{k_{j^n}}, \sigma_{L_{j^n}}, \sigma_{T_{j^n}}, \sigma_{\tau_{j^n}^{\rightarrow}}\}$.

Chacune des probabilités suivantes sera bien sûr associée au *tracking* (f) mais nous ne l'ajoutons pas pour ne pas surcharger l'écriture.

***ii* - Probabilité associée à une statistique**

Pour chaque élément c_n de cet ensemble \mathbb{C}_n (avec \bar{c}_n le correspondant dans $\overline{\mathbb{C}_n}$, et σ_{c_n} dans $\sigma_{\mathbb{C}_n}$), il nous est maintenant possible de calculer la probabilité qui lui est associée. Nous sommes face à des valeurs à une dimension, donc d'après 3.38, la probabilité associée à une fille j^n au pas de temps $t + n$ sur la statistique c_n est définie par :

$$p_{c_j^n}^{\vec{\rightarrow}} = \frac{e^{-\frac{1}{2}\left(\frac{c_j^n - \bar{c}_n}{\sigma_{c_n}}\right)^2}}{\sigma_{c_n} \sqrt{2\pi}} \quad (3.45)$$

Comme la statistique se calcule à chaque pas de temps n après division, nous pouvons calculer la probabilité d'une *branche* fille j issue de la division par :

$$p_{c_j}^{\vec{\rightarrow}} = \prod_{n \in [t, t+\eta]} p_{c_j^n}^{\vec{\rightarrow}} \quad (3.46)$$

Et donc la probabilité associée à chacune des statistiques sur les deux filles $j \in \{i_1, i_2\}$ pour une cellule i est définie par :

$$p_{c_i}^{\vec{\rightarrow}} = p_{c_{i_1}}^{\vec{\rightarrow}} \times p_{c_{i_2}}^{\vec{\rightarrow}} \quad (3.47)$$

iii - Probabilité associée à un couple de statistiques

Nous avons aussi vu qu'il pouvait être intéressant de corrélérer les statistiques des deux filles pour arriver à une meilleure détection des cellules en division. On définit donc pour chaque élément $c_{j^n} \in \mathbb{C}_{j^n}$ un ensemble $\mathbb{P}_{c_{j^n}}$ de couple de statistique associé aux deux filles tel que $j \in \{j_1^n, j_2^n\}$ par $\mathbb{P}_{c_{j^n}} = \{(c_{j_1^n}, c_{j_2^n})\}$. On a donc $\mathbb{P}_{i^n} = \{(E_{j_1^n}, E_{j_2^n}), (m_{j_1^n}, m_{j_2^n}), (v_{j_1^n}, v_{j_2^n}), (k_{j_1^n}, k_{j_2^n}), (L_{j_1^n}, L_{j_2^n}), (T_{j_1^n}, T_{j_2^n}), (\tau_{j_1^n}^{\vec{\rightarrow}}, \tau_{j_2^n}^{\vec{\rightarrow}})\}$. Sur chacune de ces paires de statistiques $p_{c_{i^n}} \in \mathbb{P}_{c_{i^n}}$, nous pouvons calculer la valeur moyenne associée $\bar{p}_{c_{i^n}}$ et comme nous sommes dans un espace à deux dimensions, on peut en déduire la matrice de variance-covariance associée définie par $\sum_{p_{c_{i^n}}}$. Ainsi, grâce à 3.37, nous pouvons calculer la probabilité sur un couple de statistiques :

$$p_{p_{c_{i^n}}}^{\vec{\rightarrow}} = \frac{e^{-\frac{1}{2}t(p_{c_{i^n}} - \bar{p}_{c_{i^n}}) \sum_{p_{c_{i^n}}}^{-1} (p_{c_{i^n}} - \bar{p}_{c_{i^n}})}}{2\pi \sqrt{\det(\sum_{p_{c_{i^n}})}}} \quad (3.48)$$

Et de même qu'en 3.46, on a sur l'ensemble des deux branches filles la probabilité associée :

$$p_{p_{c_i}}^{\vec{\rightarrow}} = \prod_{n \in [t, t+\eta]} p_{p_{c_{i^n}}}^{\vec{\rightarrow}} \quad (3.49)$$

iv - Probabilité associée au voisinage de la cellule en division

Comme nous le verrons dans le chapitre 2, on a une forte déformation du maillage des cellules voisines d'une cellule en division. Il est donc important de prendre ce paramètre en compte dans le calcul de la probabilité d'une division. Nous avons mesuré dans la partie Estimation, la matrice de déformation moyenne

M^d associée aux voisinages d'une division, ainsi que sa matrice de variance-covariance associée $\sum_{M^d}^{-1}$.

Nous pouvons donc calculer la probabilité associée au voisinage ϑ_i^- de la cellule i . Il est bien sûr important de ne regarder que les voisines de la cellule en division sans celle-ci car elle pourrait influencer la mesure de la déformation par l'écartement de ses filles.

Soit pour $j \in \vartheta_i^-$, on a $p_{ij}^{\vec{\rightarrow}}(f)$ la probabilité associée au déformation du voisinage définie par :

$$p_{ij}^{\vec{\rightarrow}}(f) = \frac{e^{-\frac{1}{2}t \Delta_{ij}^d \sum_{M^d}^{-1} \Delta_{ij}^d}}{(2\pi)^{\frac{|X_i|}{2}} \times \sqrt{\det(\sum_{M^d})}} \quad (3.50)$$

où $\Delta_{ij}^d = (X_{j+} - X_{i+}) - (X_j - X_i) - M^d(X_j - X_i)$.

v - Probabilité associée à une division

De 3.49, 3.47 et 3.50, nous pouvons définir la probabilité qu'une cellule i ait deux successeurs (j_1, j_2) comme la multiplication associée aux statistiques des filles et à ses paires de statistiques, sur la déformation de l'ensemble de son voisinage. Soit :

$$p_i^{\vec{\rightarrow}}(f, \vartheta_i^-) = p_{c_i}^{\vec{\rightarrow}}(f) \times p_{p_{c_i}}^{\vec{\rightarrow}}(f) \times \prod_{j \in \vartheta_i^-} p_{ij}^{\vec{\rightarrow}}(f) \quad (3.51)$$

3.5.6.3 Probabilité de disparaître $p_i^{\vec{\leftrightarrow}}$

Normalement une cellule ne devrait disparaître que dans le cas d'une mort cellulaire, qui est un événement très rare dans le développement précoce de l'embryon. Mais compte tenu de la technique d'imagerie, nous coupons l'embryon pour ne conserver qu'une calotte sphérique, et un grand nombre de filles peuvent par conséquent sortir de notre volume d'imagerie. De plus, vu les artéfacts de la détection des centres, nous avons des faux positifs ou négatifs qui impliquent parfois qu'une branche n'ait pas de suite. Il nous faut donc définir une probabilité qu'une cellule ait de disparaître. Nous allons donc énumérer les deux cas possibles de disparition et nous allons diviser cette probabilité en deux parties, dans le cas où on est près du bord du volume imagé ou dans l'autre.

Ces calculs de probabilités se font sur les statistiques de mesures faites dans la partie Estimation de l'algorithme. On a donc un certain pourcentage de cellules qui peuvent disparaître, qui correspond à la probabilité associée. Soit $p_i^{b \leftrightarrow}$ la probabilité de disparaître pour une cellule près du bord et $p_i^{i \leftrightarrow}$ pour une cellule à l'intérieur du volume imagé.

La probabilité $p_i^{\vec{\leftrightarrow}}$ sera donc choisie en fonction de la mesure de distance au bord dans l'ensemble $\{p_i^{b \leftrightarrow}, p_i^{i \leftrightarrow}\}$.

3.5.7 Probabilités associées au *tracking backward*

De la même façon qu'au paragraphe 3.5.6, pour le cas de l'étude du *tracking backward* f^- , nous distinguerons deux cas :

- p_i^{\leftarrow} : la probabilité qu'une cellule ait un et un seul prédécesseur.
- $p_i^{\leftarrow\leftarrow}$: la probabilité qu'une cellule n'ait pas de prédécesseur (arrivée près du bord de l'image).

Soit

$$\begin{cases} \text{à } i^- \xleftarrow{f^-} i \text{ la probabilité associée } p_i^{\leftarrow}(f) \\ \text{à } \emptyset \xleftarrow{f^-} i \text{ la probabilité associée } p_i^{\leftarrow\leftarrow}(f) \end{cases} \quad (3.52)$$

avec $p_i^{\leftarrow}(f) + p_i^{\leftarrow\leftarrow}(f) = 1$.

3.5.7.1 Probabilité d'avoir un prédécesseur p_i^{\leftarrow}

La probabilité d'avoir un prédécesseur est définie par

$$p_i^{\leftarrow}(f, \vartheta_i) = p_i^{\leftarrow\leftarrow}(f) \times \prod_{j \in \vartheta_i} p_{ij}^{\leftarrow M_i}(f) \quad (3.53)$$

Nous ne détaillons pas le calcul ici mais il se fait de la même façon que dans le calcul de la probabilité d'avoir un successeur en 3.40.

La différence principale (et primordiale) dans le cas de l'étude du calcul des probabilités sur le passé est, qu'au lieu de savoir par les événements passés quel sera le meilleur événement futur, d'apprendre les événements futurs à *reculons* pour être capable à l'instant présent de donner la meilleure prédiction sur l'instant passé.

3.5.7.2 Probabilité d'apparition $p_i^{\leftarrow\leftarrow}$

Jusqu'à preuve du contraire, une cellule ne peut pas se créer toute seule, mais comme dans le cas de la disparition cellulaire 3.5.6.3 nous définissons la probabilité d'apparition d'une cellule $p_i^{\leftarrow\leftarrow}$, en fonction de la mesure de distance au bord dans l'ensemble $\{p_i^{b\leftarrow\leftarrow}, p_i^{b'\leftarrow\leftarrow}\}$.

3.6 Résultats

Nous avons implémenté l'ensemble des différentes mesures présentées dans la partie Maximisation ainsi que les probabilités associées aux cellules dans la partie Estimation.

Grâce à l'annotation effectuée par les biologistes sur un très grand nombre de liens (cf 3.2.2.2), nous avons pu valider la performance de chaque module de l'algorithme et ainsi évaluer sa qualité à chaque itération.

Compte tenu de l'énorme quantité de cellules et des nombreuses mesures à effectuer nous avons *découpé* l'embryon en plusieurs boîtes spatio-temporelles. Chacune des boîtes contient un grand chevauchement avec les autres, permettant ainsi de faire les calculs sur un maximum d'informations. Chacun de ces *morceaux* d'embryon est ensuite envoyé sur un processeur. Il faut ainsi environ dix processeurs pendant une dizaine d'heures pour effectuer une itération de l'algorithme. A la fin de chaque itération nous fusionnons les boîtes afin de reconstruire l'embryon complet. Nous avons effectué des tests permettant d'évaluer ce découpage en terme de perte d'informations, et il s'avère que nous ne perdons en moyenne qu'un très faible pourcentage de validité (0.001%).

Nous ne détaillons pas ici les difficultés liées à l'implémentation mais montrons simplement les résultats obtenus afin d'améliorer le *tracking* des cellules.

Nous effectuons lors des différentes itérations de l'algorithme différents tests sur le paramètre η . Il permet de définir la taille de la fenêtre temporelle à observer lors du calcul des probabilités.

Les résultats des différentes itérations sont représentées sur 4 figures :

- La figure 3.7 représente le pourcentage de cellules validées. Elle nous donne ainsi la validité générale du tracking à chaque itération en fonction de η .
- La figure 3.8 représente le pourcentage de cellules qui ont une probabilité certaine par rapport à l'ensemble des cellules.
- La figure 3.9 représente la validité de nos certitudes.
- La figure 3.10 représente la validité de la détection des mitoses à partir de la 4^{ème} itération.

Dans les quatre premières phases itératives, nous utilisons une version partielle de l'algorithme et y ajoutons à chaque itération différentes contraintes permettant de vérifier le paramètre η . Nous effectuons ensuite plusieurs itérations sur l'algorithme complet.

- **1^{ère} itération** : En première itération de l'algorithme, nous avons, à partir du tracking obtenu par registration, calculé pour différents η les différentes probabilités considérant un successeur. Nous n'avons donc pas pris en compte dans cette première itération les divisions cellulaires. Nous avons calculé et affecté le lien ayant la meilleure probabilité en itérant sur toutes les cellules dans un ordre aléatoire. Si une cellule c_1 trouve un lien temporel futur déjà occupé par une cellule c_2 mais que la probabilité de c_1 est meilleure, alors ce lien lui est affecté. Ainsi la cellule c_2 ne possède plus aucun lien temporel futur.

Le meilleur tracking est obtenu pour $\eta = 2$ et validé à 97.28%. Le nombre de cellules certaines est au maximum de 75.17% et leur validité est assurée à 99.84%. Cela revient à regarder 2 pas de temps dans le futur et 2 pas de temps dans le passé de la cellule. On constate sur la figure 3.7 que c'est un optimum certain, car en dehors de cette valeur, les résultats décroissent fortement.

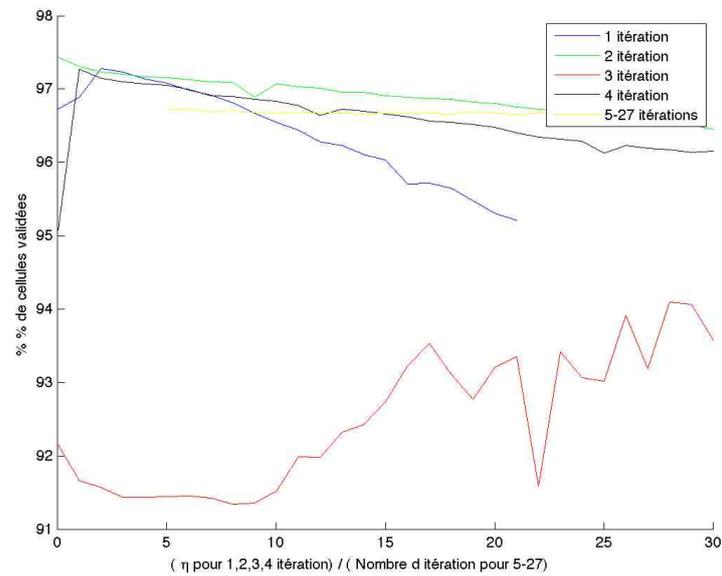


FIG. 3.7 – Pourcentage de cellules validées sur l'ensemble du tracking en ordonnée. L'axe des abscisses représente le paramètre η pour les 4 premières itérations et correspond au nombre d'itérations à partir de la 5^{ème}.

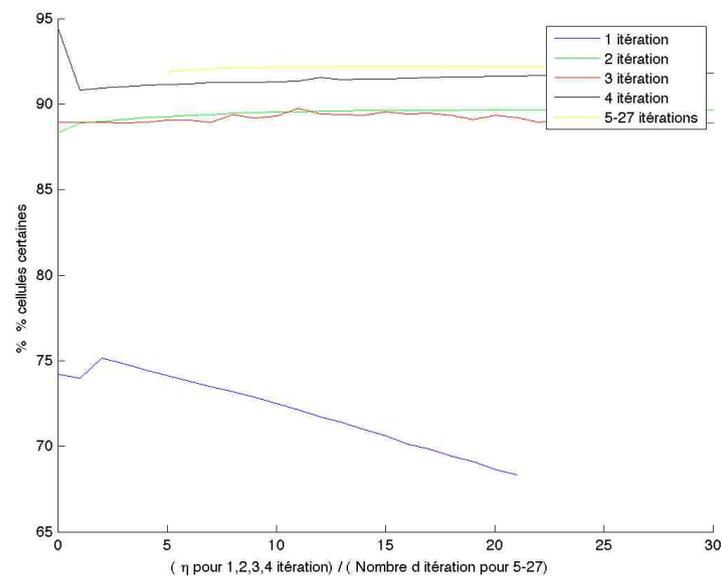


FIG. 3.8 – Pourcentage de cellules certaines en ordonnée. L'axe des abscisses représente le paramètre η pour les 4 premières itérations et correspond au nombre d'itérations à partir de la 5^{ème}.

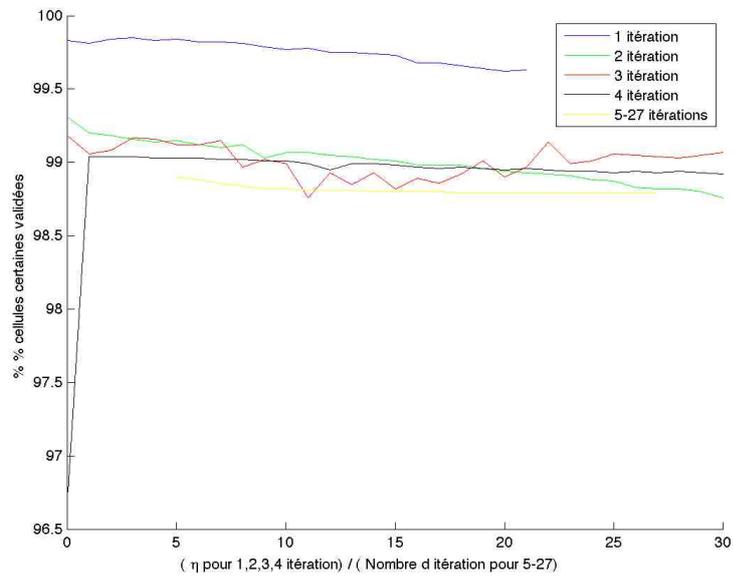


FIG. 3.9 – Pourcentage de validité des cellules certaines en ordonnée. L'axe des abscisses représente le paramètre η pour les 4 premières itérations et correspond au nombre d'itérations à partir de la 5^{ème}.

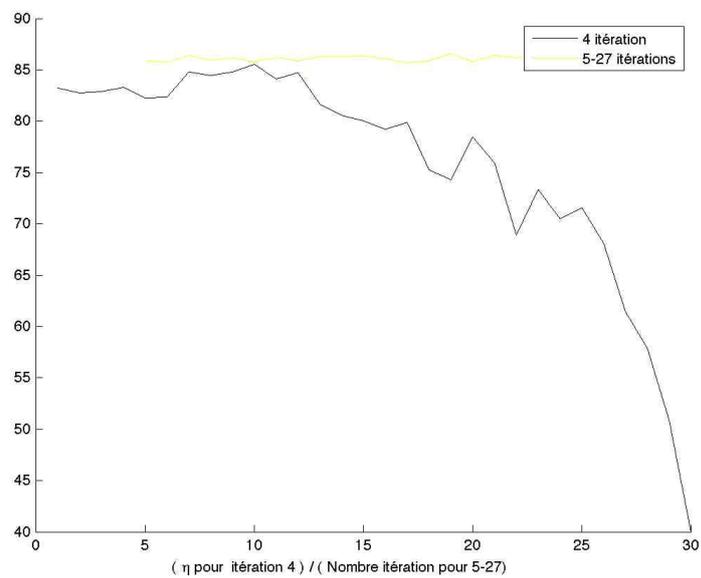


FIG. 3.10 – Pourcentage de validité des cellules en division en ordonnée. L'axe des abscisses représente le paramètre η pour la 4^{ème} itération et correspond au nombre d'itérations à partir de la 5^{ème}.

- **2^{ème} itération** : La deuxième itération prend en entrée le premier meilleur tracking obtenu en sortie de la première itération, soit pour $\eta = 2$. Nous continuons à ne travailler que sur un seul successeur. Nous cherchons, dans cette itération, à propager les certitudes de façon récursive. Ainsi l'ensemble des cellules certaines à la 1^{ère} itération (75.17%) sont fixées et ne peuvent plus changer leur lien. De cette information, leurs cellules voisines voient réduire les possibilités de liens temporels et ainsi peuvent devenir certaines. La récursivité permet de remettre dans la liste des cellules à *affecter* des cellules qui se verraient *prendre* leur liens temporels par une autre cellule. Nous obtenons un meilleur tracking pour $\eta = 0$, il est valide à 97.43%. Le nombre de cellules certaines a fortement augmenté à 89.66% mais par contre leur validité a légèrement baissé à 99.31%.
- **3^{ème} itération** : Nous ajoutons, à partir du meilleur tracking obtenu en 2^{ème} itération, les mitoses détectées par la méthode de l'*Active Machine Learning* proposé dans le chapitre 2. Certaines de ces divisions cellulaires ont déjà été validées par les biologistes, mais afin de vérifier si l'algorithme les détecte correctement, nous ne rajoutons ici que celles qui n'ont pas été annotées. Nous fixons dans un premier temps ces doubles liens et effectuons une itération en ne cherchant toujours qu'un successeur à partir de ce jeu de données complété. Nous constatons ici que le tracking dépend énormément de η puisque nous obtenons le meilleur résultat pour $\eta = 29$:
 - Cellules validées : 94.09%.
 - Cellules certaines : 89 %.
 - Cellules certaines validées : 99.03%.
- **4^{ème} itération** : Dans cette itération nous ajoutons le module de détection des mitoses. On constate que pour $\eta < 2$ les résultats de détection des cellules décroissent énormément. Par contre, la détection des divisions cellulaires est optimal pour $\eta = 11$, ce qui signifie que l'algorithme a besoin de connaître l'information temporelle sur une vingtaine de pas temps pour arriver à mieux détecter les mitoses Nous prenons le meilleur résultat pour le tracking global pour $\eta = 2$:
 - Cellules validées : 97.27%.
 - Cellules certaines : 90.8 %.
 - Cellules certaines validées : 99.04%.
 - Mitosies validées : 84.03%.
- **5^{ème} – 27^{ème} itérations** : Nous avons fixé $\eta = 2$ pour les itérations suivantes qui semble être la meilleure valeur du paramètre au vu des tests effectués sur les itérations précédentes. L'axe des abscisses représente le nombre d'itérations pour la courbe jaunes sur les 3 figures : 3.7, 3.8, 3.9 . A partir de la 5^{ème} itération, l'algorithme fonctionne dans sa version complète. Il prend toujours en entrée le tracking en sortie de l'itération précédente. L'algorithme converge au meilleur résultat \hat{f} après 24 itérations au total en obtenant les résultats suivants :

- Cellules validées : 96.73%.
- Cellules certaines : 91.86 %.
- Cellules certaines validées : 98.9%.

Le pourcentage de cellules certaines augmente fortement au cours des différentes itérations, bien que leur validité diminue légèrement (figure 3.11). Il est de plus intéressant de constater qu'à la première itération l'ensemble des cellules certaines sont presque totalement validées. Nous pouvons donc à partir de cette observation classer les certitudes en fonction des itérations. Il est intuitivement naturel qu'une cellule certaine à la $n^{\text{ème}}$ itération, pour n grand, et ne l'étant pas à la $(n - 1)^{\text{ème}}$ itération, n'ait pas autant de certitude que celles certaines à la 1^{ère} itération.

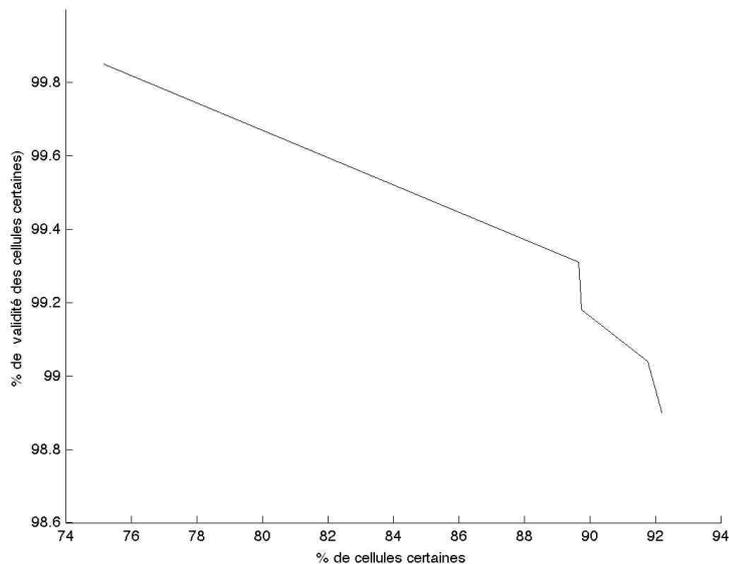


FIG. 3.11 – Pourcentage de validité certaines au cours de l'évolution de l'algorithme par rapport à leur validité.

Remarque : Il est intéressant de constater que nous n'obtenons jamais 100% de validité sur chacune des itérations. Pour avoir participé personnellement à la validation, au côté des biologistes, d'un grand nombre de liens, il arrive que même un expert ne sache pas réellement quel est le futur ou le passé de la cellule et donc décide de la validité d'un lien avec peu de conviction. De plus la validation de l'ensemble des 80000 liens, même effectué par plusieurs personnes, demande une bonne concentration de nombreuses heures. Aussi, nous pouvons nous poser cette question : "Est-ce que l'on est sûr que tous les liens validés sont valides ?".

3.7 Le langage MorphoTrack

Toute la procédure E'M' montre que son résultat est probabiliste :

- la trajectoire $X_i(t)$ de chaque cellule est estimée à partir d'un vecteur vitesse déterministe V_i et la partie stochastique de la dynamique est exprimée par une matrice de variance-covariance \sum_i .
- le lignage proposé vise à atteindre le maximum de vraisemblance mais d'autres lignages pourraient être également considérés.

Le langage MorphoTrack a pour objectif de représenter les résultats du tracking comme un lignage cellulaire spatiotemporel. Il ne fait que reprendre les différentes variables

La grammaire du langage est défini par : $\{A\}$ est une liste de A ; $(.,.)$ est un couple ou plus généralement un n-uple d'un produit cartésien ; $|$ est une définition par cas, $(* *)$ représente des commentaires. On obtient :

- Lineage = {Cell}
- Cell={Identity, StateVector, Velocity, VarianceCovariance, DeformationMatrix, Domain, Previous, Successors}
- Domain={NameField,{Domain}} (* un domain can have recursively subdomains *)
- Previous=(Identity, Probability) | (Outside, Probability)
- Successors=(Outside, Probability) | (Identity, Probability) | Division
- Division=(Identity, Probability) | (Identity, Probability)

Variante pour prendre en compte les autres cellules qui auraient pu être sélectionnées :

- PossibleSuccessors={successors}
- PossiblePrevious :{Previous}

3.8 Interprétation et Conclusion

Nous avons proposé pour le tracking un nouvel algorithm, *MorphoTrack*, qui appartient à une famille d'algorithmes E'M' qui peuvent être considérés comme des ε -algorithmes EM. Avec l'amélioration des images et des traitements, il est conjecturé que l'algorithm E'M' converge vers un algorithm EM. Le résultat de *MorphoTrack* est un lignage cellulaire spatiotemporel où chaque cellule à chaque instant est décrite non seulement par son vecteur d'état mais aussi par les lois de probabilité de son déplacement (une gaussienne avec une matrice de variance-covariance autour d'une vitesse moyenne), par la matrice de déformation de son voisinage. Le résultat, comme annoncé dans l'introduction n'est pas un lignage spatiotemporel certain mais seulement un maximum de vraisemblance, où des variantes de lignages ont aussi une probabilité d'être le bon. Le langage *MorphoTrack* exprime de façon rigoureuse les incertitudes qui subsistent.

Nous pouvons interpréter les résultats ainsi : l'algorithm *MorphoTrack* fonc-

tionne comme un algorithme de programmation par contrainte alors qu'il calcule des probabilités. Chaque fois que le calcul fait apparaître une probabilité de 1 (en double précision) pour la transition $(i, t) \rightarrow (j, t + 1)$, un lien "sûr" $f(i) = j$ est créé et cette certitude est propagée, créant d'autres liens "sûrs". Ce processus de propagation des certitudes fonctionne jusqu'à créer 95% du tracking.

Sur la base de cette interprétation, nous pouvons dégager les remarques suivantes :

- Avec l'amélioration future de l'imagerie et des algorithmes de détection de centres, il est conjecturé que la propagation de certitude ira jusqu'à 99%. L'épsilon de l'algorithme E'M' (cf. 3.3.1.3) convergera vers 0 et, en ce cas, l'algorithme E'M' convergera vers l'algorithme EM.
- Il faut encore noter que l'algorithme *MorphoTrack* n'utilise pas les trois hypothèses d'indépendance (cf 3.3.3). Il se contente dans $p(f)$ de sélectionner tour à tour, des milliers de fois, des fenêtres temporelles $[\eta; \eta]$ autour de la cellule i pour ôter ce qui n'est pas possible et parfois conclure qu'il ne reste qu'une solution unique pour $f(i)$, et propager à nouveau.
- Il convient finalement de remarquer que la fin du travail probabiliste ne concerne que très peu de cellules disséminées, pour lesquelles les hypothèses 1, 2, 3 deviennent peu contestables.

En conclusions, nous pouvons dégager les principaux avantages de *MorphoTrack* :

- il peut s'adapter à tout type de *tracking* d'éléments.
- il est probabiliste au sens ou, à chaque *tracking* effectué, il donnera une valeur de probabilité associé au *tracking*, ce qui permet très rapidement de pouvoir faire des comparaisons de résultat.
- des probabilités sont aussi associées à chaque cellule, ainsi nous aurons des *branches* du lignage cellulaire de la validité desquelles nous pourrons être sûr et des branches sur lesquelles nous pourrons avoir des doutes.
- maximisation de la vraisemblance : nous allons utiliser un algorithme Estimation-Maximisation pour qu'à chaque itération on puisse estimer dans des zones statistiquement homogènes certains paramètres (mouvement cellulaire, déformation du voisinage, caractéristique des divisions cellulaires, etc.) et maximiser leur vraisemblance par la suite.
- nous pouvons extraire les dynamiques stochastiques de ce *tracking* pour tout déviation à la vraisemblance.

3.9 Démonstration

Soit A_i la matrice telle que

$$\begin{aligned}
 A_i &= \frac{\sum_{j \in \vartheta_i} \delta_{X_{ij}} \otimes \delta_{X_{ij}}}{|\vartheta_i|} \\
 &= \frac{\sum_{j \in \vartheta_i} (\delta_{X_{ij}} - \overline{\delta_{X_{ij}}}) \otimes (\delta_{X_{ij}} - \overline{\delta_{X_{ij}}})}{|\vartheta_i|} + \overline{\delta_{X_{ij}}} \otimes \delta_{X_{ij}} + \delta_{X_{ij}} \otimes \overline{\delta_{X_{ij}}} - \overline{\delta_{X_{ij}}} \otimes \overline{\delta_{X_{ij}}} \\
 &= \overline{A_i} + 2\overline{\delta_{X_{ij}}} \otimes \frac{\sum_{j \in \vartheta_i} \delta_{X_{ij}}}{|\vartheta_i|} - \overline{\delta_{X_{ij}}} \otimes \overline{\delta_{X_{ij}}} \\
 &= \overline{A_i} + \overline{\delta_{X_{ij}}} \otimes \overline{\delta_{X_{ij}}}
 \end{aligned}$$

3.10 Symboles

- $\sum_{i \in B} A_i$ la somme des éléments A_i pour i appartenant à l'ensemble B .
- $\prod_{i \in B} A_i$ le produit des éléments A_i pour i appartenant à l'ensemble B .
- $A \otimes B$ produit tensoriel ou produit de Kronecker de deux matrices A et B .
- Soit A^{-1} l'inverse de la matrice A .
- ${}^t A$ la transposée de la matrice A .
- $\det(A)$ le déterminant de la matrice A .
- Soit \overline{A} la moyenne de la matrice (ou du vecteur) A .
- Soit \hat{A} la meilleure matrice A .
- $t : \dots$ représente l'abstracteur de Russell (t tel que...)
- a^{\rightarrow} représente une mesure a sur une cellule qui a un successeur.
- a^{\rightrightarrows} représente une mesure a sur une cellule qui a deux successeurs.
- $a^{\rightarrow\rightarrow}$ représente une mesure a sur une cellule qui n'a pas de successeur.
- a^{\leftarrow} représente une mesure a sur une cellule qui a un prédécesseur.
- $a^{\leftarrow\leftarrow}$ représente une mesure a sur une cellule qui n'a pas de prédécesseur.

Cell division orientation in zebrafish brain early embryogenesis : a systematic analysis based on *in toto* imaging and automated cell tracking

Résumé

A partir de données d'imagerie 4D obtenues en microscopie 2-photons de l'embryon de poisson zèbre depuis la gastrulation jusqu'à mi-somitogénèse, nous avons identifié les paramètres mesurables pertinents de la division cellulaire permettant de caractériser la contribution des divisions cellulaires à la morphogénèse précoce de l'embryon. Jusqu'à aujourd'hui, l'ensemble des études sur ces phénomènes restent très qualitatives, hétérogènes au plan spatio-temporel et basés sur des stratégies de mesures diverses qui rendent difficiles les comparaisons et la généralisation. Nous proposons un protocole standard et automatisé de mesure des caractéristiques de la division cellulaire au cours de la morphogénèse, fondé sur des combinaisons d'algorithmes utilisant des positions cellulaires, leur tracking et un référentiel. Nous avons mesuré des caractéristiques en termes de vitesse, d'orientation du plan de division et de persistance du voisinage. Nous avons trouvé dans ce contexte des relations entre le plan de division et la couche cellulaire à laquelle appartient la mère, ainsi que des orientations préférentielles par rapport à son mouvement. Lorsque la vitesse de la cellule mère est relativement importante ($> 1\mu m$), le dipôle créé par les positions des cellules filles s'oriente perpendiculairement au mouvement de la mère. Nous constatons dans les minutes qui suivent la cytokinèse une ré-orientation de ce dipôle dans l'alignement du vecteur défini par le mouvement. Nous avons corréler l'orientation de ce dipôle par rapport à la couche cellulaire, et constatons une préférence forte d'orientation de division dans la couche au début de la gastrulation. Nous proposons un modèle basé sur les équations de conservations d'énergie et de mouvements qui confirme ces mesures.

Chapitre 4. Cell division orientation in zebrafish brain early embryogenesis : a 108 systematic analysis based on *in toto* imaging and automated cell tracking

Authors : Emmanuel Faure^{1,2}, Raphaël Keller^{1,2}, Benoit Lombardot^{1,2}, Thierry Say^{1,2}, Miguel Luengo-Oroz^{1,3}, Louise Duloquin^{1,4}, Paul Bourgine^{1,2} and Nadine Peyri ras^{1,4}.

- ¹ ISCPIF - Complex Systems Institute of Paris Ile de France, ISC-PIF, 57-59 rue Lhommond, 75005 Paris, France
- ² Centre de Recherche en Epist mologie Appliqu e - CNRS - Ecole Polytechnique, 32 Boulevard Victor, 75011 Paris, France
- ³ Biomedical Image Technologies Lab of ETSI Telecomunicacion, Universidad Politecnica de Madrid, Av. Complutense SN, 28040 Madrid, Spain.
- ⁴ D veloppement, Evolution, Plasticit  du Syst me Nerveux - CNRS , Institut de Neurobiologie Alfred Fessard, Avenue de la Terrasse, 91198 Gif Sur Yvette, France

Sommaire

4.1 Results	110
4.1.1 Cell behaviours analysis from the phenomenological reconstruction of 4D image data sets	110
4.1.2 Cell morphodynamics characteristics and their correlation with the cell layers topology	111
4.1.3 Quantitative measurements defining cell division characteristic features	111
4.1.4 Cell division plane orientation is biased during gastrulation and correlates with mother cell average speed	112
4.2 Discussion	112
4.3 Figure	113
4.3.1 Figure1	113
4.3.2 Figure2	115
4.3.3 Figure3	117
4.3.4 Figure4	119
4.4 Supplementary Materials	121
4.4.1 The different embryos	121
4.4.2 Setting up a spatial and temporal frame of embryogenesis	121
4.4.3 Measuring angles	126
4.4.4 Study of cell movement relative to the layer	126
4.4.5 Study of neighborhood deformation	127
4.4.6 Cells undergoing a division	127
4.4.7 Model of cellular division	128
4.4.8 Mitosis rates	129

Embryogenesis of vertebrate embryos involves high rate proliferation. In the zebrafish, the first 24 hours of development at 28 C lead to some 50 000 cells (Kane (1999)). In addition to increasing the overall cell number, cell division contributes to shaping the tissues through the orientation of the division plane (Ahringer

(2003); Sausedo *et al.* (1997), Schoenwolf & Alvarez (1989); Wei & Mikawa (2000)). Demonstrating a bias in this respect requires accurate and precise measurements of solid angles in a well defined referential and a systematic analysis of cell divisions through time and space. Such approach has been so far largely limited by available measurements thus focusing on limited spatio temporal sequences (Concha & Adams (1998), Woo & Fraser (1995), Geldmacher-Voss *et al.* (2003), Poggi *et al.* (2005)).

Furthermore, before mitosis, cells undergo dramatic changes in terms of adhesion, shape and mobility thus perturbing their neighborhood (Thery & Bornens (2006)). Their reintegration into the tissue pattern after mitosis is crucial (Lecuit (2008)).

We investigated these features and their correlation with morphogenetic events in the zebrafish anterior brain imaged by two-photon laser scanning microscopy, starting prior gastrulation until mid-somitogenesis (Kimmel *et al.* (1995)). The obtained 4D data sets were used to assess cell position, cell trajectories and cell division characteristics, using the most efficient algorithmic strategies currently available (England *et al.* (2006), Melani *et al.* (2007)). This unique type of data allowed us to systematically investigate cell division plane orientation and cell neighbourhood persistence at the time of cell division as well as daughter cells trajectories compared to their mother cell.

Statistical analysis of cell division features requires defining a referential for the embryo and its morphological characteristics (Adams (1996)). During gastrulation, cell movements are confined within cell layers known to correspond to different fates and shown here to have different characteristics in terms of average cell speed and neighbourhood persistence.

In this context, cell division is both characterized by intrinsic features of mitoses and their cell neighbourhood context. For the explored developmental sequences, covering brain regions during pre-gastrulation, gastrulation and early somitogenesis, biased cell division orientation correlates with cell velocity and, above a speed of 1micron per minute, cell divisions preferentially send daughter cells perpendicular to their mother cell trajectory. In addition, daughter cells are confined within their mother's layer during gastrulation for all layers, and through the whole developmental sequence for the embryo surface cell layer - the so-called enveloping layer or EVL. This confinement is even reinforced in outer layers during gastrulation by daughter cells displacement during the first 20 minutes after cytokinesis. Finally, our quantitative strategy allowed showing that *zoep* homozygous mutants defective in nodal signalling pathway transduction, significantly differ from wild type embryos by cell velocity that does not rise above 1,2 microns per minute. Correlating with this low speed, division orientation is random and cell displacement after division does not significantly align daughter cells with their mother's trajectory.

4.1 Results

4.1.1 Cell behaviours analysis from the phenomenological reconstruction of 4D image data sets

The systematic analysis of cell division features underlying morphogenetic processes was achieved from measurements provided by the algorithmic processing of 4D image data sets. The latter were obtained from wild type (wt) or mutant (*zoep*) zebrafish embryos with stained nuclei (H2B/mcherry fusion) and membranes (farnesylated eGFP) expressed through mRNA injection at the one cell stage. 4D imaging was achieved by two-photon laser-scanning microscopy from the sphere stage, i.e. 4 hours post fertilization (hpf) at 28°C, until one third of somitogenesis (10-12 somites) (Figure 1A-B, Video S1-4). 4D data sets were submitted to a chain of algorithms achieving the automated detection of nucleus centre and cell tracking, providing systematic measurements for cell position, cell trajectory and cell division. We named augmented phenomenology (Video S1-5) the superimposed raw (Figure 1A-B) and reconstructed data (Figure 1C-D) providing powerful means for validating and analysing *in vivo* cell behaviours. Measurement of cell division features and comparison of data obtained from different embryos required defining a common spatial and temporal referential.

This was achieved by positioning the imaged volume on a sphere approximating the embryo shape oriented according to the animal pole-vegetal pole and dorsal-ventral axis (Figure 1E). Five different imaged data sets corresponding to five different embryos (3 wt and 2 *zoep* mutants) were matched through 150 microns depth from the animal pole (Figure 1E) encompassing the zebrafish presumptive ventral forebrain (reference). Because cell diversity during gastrulation is organized in concentric layers, single cell layers were calculated from the inter nucleus distance and displayed in different colours (Figure 1F, G). From the onset of gastrulation until the end of the time lapse, the embryo showed a so-called bilateral symmetry plane, rather identified as a surface separating two sides at the level of the midline (Figure 1A-D, 1F). The cell layers number changes through gastrulation movements of epiboly and convergence extension. As shown on a sagittal section during gastrulation, the involution of hypoblast cells leads to the addition of 2-3 cell layers (Figure 1G and Figure 2). Furthermore, single cell layers thickness changed through time as shown in Figure 1H as cell division proceeded decreasing the distance between nuclei. As shown in Figure 1H, I, the distance between nuclei changed throughout gastrulation until a minimal distance of about 10 microns. Wild type and homozygous *zoep* mutants did not significantly differ in this respect indicating similar average proliferation rate throughout the explored spatio temporal developmental sequence (Figure 1H, I). As the imaging procedure started at the sphere stage (4 hpf), embryonic cells were expected to be within their 11th-12th division cycle.

4.1.2 Cell morphodynamics characteristics and their correlation with the cell layers topology

Cell morphodynamics measurements were derived from the phenomenological reconstruction of 4D data sets. They provide a quantitative description of cell layers characteristic features in space and time. The latter include average cell speed within the layers, cell trajectory orientation relative to the plane tangent to the sphere (Figure 1G), and cell neighbourhood deformation calculated from a deformation tensor (Supplementary Materials). Cell speed calculated from its displacement during the past 20 minutes was averaged within each layer (Figure 2A, B). Averaged cell speed varied from 0,5 microns to 2 microns per minute in wild-type embryos and did not rise above 1, 2 microns per minute in *zoep* mutants. Cell speed distribution in time and space revealed the highest speed of hypoblast cells throughout gastrulation and the acceleration of epiblast by the end of gastrulation (Figure 2A). These features were dramatically dampened in *zoep* homozygous mutants (Figure 2B). Cell displacement orientation showed that, during gastrulation, cells were confined within a layer (Figure 2B), contrasting with movements across layers before the onset of gastrulation and during early somitogenesis (Figure 2A). *zoep* mutants were characterized by an overall average orientation largely confined within layers throughout the spatiotemporal sequence, except for very early steps of epiboly. However, averaging might hide noisy behaviour. This hypothesis is supported by the observation of cell neighbourhood deformation : At least prior gastrulation, deformation was higher in *zoep* compared to wild type embryos (Figure 2E, F). At later stages, this difference was at least significant for the EVL and the most superficial layer of the epiblast. The above characterization of cell morphodynamics changes in space and time provided a framework to correlate cell division features and tissue morphogenesis.

4.1.3 Quantitative measurements defining cell division characteristic features

Mitoses were systematically identified by detection and tracking algorithms (Figure 3A). Cell division features were described using the vector V characterizing the mother cell displacement and the vector W characterizing the daughter cells relative position. In addition, cell neighbourhood deformation at the time of cell division was analyzed as in Figure 2 E, F through the deformation of a tensor calculated with a Delaunay triangulation (Figure 3C). The norm of vector W increased very rapidly at the time of cytokinesis and small values of W over cell diameter (Wd) were observed only in time-lapse schemes with a very small time step (10 seconds). These values were not revealed anymore when fit was larger than 2 minutes (Figure D, E). Wd plotted as a function of time increased rapidly within the first 10-12 minutes after cytokinesis to attain its maximum (Figure 3F). During the next 10 minutes, Wd plotted as a function of time consistently diminished before getting into a third phase of slow increase. These features did not

differ between wild type and *zoep* mutants embryos (Figure 3F and 3G). These observations suggest that the shape of the curve Wd as a function of time might be a robust feature of the cell division process.

4.1.4 Cell division plane orientation is biased during gastrulation and correlates with mother cell average speed

Assessing the orientation of cell division, calculated relatively to mother cell speed, was achieved by plotting the percentage of cell divisions with V,W angle exceeding 45° as a function of Wd , revealing that cell division orientation was biased towards being orthogonal to mother cell orientation when mother cell speed was over 1 micron per minute (Figure 4A, B). Furthermore, as Wd increased over time after cell division, daughter cells tended to align along their mother's trajectory, again for the highest speed values (Figure 4A). In addition, daughter cells were confined within their mother's layer, for all layers during gastrulation and through the whole developmental sequence for the EVL (Figure 4C, D). This confinement was even reinforced in outer layers at gastrulation stages by daughter cells displacement during the first 20 minutes after cytokinesis (Figure 4E, F). The major difference between wild type and *zoep* mutants revealed by these measurements relates to cell speed, which is much less in *zoep* mutants compared to wild type, correlating with largely random division orientation relative to the mother cell trajectory. In addition, mother cell and daughter cell neighbourhood deformation suggests that in *zoep* mutant embryos, changes in cell neighbourhood within the tissue were more extensive in mutants compared to wild type mitosis, indicating a difference in the biomechanical properties of the tissues.

4.2 Discussion

We describe here original methodologies and tools to achieve a systematic analysis of cell division features from a type of data unavailable so far. We have for this study 5 data sets with a measured precision and accuracy for cell detection and tracking that allows for the first time a systematic statistical analysis of cell division orientation relative to mother cell direction and speed. In addition, we show that a spatial and temporal referential can be defined that refers to embryonic axis and cell layers, automatically defined, allowing to compare data sets. We develop in supplementary data a theoretical model for cell division describing it as the separation of objects in a complex fluid. In this framework, above a certain cell velocity, when the mother cell divides in daughter cells of similar size, the conservation of movement quantity and mechanical energy is indeed expected to lead to an orientation of the daughter cell dipole perpendicular to the mother direction. In case of a slight asymmetry in cell volume, the angle formed by the daughter cells dipole is expected to slightly deviate from the orthogonal orientation. This model fits with the observation that for the highest mother cell velocity, a

bias in the orientation of the division plane is observed with the daughter's dipole oriented orthogonally to the mother's direction. The other main spatial temporal correlation between cell division features and morphogenetic events relates to the organization of the embryo in cell layers with characteristic average speed and neighbourhood persistence. Gastrulation movements are characterized by cell displacement within the layers, cell division orthogonal to mother's trajectory, confinement within the layer of daughters cells and even reorientation of the daughter cells dipole to align it with the mother trajectory. These movements are consistent with an overall intercalation of cells close to the midline where cell velocity is the highest and arrangement of cell clones in what was described in previous studies as clonal strings.

4.3 Figure

4.3.1 Figure1

- **A : wildtype, B : *zoep*** : Embryo rawdata at 12 somites. View of bilateral symmetry plane (BSP) as a white line from from animal pole at the beginning of somitogenesis. 3D volume rendering of both membranes and nuclei. The volume is cut at $100 \mu m$ from the animal pole.
- **C : wildtype, D : *zoep*** : Embryo reconstructed at 75% Epiboly. One hour of trajectory is plot. The colour corresponds to the speed, from green for lower speeds to red for the highest.
- **E** : Global referential definition. Embryos are focused in the center of the yolk (represented as a yellow sphere), reoriented along the BSP plane. At 4hpf cells nuclei and membranes are superposed on the yolk. Each embryo is recalibrated in space cut at $150 \mu m$ from animal pole (AP)
- **F** : Ortho-slice at $150 \mu m$ superposed reconstructed centers of nuclei. Each colour corresponds to a layer detection. Yellow : EVL. Blue : hypoblast.
- **G** : $10 \mu m$ of 3D volume rendering in XY with coloration by cell layer. Oriented in XZ of the BSP at the same moment of F.
- **H : wildtype, I : *zoep*** : Average cell radius depending on embryo development. The radius is calculated by measuring distance to neighbors. The neighbours are calculated from the detected centers with Delaunay triangulation. Hypoblast enters the volume 2 hours after beginning of image acquisition.

Chapitre 4. Cell division orientation in zebrafish brain early embryogenesis : a
 114 systematic analysis based on *in toto* imaging and automated cell tracking

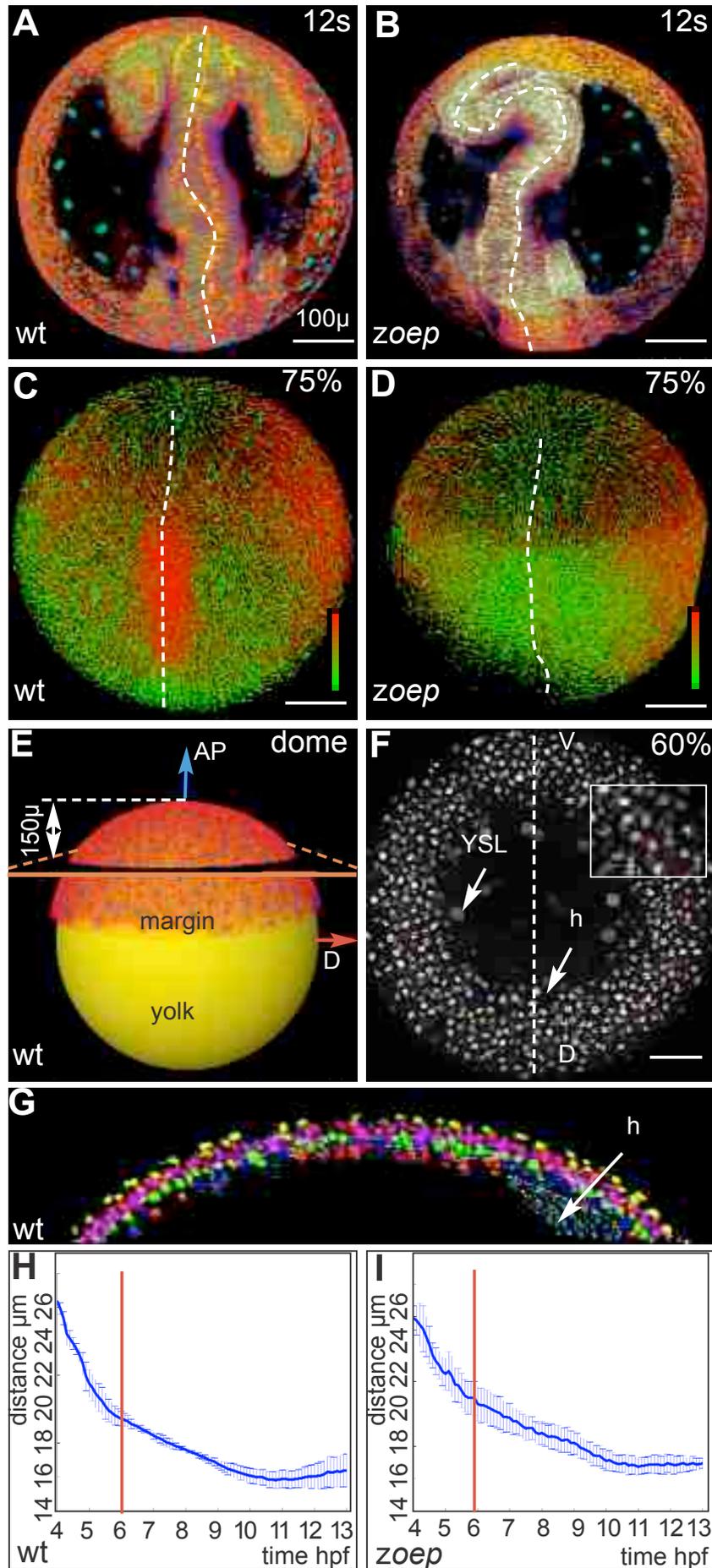


FIG. 4.1 – Context of experiences

4.3.2 Figure2

- **A : wildtype, B : zoep** : Cell speed during embryo development. X axis corresponds to embryo development time in hours post-fecondation. Y axis corresponds to different layers. From the first, the EVL, to the tenth and deepest layer. The colour represents the velocity in $\mu m/minute$.
- **C : wildtype, D : zoep** : Cell movement orientation during embryo development. X,Y correspond to the same axes at A,B. We measure the angle between cell movement averaged on 20 minutes and the tangent plane, so we study the cells movement in their layers. The colour represent the angle in degrees.
- **E : wildtype, F : zoep** : Cell neighbours deformation during embryo development. X axis corresponds to the embryo development time in hours post-fecondation. Y corresponds to the norm of the deformation matrix. Curves with different colours represent different layers. From the first, the EVL, to the tenth, the deepest layer.

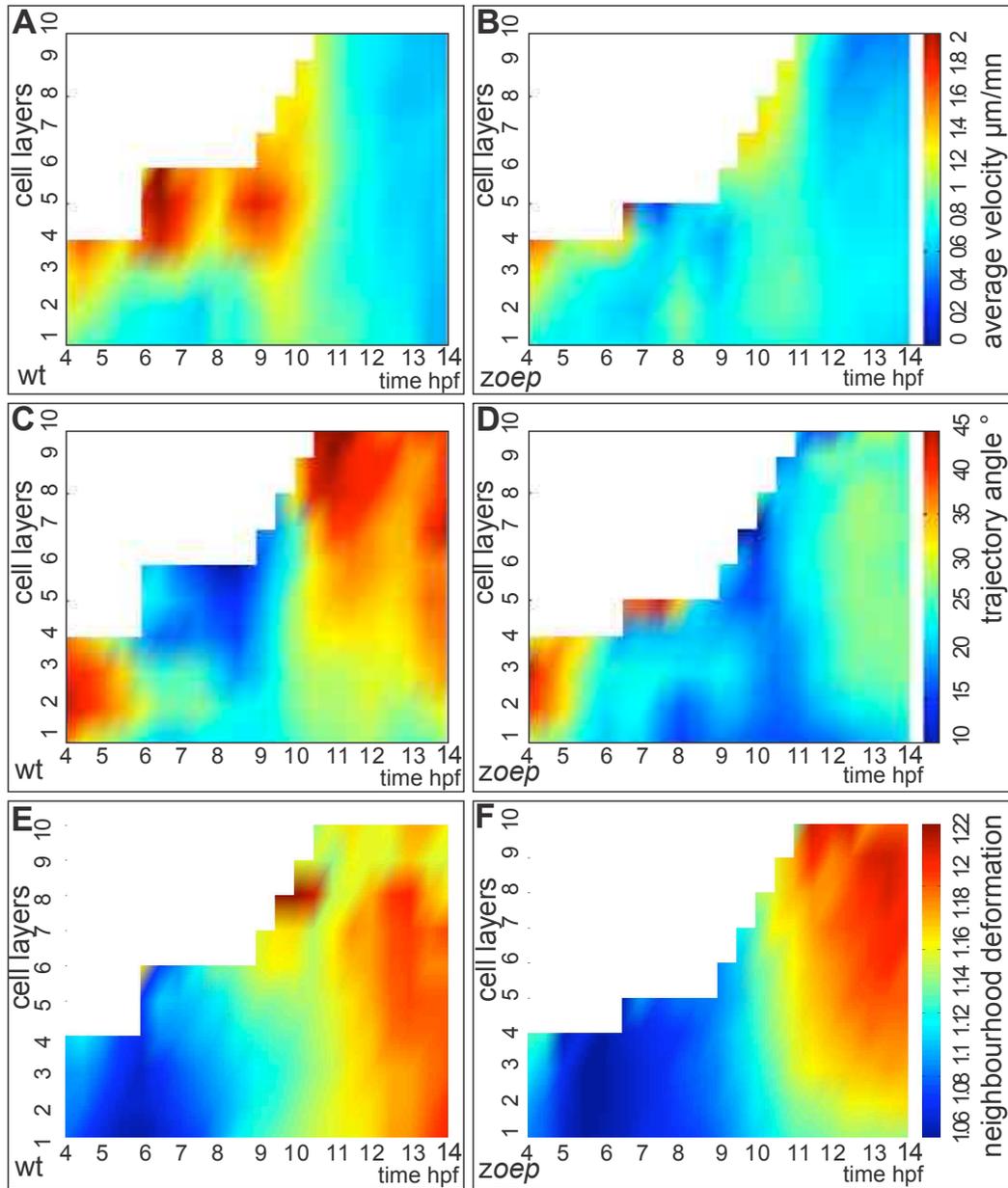


FIG. 4.2 – Cells measurement, Velocity, Orientation, Deformation

4.3.3 Figure3

- **A** : Trajectory of cell division each $\Delta t = 10s$. Nucleus cells are represented in 3D volume rendering in yellow. Yellow squares represent the nucleus center of the two daughter cells 10 seconds after division. Line colours from red to yellow represent cells speed.
- **B** : Graphic representation of a division. Trajectory in blue. Yellow spheres represent cells.
 - $\rightarrow V$ cell movement average.
 - $\rightarrow n$ normal vector to the tangent plan of the yolk at each cell
 - $\rightarrow W$ stands for dipole segments between the two daughters just after division.
- **C** : Mesh neighborhoods visualization before and after division. The green circles represent nuclei. Blue line correspond to the link between each neighbors of the cell division.
- **D** : Histogram of the dipole length. The dipole is taken Δt after division. On the X axis, Wd corresponds to dipole length normalized by the average size of the cells at each timestep. 8 experiments are represented by different histogram colours with their associated Δt . In the smaller figure inside D. X axes represent the time after division in minutes. Wd is measure in Y. All experiences with different Δt represented their pic with a green circle are link in same blue lines.
- **E** : Dipole length during the first 100 minutes after division for each experiences. X axes correspond the the logarithm of the time after division in minutes. Y axes correspond the logarithm of dipole length normalize by cell size. Wildtype in red and *zoep* in blue.
- **F** : Dipole lgent during the first hour after division for validate cells. In red the extension of the dipole with their associate the associate fitting curve in black.. In green is represented the Standard deviation the associate fitting curve in black.

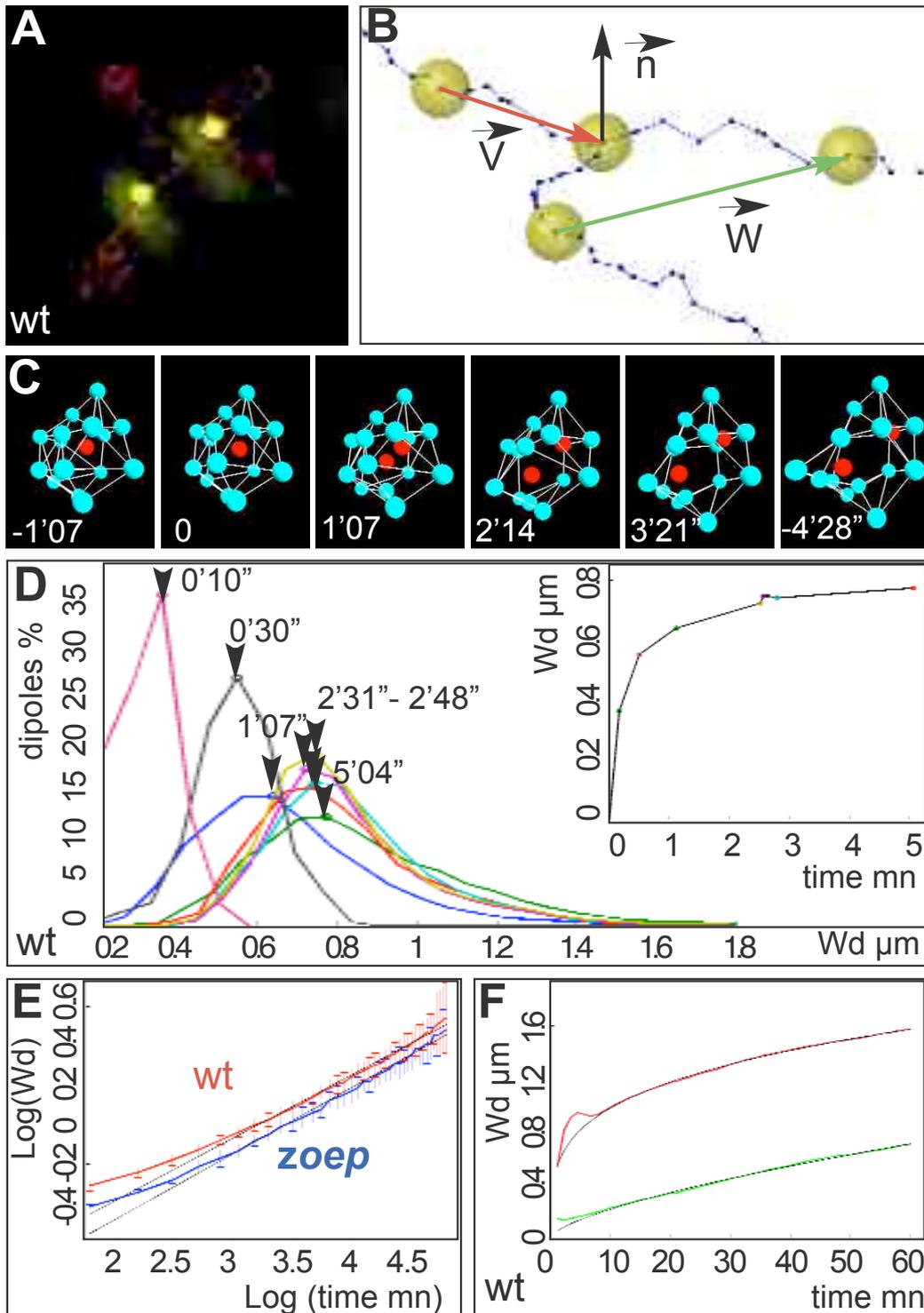


FIG. 4.3 – Cells divisions descriptions

4.3.4 Figure4

- **A : wildtype, B : zoep** : Influence of velocity on dipole orientation. X axis corresponds to Wd. We measure the angle between the dipole created by the two daughter cells fit after division and the vector velocity of the mother cell. Y axis corresponds to the percentage of this solid angle above 45° . The different lines coloured depending on speed cell movement are estimated from a linear regression. The colour bar stands for the velocity in $\mu\text{m}/\text{minute}$.
- **C : wildtype, D : zoep** : Dipole orientation during embryo development. X axis corresponds to the time embryo development in hours post-fecundation. Y axis corresponds to the different layers from the 1st, the EVL, to the tenth and deepest layer. We measure the angle between the dipole just after division and the tangent plane. The colour represent the angle in degrees.
- **E : wildtype, F : zoep** : Dipole reorientation after division. X,Y correspond to the same axes at A,B. We measure the possible reorientation of the angle in C,D one hour after division. The colour represents the ratio of reorientation. This ratio corresponds to the slope from the linear regression on the different angles of the dipole until 20 minutes after division. (0 correspond to no-reorientation, >0 increase of angle value, <0 decrease of angle value).
- **G : wildtype, H : zoep** : Cell neighbourhood deformation during the division. X axis corresponds to time in minutes before and after division. Y corresponds to the norm of the deformation matrix. Curves and respective error bars in different colours represent different layers.

Chapitre 4. Cell division orientation in zebrafish brain early embryogenesis : a 120 systematic analysis based on *in toto* imaging and automated cell tracking

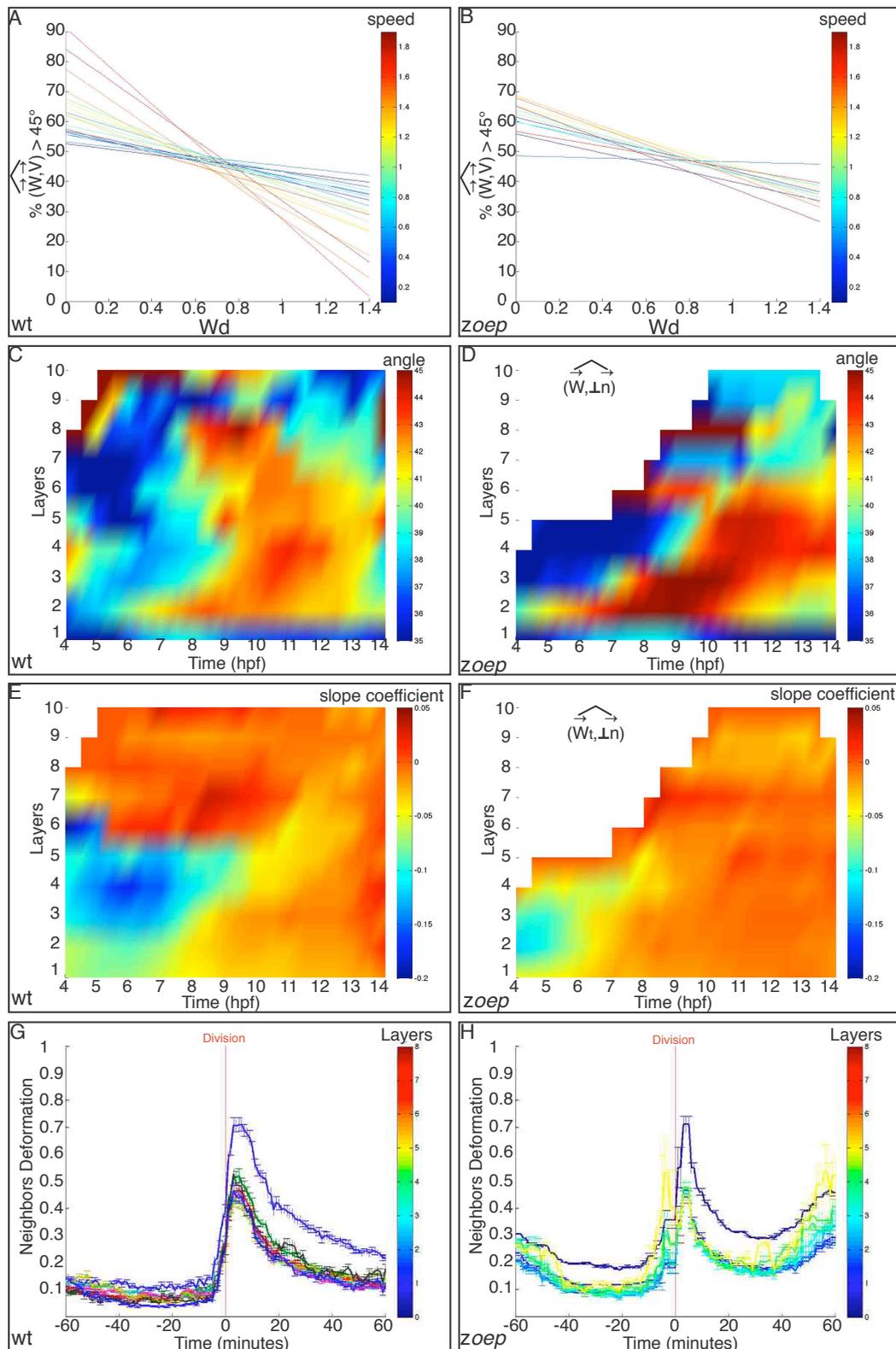


FIG. 4.4 – Cells divisions measurement, Velocity, Orientation, Deformation

4.4 Supplementary Materials

4.4.1 The different embryos

We are working here on three similar (*wildtype*) embryos, taken from 4 hours of development as well as two mutant cyclopes (*zoep*). The characteristics of each of the (*wildtype*) embryos is represented on figure 4.5 and those of the *zoep* embryos on figure 4.6.

EXPERIMENT DEFINITION Name: 070418a Status: Sent Operator: Peyrieras Nadine SRB path: /bloemerg/groups/rawdata/Im2007/Trim2/070418a/VTK Staining: RNA injection 1 cell stage H2Bmcherry eGFPras Mounting: PTFE old series 1 + agarose Comments: same scheme as 060303 but higher speed (one stack in 1 minute 7 seconds) at the expense of resolution and signal to noise ratio	EXPERIMENT DEFINITION Name: 071126aF Status: New Operator: Peyrieras Nadine SRB path: /bloemerg/groups/rawdata/Im2007/Trim4/071126aF/VTK Staining: transgenics H2Bmcherry YFPmem inj H2Bmcherry eGFPras Mounting: PTFE mold series 3 Comments:
ANIMAL Species: danio rerio Treatment: RNAinj Fishline: WT/Tu x WT/Tu	ANIMAL Species: danio rerio Treatment: trans+inj Fishline: H2B/cherry;eYFPmem x H2B/cherry;eYFPmem
MICROSCOPE Type: SP5 Excitation: 1030 fiber + dichroic Objective: 20x/0.95NA W Zoom: Temperature: room 18C Speed: 400 Hz Channels(n,g): 0, 1 Averaging:	MICROSCOPE Type: SPSF Excitation: 1030+980 Objective: 20X/1NA W Zoom: 1x Temperature: 22°C in the room Speed: -400 Hz Channels(n,g): 0, 1 Averaging: 1
IMAGING SCHEME Voxel: 1.37x1.37x1.37 μ^3 Type: AP4 Volume: 512x512x104 px ³ (761 stacks) 699.9x699.9x141.11 μ^3 Duration: 14:09:47.000 (01:07.000 between stacks) Stage 0: sphere End stage: Comments: Checking:	IMAGING SCHEME Voxel: 1.51x1.51x1.51 μ^3 Type: AP4 Volume: 512x512x234 px ³ (306 stacks) 773.12x773.12x351.83 μ^3 Duration: 14:17:59.910 (02:48.235 between stacks) Stage 0: End stage: Comments: Checking:
EXPERIMENT DEFINITION Name: 071226a Status: Sent Operator: Peyrieras Nadine SRB path: /bloemerg/groups/rawdata/Im2007/Trim4/071226a/VTK Staining: RNA inj Red nuclei green membranes Mounting: PTFE mold series 3 + agarose Comments: signal to noise ratio favoured at the expense of temporal resolution	
ANIMAL Species: danio rerio Treatment: RNAinj Fishline: WT/Tu x WT/Tu	
MICROSCOPE Type: SP5 Excitation: 1030+980 Objective: 20x/0.95NA W Zoom: 1x Temperature: 21° in the room Speed: 200 Hz Channels(n,g): 0, 1 Averaging:	
IMAGING SCHEME Voxel: 1.37x1.37x1.37 μ^3 Type: AP4 Volume: 512x512x120 px ³ (231 stacks) 699.9x699.9x163.03 μ^3 Duration: 09:52:39.909 (02:33.939 between stacks) Stage 0: between high and sphere End stage: 10-12 somites Comments: no damage detected however constrained at the end of imaging Checking: eye inspection after imaging: no visible damage	

FIG. 4.5 – 3 different embryos 070418a, 071126aF, 071226a

Tracking of cells and detection of mitoses was performed through the methodology described in Melani *et al.* (2007).

4.4.2 Setting up a spatial and temporal frame of embryogenesis

We have defined a spatial and temporal frame for each embryo so as to be able to compare one to another.

Chapitre 4. Cell division orientation in zebrafish brain early embryogenesis : a systematic analysis based on *in toto* imaging and automated cell tracking

EXPERIMENT DEFINITION		EXPERIMENT DEFINITION	
Name:	081018a	Name:	081025a
Status:	Sent	Status:	Sent
Operator:	Peyrieras Nadine	Operator:	Peyrieras Nadine
SRB path:	/bioemerg/groups/rawdata/Im2008/Trim4/081018a/VTK	SRB path:	/bioemerg/groups/rawdata/Im2008/Trim4/081025a/VTK
Staining:	RNA inj mcherry H2B eGFP-F 200 pg	Staining:	RNA inj mcherry H2B eGFP-F 200 pg
Mounting:	PTFE mold 3rd generation + 0.5 % LMP	Mounting:	PTFE mold 3rd generation + 0.5 % LMP
Comments:	female oeptz57-/+ cross male oeptz57 -/- -/- mutant sibling -/+ 081018aF	Comments:	female oeptz57-/+ cross male oeptz57 -/- -/- mutant sibling -/+ 081025aF
ANIMAL		ANIMAL	
Species:	danio rerio	Species:	danio rerio
Treatment:	oeptz57-/-	Treatment:	oeptz57-/-
Fishline:	oeptz57-/+ x oeptz57-/-	Fishline:	oeptz57-/+ x oeptz57-/-
MICROSCOPE		MICROSCOPE	
Type:	SP5	Type:	SP5
Excitation:	1030+980	Excitation:	1030+980
Objective:	20X/1NA W	Objective:	20X/1NA W
Zoom:	1x	Zoom:	1x
Temperature:	22°C in the room	Temperature:	22°C in the room
Speed:	300 Hz	Speed:	300 Hz
Channels(n,g):	0, 1	Channels(n,g):	0, 1
Averaging:	1	Averaging:	1
IMAGING SCHEME		IMAGING SCHEME	
Voxel:	1.31x1.31x1.31 μ^3	Type:	AP4
Volume:	512x512x171 px ³ (433 stacks)	Volume:	512x512x179 px ³ (406 stacks)
Duration:	18:10:08.114 (02:31.058 between stacks)	Duration:	17:49:58.750 (02:38.125 between stacks)
Stage 0:	about high	Stage 0:	about high
End stage:	20 somites	End stage:	20-22 somites
Comments:	image processing should stop at t 414 somewhat constrained by agarose mounting recovered	Comments:	the tail was constrained and the somites did not converge completely recovered
Checking:	no visible cell damage	Checking:	no visible cell damage

FIG. 4.6 – 2 different mutants 081018a, 081025a

4.4.2.1 Detection of the embryo centre

Zebrafish develop over a round vitellus, which provides them with nutrients. Throughout epiboly and gastrulation, which are the main stages of development followed here, the mass of embryonic cells spreads over the vitellus from the animal pole, embracing the close-to-spheric shape of the vitellus. We have approximated the embryo with a sphere, automatically detecting its centre and its radius, using least squares to find parameters (x_0, y_0, z_0, R) which optimally fit the centres of all of the cells' nuclei at a given timestep with the equation of the sphere. This procedure is repeated for each embryo every time step and centres and radii are averaged over time. The centre of the embryo is used as the new origin of the co-ordinate frame.

4.4.2.2 Detection of embryo's axes

In order to place the embryos in a general frame, we have chosen fixed axes which we calculate for each individual. Of the main axes goes through animal and the vegetal pole (figure 1.12). We have chosen that axis to represent depth. The bilateral symmetry plane visible with the arrival of the hypoblast represents the second axis. The third is chosen to build an orthogonal frame with the two latter (figure 4.1).

4.4.2.3 Spherical cap

In order to be able to compare the embryos in an equivalent way, we have to work on the exact same *portion* of the embryo. We have therefore taken a spherical cap corresponding to a $150\mu m$ cut from the animal pole (figures 4.1 E,H).

4.4.2.4 Temporal frame

Each embryo was put under the microscope approximately at the same stage of development, but in order to be able to compare the embryos, perfect synchrony of evolution is needed. Given external and internal factors, peculiar respectively to the experiment and to the individual, the notion of continuous time scale is variable. We have chosen the time 0 corresponding to the arrival of the hypoblast in the image (figure 4.1 H).

4.4.2.5 Approximation of cell diameter

For each cell at each time step, neighborhood relationships are based on Delaunay triangulation in 3 dimensions (?). The distance between two cell nuclei is interpreted as the sum of each cell's radius. This measure allows us to get the approximate diameter of the cells.

Figure 4.7 represents the average size of cells over time for all embryos. Maximum and minimum are represented over time by black strokes, average in bold black and variance in the color associated with the embryo.

4.4.2.6 Cellular layers

During gastrulation, a layer of cells rushes back upwards into the image, packed between the vitellus and the still descending outer layer, which is taken aback by the former. In general, the layer-based approach seems the most relevant to address the data. Layers are defined as superposed single-cell caps centred in the embryo's centre. Cell size decreasing over time, layers become ever thinner, so layer size is recalculated at each time step.

Video [Layers for 070418a](#) illustrates for an experiment the different layers in the course of time for a 10 microns cut, as on figure 4.1.

4.4.2.7 Local frame of the cell

In order to be able to compare cellular behaviour, some characteristics distinctive of each cell must be defined. For each cell we calculate the plane tangent to the sphere S centred in the embryo's centre, so as to observe the cell orientation relative to the layer. We use the vector \vec{n} normal to the tangent plane as the axis.

We then define the vector corresponding to the average movement (\vec{V}_i) of the cell over time. Due to the stochasticity of cell movement (which arises from variance in centre detection and real biological stochasticity), it is important to have the best approximation of the actual movement of the cell. Figure 4.8 schematically represents the average trajectory of a cell.

We must choose a time average which eliminates the stochasticity of cell movement without losing trajectory changes of the cells. We define for this, according to

Chapitre 4. Cell division orientation in zebrafish brain early embryogenesis : a
124 systematic analysis based on *in toto* imaging and automated cell tracking

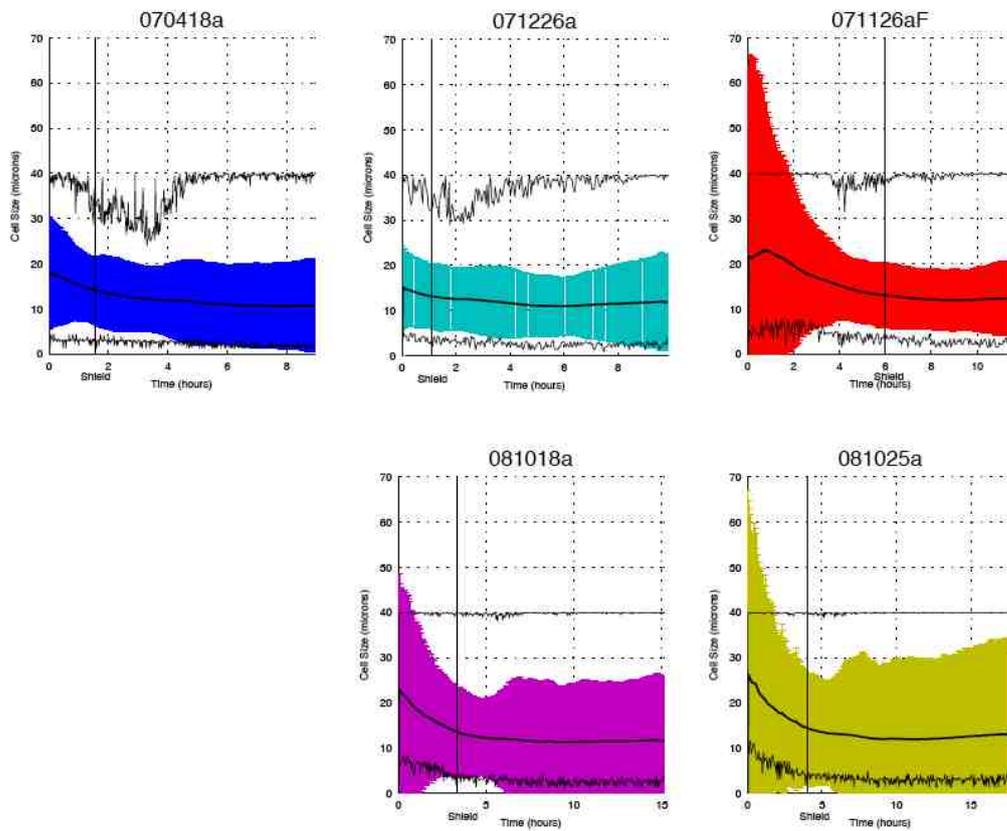


FIG. 4.7 – Evolution of cell size during development of *wildtype* embryos above and *zoep* embryos below.

the terminology of hermeneutic states (cf tracking-hermeneutic) that the best temporal trajectory is such that the angle α between past and future movement of the cell tends to π .

Figure 4.9 represents the trajectory of a cell along 20 time steps. We therefore define $\vec{V}_t = \vec{V}_{t/2}^+ + \vec{V}_{t/2}^-$ where $\vec{V}_{t/2}^+$ and $\vec{V}_{t/2}^-$ respectively represent the vector of past and of future movement of the cell along $t/2$ time steps.

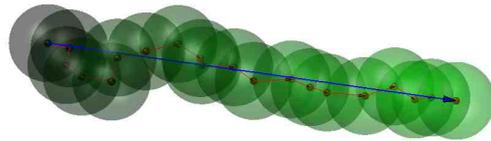


FIG. 4.8 – Schematic representation of the trajectory of a cell. The green sphere represents the cell membrane over time. Its movements between each time step are represented by a red directional arrow, its average displacement by a blue segment.

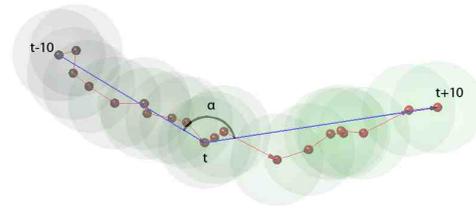


FIG. 4.9 – Schematic representation of the past and future trajectory of a cell.

Video [Orientation Hermeneutique Mouvement](#) shows, for all experiments, a spherical histogram (cf 4.4.3) of the future vector relative to the past vector, depending on the temporal average value of the movement.

All cells underwent a rotation to be centred and directed on the same axis. The red arrow represents the past movement $\vec{V}_{t/2}^-$ and the spherical histogram shows in 3D the direction of future movement. For a relatively low value of t , the distribution of vectors is spread over the sphere, which indicates average movement is quite stochastic. For a very large value of t , the future movement is unhinged relatively to past motion. For $t = 20$ minutes, we get a minimal variance without losing track changes. In figures we calculated (fig-chap2) A-Figure2 B the speed of cells obtained from this average.

La vidéo [Orientation Mouvement](#) représente pour toutes les expériences un histogramme sphérique (cf 4.4.3) du vecteur futur par rapport au vecteur passé en fonction de la valeur de moyennage temporelle du mouvement. L'ensemble des cellules a subi une rotation pour être centré et dirigé sur le même axe. La flèche rouge représente le mouvement passé, $\vec{V}_{t/2}^-$ et l'histogramme sphérique représente en 3D la direction du mouvement futur. Pour une moyenne tt relativement faible, la distribution des vecteurs est étalée sur la sphère donc le mouvement moyen est assez stochastique. Pour une valeur très grande de t , le mouvement futur se désaxe par rapport au mouvement passé. Pour $t = 20$ minutes, on obtient une variance de l'étalement minimale sans perdre les changements de trajectoires. Nous calculons dans les figures 4.2A,B la vitesse des cellules obtenues à partir de cette moyenne.

4.4.3 Measuring angles

The angle distribution between two unit vectors in three dimensions, one fixed, the other random, is not equiprobable, but follows a sinus law, which tends to 0 for colinearity. It is possible a priori to weight the measured angle distribution by the predicted distribution, but, working on the data, it creates artifacts when probabilities are low.

To solve this problem, we have used a spherical histogram. After discretizing a sphere's surface as a triangular mesh, one of the vectors was centred on the main axis and the other projected on the surface units of the spherical histogram.

Figure 4.10 is an example of a spherical histogram. The red vector corresponds to the main axis, the color (from blue to red) represents the projection of all the vectors and the bold black arrow represents the maximum value of the histogram. Figure 4.11 represents the same histogram but distorted. The distance of the unit area to the centre of the sphere is weighted by its number of elements. Figure 4.12 represents a distorted and interpolated spherical histogram in 3D color space. These last two representations are often useful to have a better visualization of observed phenomena.

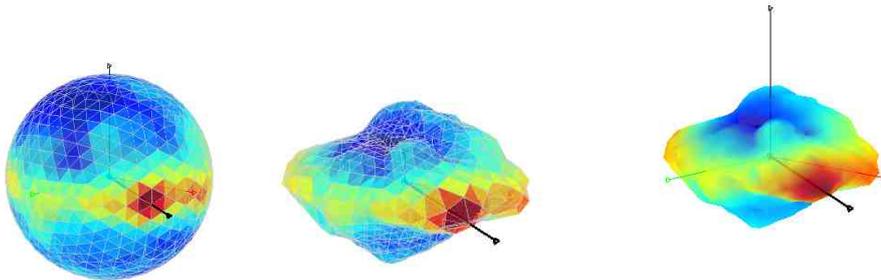


FIG. 4.10 – Spherical histogram

FIG. 4.11 – Distorted spherical histogram

FIG. 4.12 – Distorted and interpolated spherical histogram

Starting from this spherical histogram, measures are performed as on a classical 2D histogram. The histogram must sometimes be averaged. Here it suffices to weight every value associated with a triangle of the mesh by the value of neighbors triangles. A maximum may be extracted, and a distance measured between that maximum and an expected vector. Deviations from this maximum may be measured.

4.4.4 Study of cell movement relative to the layer

We use in this study (figures 4.2C,D) the temporal average value calculated in 4.4.2.7. All cells are re-centred at the centre of the spherical histogram, the plane formed by the XY axes being the plane tangent to the cell. A point on the equator

of the sphere represents a movement on the tangent plane and thus happening in the same layer. A point at the pole represents a movement perpendicular to the tangent plane and thus in the direction opposite to the layer.

Video [Vector Mother Orientation Deforme](#) represents the spherical histogram distorted over time for each of the experiments. Here time is readjusted so that $t=0$ corresponds, for each experiment, to arrival of shield.

4.4.5 Study of neighborhood deformation

The study of cell neighborhood deformation is based on the calculation of the deformation matrix studied in 3.4.4. In figures 4.1 E,F and 4.4 G,H,

4.4.6 Cells undergoing a division

Let c be a cell just before division and $c_{1,t}$ and $c_{2,t}$ the two daughters of c , t minutes after division. We define the vector between $c_{1,t}$ and $c_{2,t}$ at time t by \vec{W}_t . (figure 4.3B) and $W_t = \|\vec{W}_t\|$. We also use $W = \|\vec{W}_1\|$.

Figures 4.13,4.14, show the evolution of the number of mitoses over time for *wildtype* and *zoep* embryos. Time 0 correspond to the arrival of the hypoblast into the image. Overall, we have studied 134 110 cellular divisions.

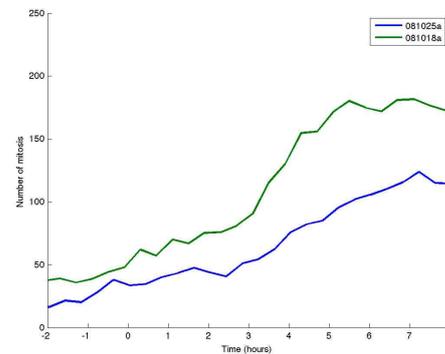
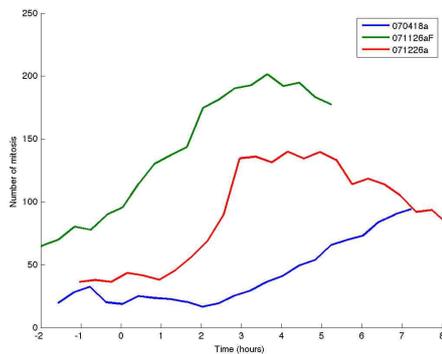


FIG. 4.13 – Evolution of the number of mitoses over time in *wildtype* embryos.

FIG. 4.14 – Evolution of the number of mitoses over time in *zoep* embryos.

For the study of the angle between the dipole and the tangent plane (Figures ?? C, D, E, F), each mother cell of a division is centred on the unit sphere, and the plane formed by the XY axes represents the plane tangent to the cell. The division vector created by the two daughters is then projected on the spherical histogram. We can thus measure the angles between the division vector and the cell layer represented by the XY plane.

In figures 4.4C,D, we measure that angle for W obtained just after division. In figures 4.4EF, we measure the reorientation of this angle during the first 20 minutes. The slope of the line shown is calculated by a linear regression on the data.

For the study of the angle between the division vector and the displacement vector of the cell (figures 4.3A,B), we take the average speed vector as the main axis of the spherical histogram.

4.4.7 Model of cellular division

We consider a cell just before division and its two daughters after division. We have conservation of momentum and energy. During mitosis, chemical energy accumulates in the cell that converts into two mechanical energies E_1 and E_2 , that propel the daughter cells one away from the other into the intracellular environment.

- Total mass : $m = m_1 + m_2$
- Reduced mass $\frac{1}{M} = \frac{1}{m_1} + \frac{1}{m_2}$
- Initial cell speed : \vec{V}
- Ejection speed of each daughter cell : \vec{v}_1 et \vec{v}_2
- We set $\vec{W} = \vec{v}_2 - \vec{v}_1$

We have the following equation of conservation :

$$(m_1 + m_2)\vec{V} = m_1\vec{v}_1 + m_2\vec{v}_2 \quad (4.1)$$

We now change variables, measuring the increase in kinetic energy.

$$\begin{cases} E_1 = \Delta E_{c_1} = \frac{1}{2}m_1\vec{v}_1^2 - \frac{1}{2}m_1\vec{V}^2 \\ E_2 = \Delta E_{c_2} = \frac{1}{2}m_1\vec{v}_2^2 - \frac{1}{2}m_1\vec{V}^2 \end{cases} \quad (4.2)$$

By equation 4.1, we have

$$\frac{\vec{V} - \vec{v}_1}{1/m_1} = \frac{\vec{v}_2 - \vec{V}}{1/m_2} = \frac{\vec{v}_2 - \vec{v}_1}{1/m_1 + 1/m_2} = M\vec{W} \quad (4.3)$$

and also

$$E_1 = \frac{1}{2}m_1(\vec{v}_1 - \vec{V})(\vec{v}_1 + \vec{V}) = -\frac{1}{2}M\vec{W}(\vec{v}_1 + \vec{V}) \quad (4.4)$$

and

$$E_2 = \frac{1}{2}M\vec{W}(\vec{v}_2 + \vec{V}). \quad (4.5)$$

So

$$\frac{2}{M}(E_1 + E_2) = W^2 \quad (4.6)$$

and

$$\frac{E_2}{m_1} - \frac{E_1}{m_2} = \vec{W} \cdot \vec{V}. \quad (4.7)$$

If we suppose that the daughter cells have same mass $m_1 \simeq m_2$ and energy $E_1 \simeq E_2$ then

$$\vec{W} \cdot \vec{V} \simeq 0 \quad (4.8)$$

that is that the movement of the mother cell is perpendicular to the daughter dipole.

If we suppose a slight asymmetry in the divisions, that is $m_1 = \frac{m}{2} - \Delta m$, $m_2 = \frac{m}{2} + \Delta m$ and $E_1 = \frac{E}{2} - \Delta E$, $E_2 = \frac{E}{2} + \Delta E$ then :

$$\vec{W} \cdot \vec{V} = \frac{4\Delta(Em)}{m^2} \quad (4.9)$$

which means a slight deviation from orthogonality mentioned above.

4.4.8 Mitosis rates

Let us introduce the average radius of a cell at time t , where $\bar{r}(t) = \frac{1}{n(t)} \sum_{\text{cells } i} r_i(t)$ is the radius of cell i , understood as a sphere, and $n(t)$ the number of cells of the embryo at time t . From time t to $t + \Delta t$, some cells have divided, changing their contribution to the average radius from $r_i(t)$ to $2\frac{r_i(t)}{\sqrt[3]{2}}$, if we assume volume distributes symmetrically between daughter cells, which need be true only on average, and, more importantly, cells do not regrow after division or change volume significantly over time. Then :

$$\sum_{\text{cells } i} r_i(t + \Delta t) = \frac{2}{\sqrt[3]{2}} \sum_{\text{dividing}} r_i(t) + \sum_{\text{non div.}} r_i(t) \quad (4.10)$$

$$= \frac{2}{\sqrt[3]{2}} \sum_{\text{all}} r_i(t) + \left(1 - \frac{2}{\sqrt[3]{2}}\right) \sum_{\text{non div.}} r_i(t) \quad (4.11)$$

Since non-dividing cells make up for almost all the cells, we may write

$$\sum_{\text{non div.}} r_i(t) \approx n_{\text{non div.}}(t) \bar{r}(t) \quad (4.12)$$

$$= (n(t) - n_{\text{div.}}(t)) \bar{r}(t) \quad (4.13)$$

$$= n(t)(1 - m(t)) \bar{r}(t) \quad (4.14)$$

where $m(t) = \frac{n_{\text{div.}}(t)}{n(t)}$ is the mitotic rate at time t .

Hence

$$\sum_{\text{cells } i} r_i(t + \Delta t) = n(t) \bar{r}(t) \left(\frac{2}{\sqrt[3]{2}} + \left(1 - \frac{2}{\sqrt[3]{2}}\right) (1 - m(t)) \right) \quad (4.15)$$

$$= n(t) \bar{r}(t) \left(1 + \left(\frac{2}{\sqrt[3]{2}} - 1 \right) m(t) \right) \quad (4.16)$$

Since $n(t + \Delta t) = n(t)(1 + m(t))$, we get $\bar{r}(t + \Delta t) = \bar{r}(t) \frac{1 + (\frac{2}{\sqrt[3]{2}} - 1)m(t)}{1 + m(t)}$ and conversely $m(t) = -\frac{\delta \bar{r}(t)}{2(1 - \frac{1}{\sqrt[3]{2}}) + \delta \bar{r}(t)}$, where $\delta \bar{r}(t) = \frac{\Delta \bar{r}(t)}{\bar{r}(t)}$ is the decay rate of the average radius.

Furthermore, we considered that the average cell radius was related to the average distance $\bar{d}(t)$ between cells by a compacity factor we considered constant. Then $\delta \bar{r}(t) = \delta \bar{d}(t)$. Finally, the data has shown that even for a timestep $\Delta t = 5$ minutes, $\delta \bar{d}(t) \ll 0.02$ always, which leads to the approximation $m(t) \approx -2.42 \delta \bar{d}(t)$.

Conclusion, Perspectives

La compréhension des dynamiques sous-jacentes à l'embryogenèse animale est un des principaux challenges de l'ère post-génomique. Intégrer l'ensemble des informations par une approche multi-échelles de l'embryogenèse animale doit permettre d'aboutir à la reconstruction et la caractérisation quantitative de l'organogenèse au cours du développement des organismes vivants.

Depuis les données d'imagerie obtenues par acquisition en microscopie multiphotonique, jusqu'à la modélisation des comportements cellulaires, cette thèse se situe dans un cadre d'interdisciplinarité. A travers l'approche des systèmes complexes, nous avons visé à reconstruire une réalité morcelée artificiellement par le cloisonnement des disciplines et ainsi acquis des compétences transversales.

Nous avons mis en place des protocoles de mesures sur des données 4D de développement du poisson zèbre. Grâce aux différentes mesures réalisées nous avons eu accès à des paramètres de natures géométriques et cinématiques permettant de décrire les interactions cellulaires et la formation des structures à différentes échelles spatio-temporelles.

La reconstruction phénoménologique de la morphogenèse animale à partir d'imagerie *in toto* d'embryons de poisson zèbre a nécessité la mise en place de méthodes mathématiques et d'outils informatiques adaptés et automatisés (Annexe A) Du filtrage des images à la reconstruction de l'arbre du lignage cellulaire, en passant par des outils statistiques, chacune des étapes du traitement des données d'imagerie a été imbriquée afin d'obtenir une reconstruction complète de l'embryogenèse. Nous avons mis en oeuvre une plateforme de traitement des données permettant d'obtenir la reconstruction de l'embryogenèse de façon rapide, simple, standardisée et complètement automatisée. Les biologistes peuvent maintenant accéder à la reconstruction d'un embryon tout juste imagé par l'intermédiaire d'une simple interface web.

Une étape importante est de caractériser et de mesurer les similitudes et les différences de développement entre différents embryons, qu'ils soient sauvages, mutants, ou traités par une drogue. A plus long terme cette systématisation peut contribuer à la mise en place d'une médecine personnalisée adaptée.

À l'avenir, les traitements médicaux seront fonction de certaines prédispositions génétiques et épigénétiques du patient, dont la prise en compte permettra d'optimiser les traitements individuels et les plans de prévention dans le cadre de santé publique.

Nous voulons pouvoir comparer différents embryons d'une même classe sous des conditions génétiques environnementales similaires. Mais nous voulons aussi

maintenant évaluer l'impact de nouvelles drogues contre le cancer sur différents embryons, en les comparant automatiquement à des embryons non traités. Nous visons à travers ce workflow à fournir une plateforme européenne de tests pré-cliniques de nouveaux traitements contre le cancer.

A différentes échelles, la reconstruction des différents éléments constituant l'embryon est soumise à beaucoup de variabilité. Elle peut-être due à leurs fonctions biologiques intrinsèques mais aussi aux techniques d'acquisitions et aux erreurs de post-traitement. Dans le but de pouvoir détecter et caractériser certains éléments spécifiques de l'embryon, nous avons présenté une méthodologie générale de catégorisation. A travers un système d'apprentissage actif, nous avons ainsi démontré que différents types d'apprentissage combinés, permettent d'éviter un grand nombre d'annotations manuelles fastidieuses. La détection de la division cellulaire lors du développement de l'embryon étant une des principales difficultés dans la reconstruction du lignage cellulaire, nous avons appliqué cette méthodologie dans le cadre de la reconnaissance automatique des phases du cycle cellulaire.

Nous voulons maintenant appliquer ce système d'apprentissage actif à la détection d'autres caractéristiques spécifiques à l'embryogenèse afin de pouvoir les catégoriser à différentes échelles.

A partir d'une cellule unique, la genèse de la forme correspond à un processus complexe d'émergence que nous voudrions appréhender à tous les niveaux d'organisation et d'observation. La notion de structures émergentes des niveaux inférieurs étant partie intégrante de la morphogenèse,

La mise en place de telles solutions de façon plus systématique devrait permettre d'identifier des structures émergentes et cela à des niveaux multi-échelles et multi-temporels.

Malgré l'importance du phénomène de division cellulaire dans l'organogenèse, les mécanismes qui contrôlent son orientation *in vivo* sont cependant peu connus. Nous avons proposé un protocole standard et automatisé de mesure des caractéristiques de la division cellulaire au cours de la morphogenèse. Nous avons mesuré des caractéristiques en termes de vitesse, d'orientation et de persistance du voisinage. Nous avons trouvé dans ce contexte des relations entre le plan de division et la couche cellulaire de la mère, ainsi que des orientations préférentielles par rapport à son mouvement. Nous n'avons pu les mesurer que sur une partie spécifique de l'embryon entre le début de la gastrulation et jusqu'à 75% de l'épibolie, mais il serait intéressant de pouvoir vérifier ces caractéristiques tout au long de l'embryogenèse. Nous avons appliqué nos mesures sur trois embryons sauvages et deux mutants *zoep*. Grâce à la plateforme européenne, nous sommes maintenant à même de pouvoir mesurer ces caractéristiques sur d'autres types de mutants et sur des embryons traités par des drogues spécifiques.

L'une des approches les plus intéressantes de compréhension du développement des organismes multicellulaires est la reconstruction de l'histoire d'une cel-

lule et de sa descendance. A cette échelle une description du paysage épigénétique peut être réalisée par le biais d'un arbre de lignage cellulaire. Jusqu'à présent, la construction de l'arbre de lignage cellulaire était un travail fastidieux réalisé à la main qui ne pouvait être réalisé de manière exhaustive que pour des organismes simples.

Nous avons proposé une méthode de reconstruction du lignage cellulaire d'un embryon, à travers le suivi spatio-temporel des cellules. Nous avons mis au point une stratégie basée sur la reconstruction de l'arbre de façon probabiliste. Nous avons proposé de compléter les informations manquantes en utilisant la connaissance de l'histoire passée et future des cellules afin de maximiser la vraisemblance des informations en faisant des itérations sur la connaissance préalable. Ces stratégies permettent d'améliorer de façon significative le suivi des cellules dans l'espace et dans le temps et de reconstruire ainsi l'histoire clonale des cellules. Ce que nous avons gagné en précision est essentiel à l'analyse de la morphogenèse et aussi à la validité et à la robustesse des protocoles que nous mettons au point pour le test des drogues anti-cancéreuses.

Ces techniques peuvent devenir des bases systématiques permettant de caractériser différents types d'éléments biologiques, physiques, sociaux, etc. Nous avons par exemple appliqué ce type de stratégie pour étudier l'influence des blogs sur la presse pendant la campagne électorale Française en 2006 (Cointet *et al.* (2007) et qui a fait l'objet d'une communication dans le *Nouvel Observateur*).

La représentation en deux dimensions du lignage cellulaire comme arbre de branchement temporel (Figure 1.25) devient trop limité lorsque les organismes atteignent des milliers de cellules. De plus l'information spatiale, pourtant capitale dans les processus de structuration, n'y est pas représentée. Il devient donc nécessaire de créer un nouveau formalisme permettant une représentation adaptée au lignage spatio-temporel (Figure 5.2).

Nous avons, à l'échelle de l'embryon, accédé aux mouvements et aux déformations individuelles des cellules mais aussi aux événements morphogénétiques. Cela nous a permis d'envisager une description multi-échelle systématique des événements morphogénétiques : *l'embryome* (Annexe B).

Ce type d'analyse quantitative de l'embryogenèse permet de s'interroger sur les liens entre interactions cellulaires au niveau individuel et la formation de structures pluricellulaires (émergence et immergence) ainsi qu'à une échelle supérieure sur la description et la comparaison systématique des embryomes d'individus d'une même espèce.

Ce type d'approche permet d'expliquer l'embryogenèse des vertébrés en terme de mouvements collectifs de populations cohérentes de cellules qui correspondraient au concept biologique, encore méconnu, de champ morphogénétique. On peut alors se demander quels sont les rôles des environnements internes et externes impliqués dans les processus de structuration ?

Nous avons commencé à caractériser la forme des champs morphogénétiques

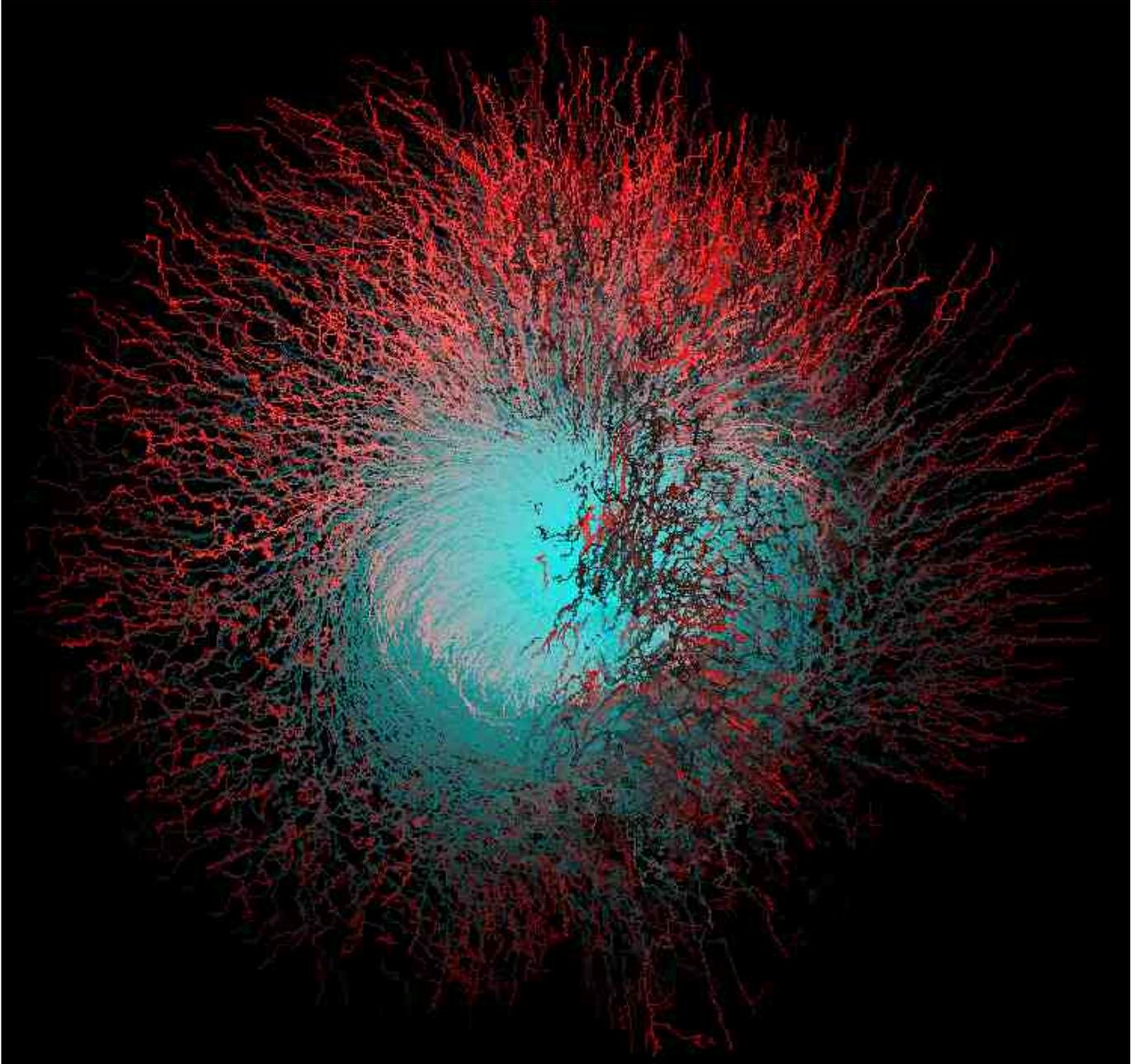


FIG. 5.1 – Représentation en deux dimensions du lignage spatio-temporel du poisson zèbre. Nous avons multiplié les trajectoires par un coefficient correspondant à l'âge de la cellule, représentée par la couleur, du bleu au rouge.

à l'aide de la méthode des *Alpha Shape* (Edelsbrunner *et al.* (1983)). Cela permet de décrire les différentes structures annotées par les biologistes par une approximation de leur forme et de pouvoir ainsi mesurer leur évolution au cours du développement.

Finalement, à partir des données observées et mesurées tout au long de cette thèse, nous avons commencé à implémenter un modèle du développement du poisson zèbre depuis la première cellule. La construction d'un tel modèle, basé sur la reconstruction multi-échelles de l'embryogenèse représentée dans la figure 1.21, permettrait de mesurer et comparer certains phénomènes modélisés avec les données réelles. Ainsi, en arrivant à reproduire virtuellement certains phénomènes spécifiques, nous pourrions faire varier quelques paramètres du modèle et mesurer les conséquences sur le développement de l'embryon. Ces paramètres pourraient être à terme identifiés et comparés à des mutations génétiques effectuées sur des embryons réels.

Ce modèle (présenté en Annexe C) nous donne un embryon virtuel représenté sur la figure 5.2 et cette vidéo [Virtual Embryo](#) représente son développement dans le temps.

Nous n'avons fait pour l'instant aucune mesure concrète mais nous observons d'ores et déjà des phénomènes émergents raisonnablement comparable à ceux observés dans le développement réel :

- des vagues de divisions (synchrones) pendant les premières heures du développement,
- un phénomène de recouvrement du vitellus selon un processus d'épibolie,
- la formation d'une couche de cellules ayant des comportements similaires à l'hypoblast.
- qui par conséquence de sa vitesse donne deux vortex comparables à ceux observés dans l'embryon réel lors de la formation des yeux,
- l'influence du choix de l'orientation des divisions sur le développement global de l'embryon,
- une déformation du voisinage pendant le processus de division,
- des comportements atypiques de certaines cellules.

Cette reconstruction automatisée du lignage cellulaire spatiotemporel rend les investigations des biologistes beaucoup plus large, profonde avec un accès aux mesures des dynamiques cellulaires. Elle ouvre des perspectives de recherche beaucoup plus audacieuses non seulement du côté de la reconstruction des modèles de l'embryogenèse mais aussi du côté de leur simulation et de l'étude de leur robustesse pour mieux les prédire. Elle permet d'étudier l'effet - à toutes les échelles? - des substances naturelles ou artificielles et de tester les effets directs et collatéraux des médicaments

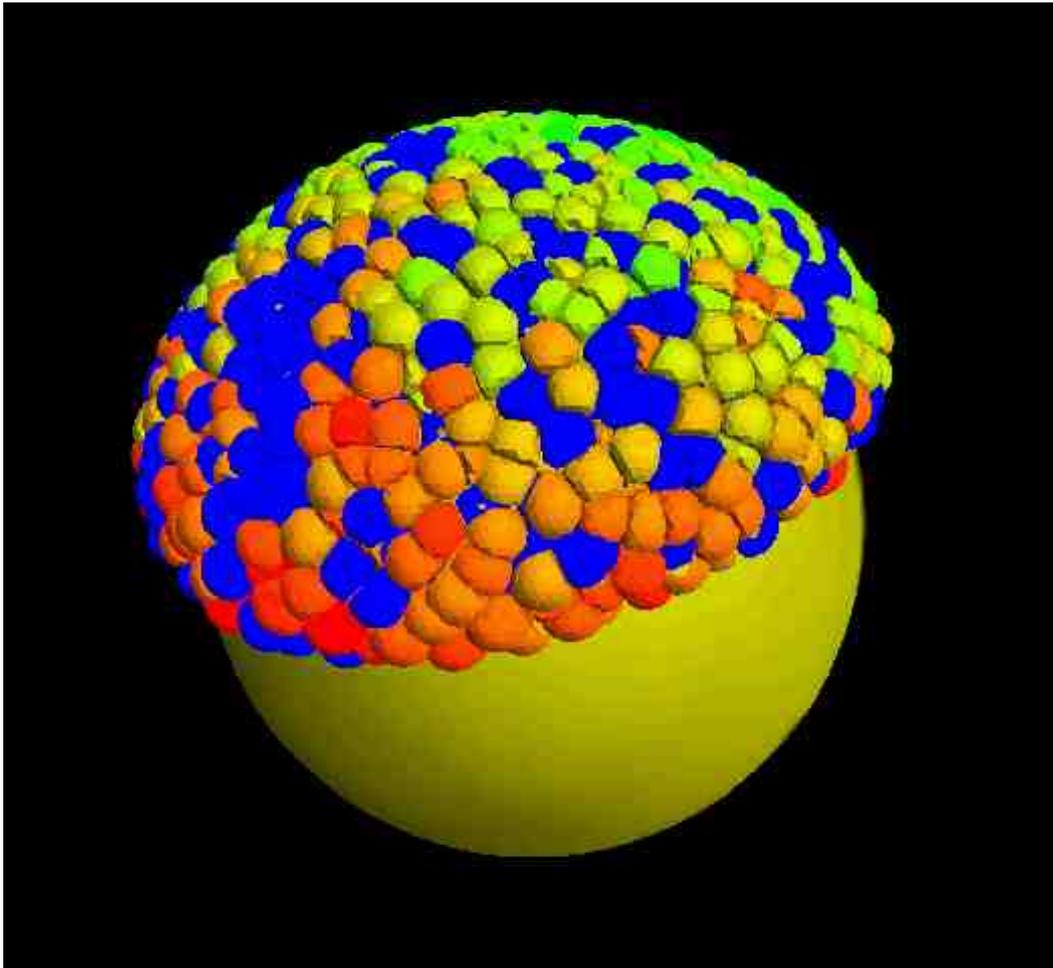


FIG. 5.2 – Embryon virtuel après quelques minutes de simulation.

Workflow Embryomics / BioEmergences

Sommaire

A.1 Un workflow pour l'embryogenèse	138
A.1.1 Comparaison d'embryons	139
A.1.2 Apprentissage des embryons et généralisation du savoir	139
A.1.3 Quantité de données	139
A.1.4 Plateforme Européenne	139
A.2 Workflow traitement des données	140
A.2.1 Filtrage des Images	140
A.2.2 Détection des centres	142
A.2.3 Mesure multi échelle des mouvements morphogénétiques	142
A.2.4 Segmentation des cellules	145
A.2.5 Détection des divisions cellulaires	148
A.2.6 Reconstruction du lignage cellulaire	149
A.3 Architecture du workflow	149
A.3.1 Implémentation	150
A.3.2 Automatisation	150
A.3.3 Utilisation	151
A.4 Ressources	151
A.4.1 Puissance de calcul	151
A.4.2 Format des données	152
A.4.3 Meta-langage	153
A.4.4 Stockage des données	154
A.4.5 Denali	155
A.5 Workflow automatique	155
A.5.1 Définition de l'expérience	155
A.5.2 Description et exécution des algorithmes	155
A.5.3 Définition des <i>pipelines</i>	158
A.5.4 Gestion des <i>jobs</i> et des erreurs	158
A.6 Conclusion	160

A.1 Un workflow pour l'embryogenèse

La reconstruction phénoménologique systématique de la morphogenèse animale à partir d'imagerie *in toto* d'embryons de poisson zèbre a nécessité la mise en place de méthodes automatisées. Nous avons développé de nouvelles stratégies qui permettent non seulement de traiter les données mais également de les interpréter. Du filtrage à la reconstruction de l'arbre de lignage cellulaire, en passant par des outils statistiques, chacune des étapes du traitement des données de l'embryon sont imbriquées pour arriver à la reconstruction totale.

Nous avons mis en oeuvre, au sein des projets Embryomics et BioEmergences, une structure permettant l'enchaînement automatique de différents algorithmes de traitement des données : le *workflow*. Par l'intermédiaire d'une interface web conviviale, les biologistes peuvent accéder à la reconstruction d'un embryon quelques instants après l'acquisition. Ainsi aux yeux des utilisateurs du workflow, la complexité technique des algorithmes de traitement devient transparente. Le workflow regroupant la collaboration de l'ensemble des partenaires permet ainsi d'accéder automatiquement à la reconstruction de l'embryogenèse (figure A.1).

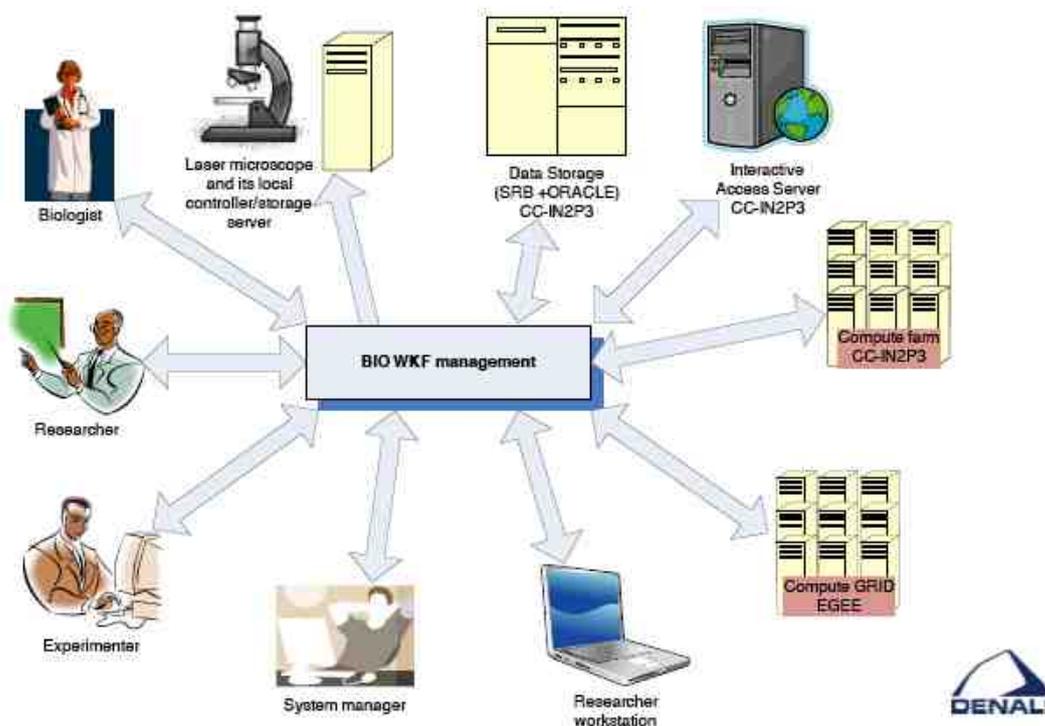


FIG. A.1 – Ensemble des interactions entre la gestion du workflow et les membres du projet

A.1.1 Comparaison d'embryons

Nous voulons dans le cadre des projets Embryomics et BioEmergences pouvoir comparer un grand nombre d'éléments, de la cellule à l'embryon en passant par l'étude des champs morphogénétiques. Des mesures descriptives et quantitatives des données permettent d'approfondir la caractérisation du développement embryonnaire de l'individu et offre la possibilité de comparer différents organismes.

Outre la difficulté méthodologique d'accéder au même élément au sens de la morphogenèse, la chaîne de traitement depuis les données brutes de la première cellule jusqu'à l'analyse statistique doit être standardisée.

Aujourd'hui, la reconstruction de l'embryogenèse est un processus qui relève d'un grand nombre d'algorithmes adaptés et optimisés en fonction des données à traiter (cf A.2). Chacun de ses algorithmes comporte un certain nombre de paramètres à ajuster afin d'obtenir le meilleur résultat possible. Afin de pouvoir identifier les éléments et les comparer de façon systématique, nous avons automatisé la chaîne de traitement des données.

A.1.2 Apprentissage des embryons et généralisation du savoir

Nous avons développé des algorithmes capables d'apprendre et de catégoriser les informations sur un ensemble d'embryons pour aider à la reconstruction des suivants. Les biologistes ont validés un grand nombre de données pour certains jeux de données. Du suivi des cellules, à la description des cartes des territoires présumés, en passant par les validations des divisions cellulaires, ces annotations sont capitales et permettent d'aider les processus de catégorisation. Afin d'automatiser l'apprentissage actif multi-échelles, nous avons développé des formalismes appropriés afin de standardiser et automatiser le traitement.

A.1.3 Quantité de données

Les images produites par les microscopes sont volumineuses, chaque volume temporel brute d'embryon se compose de plusieurs centaines de giga-octets (GO). Du filtrage, à la reconstruction de l'embryon, chacune des méthodes de traitement augmentent considérablement la quantité de données. Afin de traiter un grand nombre d'embryons, les différents algorithmes ont besoin d'importantes ressources de stockage et de calculs. De plus, le contexte européen dans lequel se situe les projets amènent chacun des membres du projet à transférer ces données vers un centre de calcul ou les rapatrier sur sa propre machine.

Nous avons rapidement dû mettre en oeuvre des solutions adaptées au stockage et au transfert d'une telle quantité d'information (cf A.4.4).

A.1.4 Plateforme Européenne

Nous visons à travers ce workflow à fournir une plateforme européenne de recherche de traitement contre le cancer (cf 1.1). Nous avons fait l'hypothèse que,

si un médicament contre le cancer est capable de faire la distinction entre les cellules cancéreuses et les cellules normales, alors il aura *in vivo* une influence sur le comportement des cellules embryonnaires. Ceci est un corollaire de la pensée actuelle que les processus de la cancérogenèse et de ceux de l'embryogenèse sont liés. Par conséquent, nous fournissons, à travers le projet *Embryomics*, un ensemble de méthodes adaptées à aux recherches de thérapies contre le cancer. Il faut signaler qu'aujourd'hui seulement trois sociétés européennes sont impliquées dans les essais pré-cliniques de médicaments contre le cancer. Elles utilisent toutes des cellules en culture de souris selon les normes actuelles. La communauté des biologistes a donc un besoin réel pour le type de technologies et de méthodologies que nous avons développé.

A.2 Workflow traitement des données

Les données en sortie des microscopes sont des volumes 3D imagés à intervalles réguliers du développement de l'embryon. Nous disposons en général de deux volumes différents un pour les noyaux et un pour les membranes des cellules. Ces données brutes autorisent une analyse des structures et des mouvements de matière à l'échelle de l'embryon. Cependant il ne paraît pas adapté de traiter les cellules individuelles directement à partir des données brutes. Nous devons donc encapsuler différents algorithmes de traitement afin de reconstruire le lignage de chacune des cellules de l'embryon.

La figure A.2 représente de façon schématique et simplifiée, les différents processus impliqués dans la reconstruction de l'embryogenèse.

Nous allons décrire dans les paragraphes suivants les principales méthodes utilisées dans le workflow. Chacun de ces algorithmes a été implémenté au sein du workflow afin de pouvoir être exécutée de façon transparente par l'interface *web*.

A.2.1 Filtrage des Images

Les images obtenues en sortie des microscopes sont bruitées. Nous avons donc, dans une première phase du workflow, développé des techniques de filtrage et de mise en valeur des structures (Perona & Malik (1990)). Nous avons mis en place, au sein des projets Embryomics et BioEmergences, plusieurs méthodes de filtrages de données en 3D et 4D (figure A.3). Compte tenu qu'il n'existe pas, aujourd'hui, de méthode universelle de pré-traitement adaptée à tous types d'images, nous avons mis en place différentes stratégies optimisées en fonction des post-traitements à effectuer.

Les projets Embryomics et BioEmergences visent à reconstruire l'embryogenèse de différents organismes vivants. De la taille à la forme des cellules, les caractéristiques cellulaires diffèrent selon le type d'animal et nous avons donc développé des stratégies adaptées. De plus certaines méthodes de filtrage sont spécifiques aux membranes des cellules alors que d'autres ont de meilleurs résultats sur

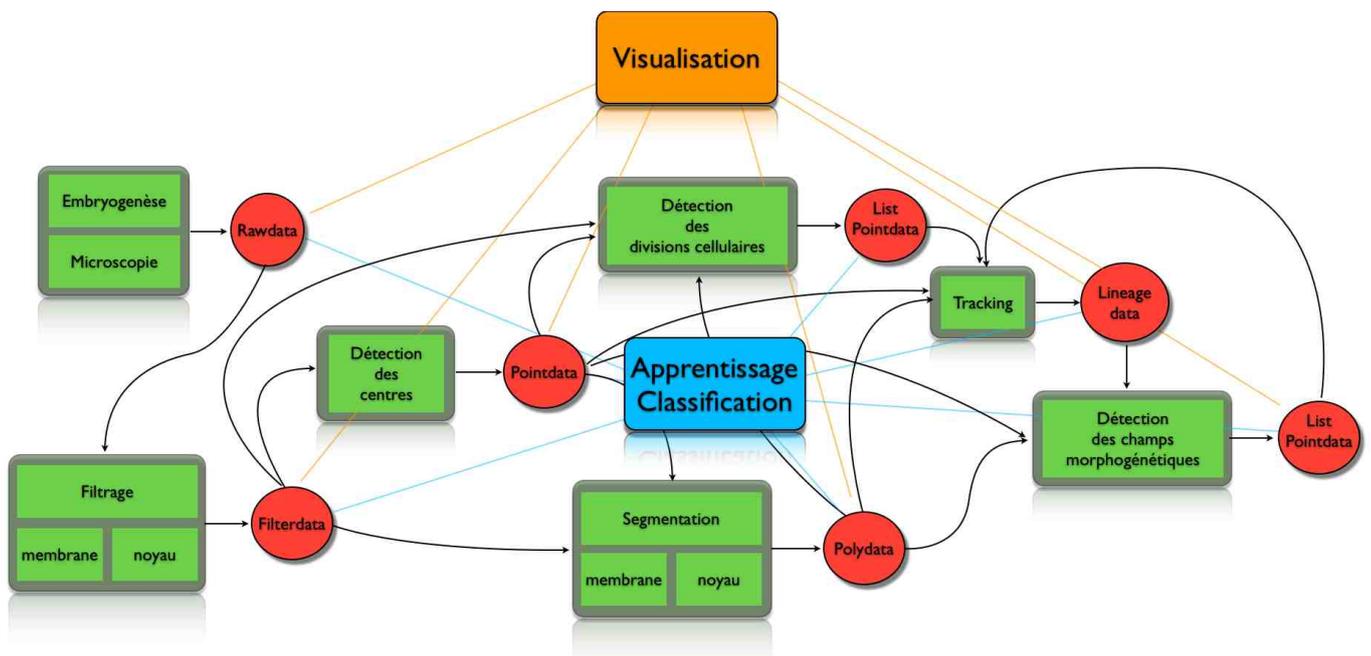


FIG. A.2 – Schéma des différents processus de traitement des données impliqués dans la reconstruction de l'embryogenèse. Les algorithmes sont représentés par des boîtes vertes, les données produites par des cercles rouges. L'interface de visualisation et les algorithmes de classification sont connectés à toutes les données produites.

les volumes de noyaux. Enfin la notion de rapidité d'exécution des algorithmes et de consommation de temps de calcul sont à prendre en compte dans la reconstruction selon le type d'informations que nous désirons étudier. Certaines méthodes de filtrage, bien qu'obtenant des résultats très performants s'avèrent très gourmandes en ressources de calcul.

Le bruit dominant dans les images produits par le microscope CLSM est sous la forme d'un bruit de Poisson plutôt que d'un bruit Gaussien (cf 1.4.3). Les techniques de filtrage non linéaires ont généralement un meilleur résultat pour le pré-traitement des images. Nous avons adapté ces techniques en utilisant des équations aux dérivées partielles (Rizzi *et al.* (2007a), Uttenweiler *et al.* (2003)), des tenseurs non linéaires de diffusion anisotrope (Drblikova (2005), Drblikova (2008)), des techniques de mathématiques appliquées à l'étude des morphologie (Luengo-Oroz *et al.* (2007)) et des techniques de seuils par analyse d'ondelettes (Boutet *et al.* (2001)).

A.2.2 Détection des centres

La technique d'acquisition consiste à prendre des volumes 3D d'images de l'embryon et de réitérer l'opération quelques instants plus tard. Suite à cette méthodologie, chaque embryon est discretisé temporellement. Il faut donc pour chaque volume d'imagerie retrouver le centre des cellules.

La détection de centres des noyaux des cellules (figure A.4) est une étape cruciale dans la reconstruction de l'embryogenèse comme nous l'avons vu dans le chapitre sur le *tracking* des cellules (cf chapitre B).

Il existe une énorme littérature sur la détection de particules en biologie, y compris l'utilisation du seuil d'optimisation (Otsu (1975)), d'opérateurs morphologiques (Chuang *et al.* (2004)) de déformation élastique (Liu *et al.* (2008)), en appliquant la transformée de Hough (Duda & Hart (1972)) ou la transformée en ondelettes discrètes (Olivo-Marin (2002)). La limite de ces méthodes apparaît rapidement quand les cellules se touchent ou lorsque le ratio de l'intensité sur le bruit est faible (généralement dans la profondeur du volume imagé) (Malpica *et al.* (1997), Wahlby *et al.* (2008)). Nous utilisons, au sein du consortium, une approche basée sur les équations d'*advection-diffusion* qui fournit des résultats très précis (Frolkovic *et al.* (2007)). L'inconvénient principal est le temps de calcul nécessaire pour optimiser les paramètres.

A.2.3 Mesure multi échelle des mouvements morphogénétiques

Afin d'étudier les mouvements de matière dans l'embryon, nous avons développé des méthodes de mesures de flots optiques en 3D entre deux pas de temps consécutifs. Nous décrivons les mouvements de matière, les déformations et les déplacements de l'embryon à l'aide de champs de vecteurs (figure A.5).

Le *tracking* des cellules pourrait fournir les données permettant de réaliser de telles mesures. Cependant, les différentes erreurs accumulées lors de chaque étape

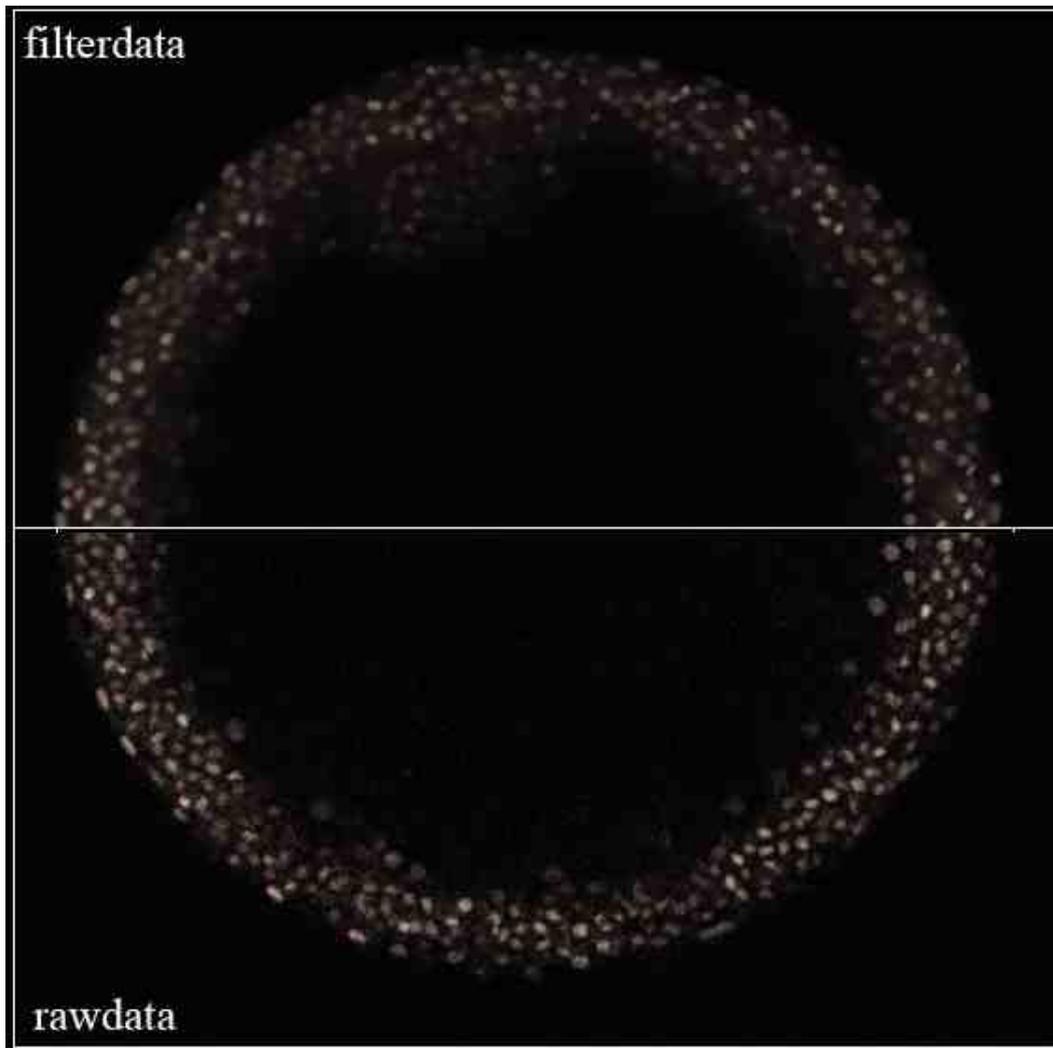


FIG. A.3 – En bas une coupe en z d’image des noyaux cellulaire à la sortie du microscope. En haut, la même coupe qui a subi un filtrage. Nous constatons que le bruit a été éliminé tout en conservant la structure des noyaux.

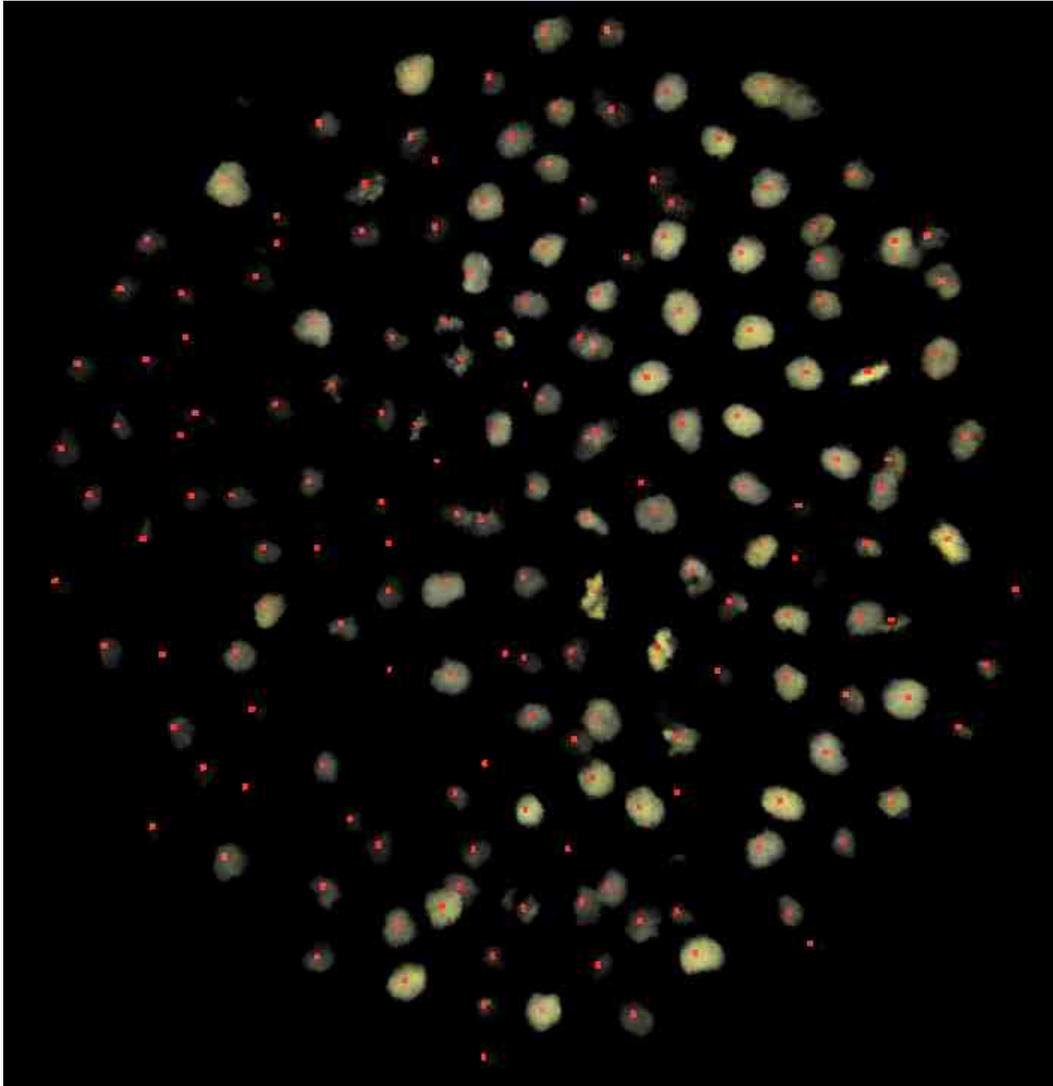


FIG. A.4 – Visualisation de la calotte d'un embryon (10 coupes en z). Les données filtrées sont affichées en projection 3D en blanc, et les centres des cellules détectées sont marqués par des carrés rouges.

de la chaîne de traitement permettant de reconstruire le lignage cellulaire, introduisent des biais dans l'étude des mouvements de matières. Ainsi ces méthodes de recalage temporel, à partir des images brutes, permettent une analyse à une échelle plus globale des mouvements morphogénétiques (Lombardot *et al.* (2008)).

Ces mesures de déplacement à l'échelle du pixel contribuent à l'amélioration du *tracking* des cellules dans le temps en donnant des directions préférentielles de mouvements.

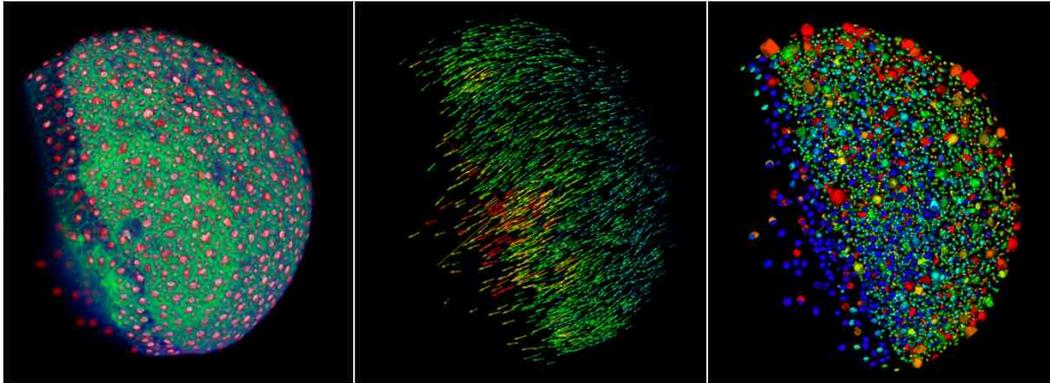


FIG. A.5 – De gauche à droite. Les images filtrées des membranes superposées à celles des noyaux des cellules. Le flot optique des déformations des pixels associées à chaque cellule. La segmentation des noyaux faite à partir des centres détectés.

A.2.4 Segmentation des cellules

Le terme de *segmentation* dans le cadre habituel du traitement d'image signifie *tracer les contours des objets*.

La segmentation des noyaux et des membranes des cellules (figure A.5 et A.6) est une condition préalable à une analyse quantitative précise de la forme et de la morphodynamique des cellules au cours de l'organogenèse afin d'obtenir une meilleure compréhension des processus biologiques.

Plusieurs méthodes de segmentation d'images ont été adaptées aux caractéristiques de la biologie cellulaire : les contours actifs (Zimmer *et al.* (2002a)), des optimisations de graphes (Meijering *et al.* (2004)), les méthodes *model-based* (Lin *et al.* (2005)) ou des méthodes utilisant des réseaux de neurones (Shitong & Min (2006)).

En fonction des mesures effectués sur les objets segmentés et des puissances de calcul disponibles, nous avons dû développer différentes méthodes de segmentation.

En initialisant l'identification des cellules par la détection des centres des noyaux, nous avons utilisé des méthodes par *watershed* (Meyer (2001)) permettant d'effectuer la segmentation en parallèle sur l'ensemble des cellules à un faible coût

de calcul (Vincent & Soille (1991)). Nous avons aussi mis en place des techniques de segmentation de membranes sur la base des voisinages créée par un algorithme de type *voronoï* (Luengo-Oroz *et al.* (2008)).

De plus, nous avons utilisé la technique des surfaces subjectives, servant à compléter les frontières manquantes (Sarti *et al.* (2000)), basée sur des équations aux dérivées partielles (Rizzi *et al.* (2007c), Rizzi *et al.* (2007b)). Ces techniques s'avèrent très gourmande en ressource de calcul mais permettent d'extraire avec précision la forme des membranes et les noyaux.

En outre, la segmentation manuelle consiste à designer manuellement les contours d'une cellule préalablement choisis. Elle a été mise en place au sein du projet pour permettre de mesurer la qualité de la segmentation automatique et obtenir un intervalle fiable sur les mesures de volume cellulaire. Des techniques de segmentation semi-manuelle ont aussi été implémentées, il s'agit de d'abord placer manuellement des marqueurs dans chaque cellule, et ensuite lancer une segmentation automatique qui utilise la position des marqueurs comme état initial.

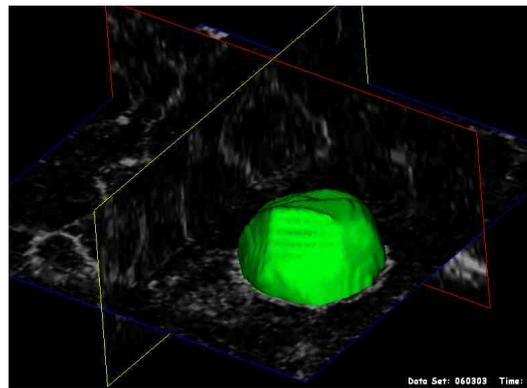


FIG. A.6 – Visualisation de la segmentation d'une membrane d'une cellule superposé aux orthoslices des images originales.

A.2.4.1 Segmentation des cellules épithéliales

Les cellules épithéliales (figure A.7) sont les premières cellules lors du développement du zebrafish à avoir un comportement morphodynamique et une forme membranaire différenciée. Afin de mieux caractériser l'embryon, il fallait pouvoir séparer ces cellules des autres cellules. On a donc utilisé des techniques spécifiques appliquées à ce champ morphogénétique. Elles se basent principalement sur leur position dans l'espace puisqu'elles sont à la périphérie de l'embryon.

A.2.4.2 Interface de visualisation

Une grande quantité d'outils de visualisation des données biomédicales est aujourd'hui disponible. *Amira*, *Paraview*, *Volview*, pour ne citer que les plus connus, chacun ayant ses avantages et ses inconvénients selon les besoins spécifiques. La

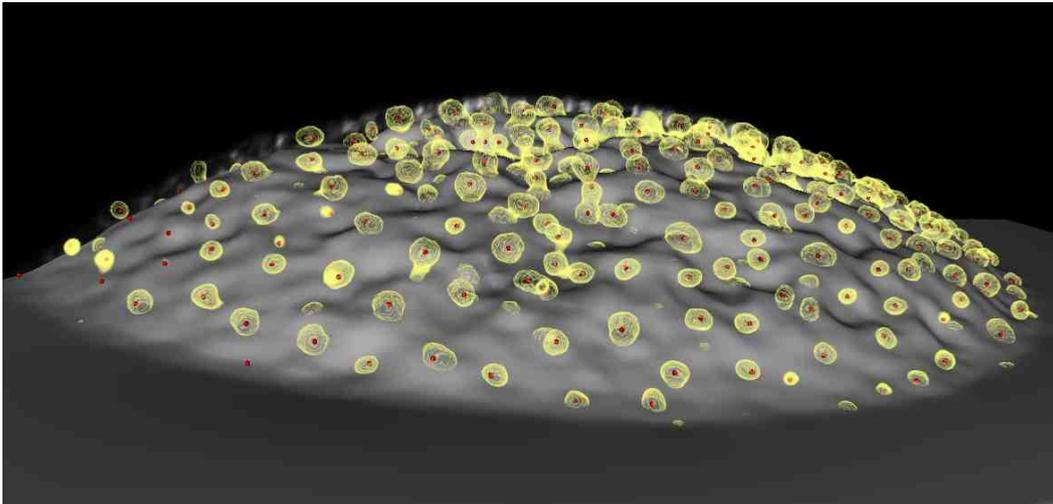


FIG. A.7 – Segmentation de cellules épithéliales représentées en jaune avec leur centre marqué par un point rouge. Le reste de l’embryon se trouve sous la calotte grise.

grande majorité des logiciels permet bien-sûr de visualiser les données brutes et parfois d’y ajouter une collection de centres et d’objets segmentés. Nous utilisons dans le cadre de notre projet, un grand nombre de différents outils de visualisation des données permettant ainsi d’obtenir des rendus spécifiques.

Dans le cas du développement de la plate-forme, nous avons développé une interface de visualisation permettant de visualiser tous les types de données produites par le workflow. De plus il a été capital de pouvoir avoir un retour des biologistes sur la reconstruction des données sous forme d’annotations. Nous avons donc rendu cette interface interactive permettant ainsi d’annoter des éléments reconstruits. Cette annotation permet d’évaluer à chaque étape la qualité de la reconstruction et ainsi de pouvoir sélectionner les techniques de reconstructions qui semblent les plus pertinentes.

Nous avons directement connecté cette interface au serveur de données ainsi qu’à la base de données, ce qui permet dans des temps extrêmement courts d’avoir accès aux données reconstruites et de les annoter (cf A.4.4). Cette interface est complètement dynamique, permettant ainsi d’accéder à n’importe quel *instant spatio-temporel* de la reconstruction. De plus, la conception de méta-langages permet d’accéder à différents niveaux de l’embryon (cf A.4.3). Ainsi cette interface permet de visualiser les objets de façon mutli-échelles. Les biologistes ont pu annoter des cellules aux comportements spécifiques, tracer les frontières des champs morphogénétiques, corriger des divisions cellulaires etc Enfin cet outil de visualisation est en perpétuelle évolution selon les besoins des différents utilisateurs.

La figure A.8 représente la visualisation par l’interface de la reconstruction du lignage cellulaire d’un embryon. La vidéo [Interface](#) donne un aperçu dynamique

du suivi de cellules avec les données brutes superposées.

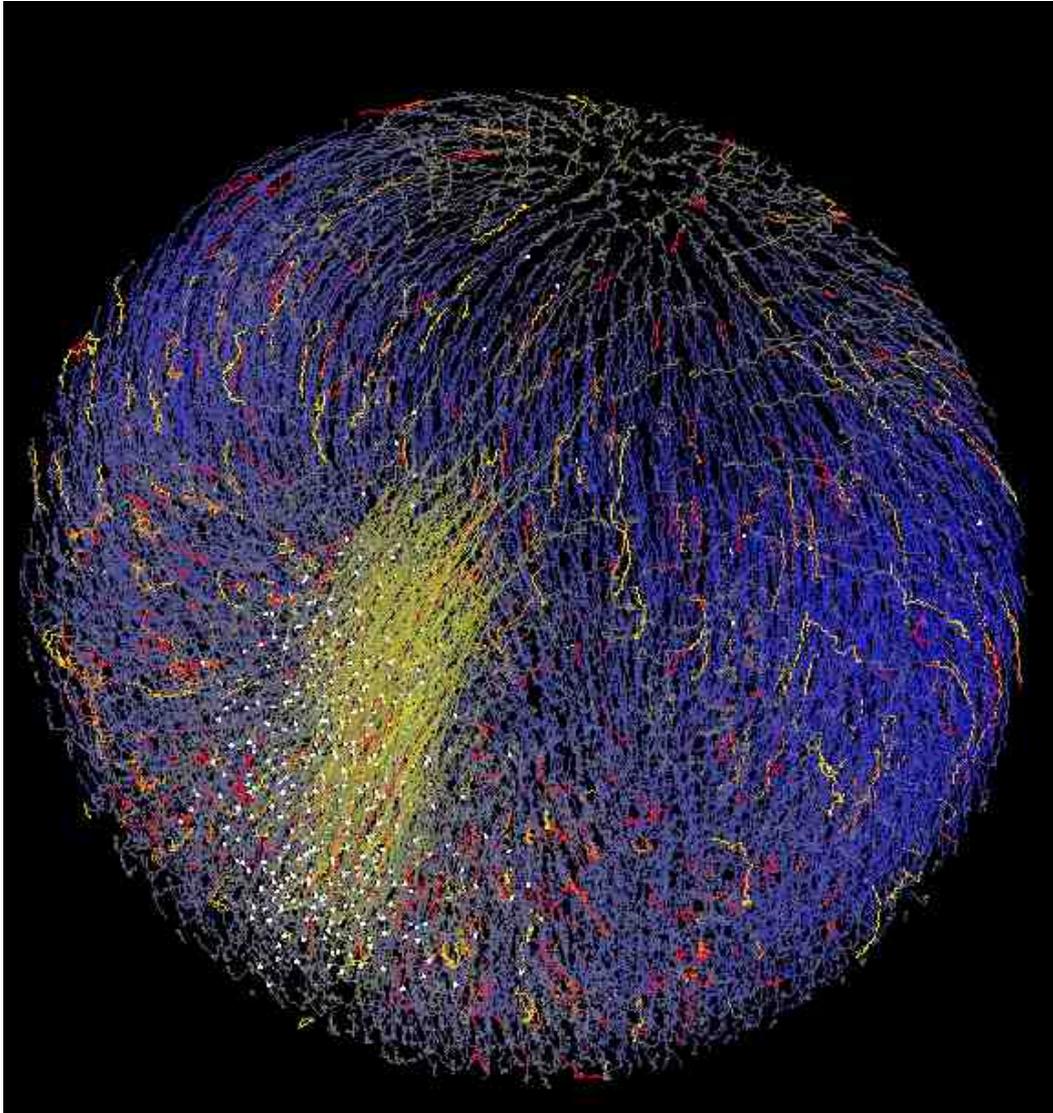


FIG. A.8 – Interface : visualisation du *tracking* de cellules. La couleur correspond à la vitesse des cellules, du plus lent en bleu au plus rapide en jaune. En rouge sont représentés les cellules en division.

A.2.5 Détection des divisions cellulaires

La détection de la division cellulaire lors du développement de l'embryon est une des principales difficultés dans la reconstruction du lignage cellulaire. L'acquisition d'imagerie par microscopie bi-photonique faite à l'échelle de l'embryon ne permet pas d'observer au niveau intra-cellulaire les différentes phases de la di-

vision cellulaire. Mais elles sont reconnaissables dès la prophase grâce à la condensation de la chromatine. Cette condensation génère une hausse de l'intensité dans l'image et une déformation de la forme du noyau. Dès lors, il est possible d'utiliser ces caractéristiques pour différencier la cellule prête à se diviser des autres. Nous avons élaboré un système d'apprentissage actif basé sur la catégorisation par les SVM et utilisant l'information de contexte temporel et spatial (cf chapitre 2). Les taux d'erreurs de cette reconnaissance automatique sont mesurés par comparaison à un standard de référence (*gold standard*) correspondant aux annotations des biologistes. Nous avons proposé, dans le chapitre 4, un protocole standard et automatisé de mesure des caractéristiques de la division cellulaire au cours de la morphogénèse, fondé sur des combinaisons d'algorithmes.

A.2.6 Reconstruction du lignage cellulaire

Après la détection du centre du noyaux des cellules par volume d'imagerie, une cellule n'existe qu'à un pas de temps donné et n'a donc aucun passé ou futur. Il faut ensuite créer les liens temporels entre les centres détectés à des temps différents pour reconstruire l'histoire des cellules.

Nous avons utilisé des stratégies de minimisation locale et globale des informations afin de reconstruire l'arbre du lignage cellulaire (Melani *et al.* (2007)). Cette reconstruction du lignage permet d'analyser certains phénomènes descriptifs qui ne nécessitent pas d'avoir un tracking complètement parfait.

Nous avons mis au point une stratégie basée sur la reconstruction de l'arbre de façon probabiliste (cf chapitre 3). Un des problèmes majeur est que parfois l'algorithme de détection du centre des noyaux des cellules n'arrive pas à en trouver certaines, ou en détecte qui n'existent pas. Nous proposons alors de compléter les informations manquantes en utilisant la connaissance de l'histoire passée et future des cellules (trajectoires, déformations, forme, etc.). Nous maximisons la vraisemblance des informations par des itérations sur la connaissance préalable. Ces stratégies permettent d'améliorer de façon significative le suivi des cellules dans l'espace et dans le temps et de reconstruire ainsi l'histoire clonale des cellules. Cette précision est ainsi essentielle à l'analyse de la morphogénèse.

A.3 Architecture du workflow

L'implémentation, l'automatisation et l'utilisation des différents processus impliqués dans la reconstruction du lignage cellulaire sont les principales étapes successives permettant la création d'un workflow. Nous avons ainsi décomposé l'architecture du workflow en trois couches :

- *Implémentation* : couche de création des algorithmes.
- *Automatisation* : couche intermédiaire de gestion automatique des algorithmes.
- *Utilisation* : couche des utilisateurs du workflow grâce à l'utilisation d'une interface *web*.

Cela permet notamment de développer et remplacer indépendamment chacun des modules internes aux couches, sans perturber le fonctionnement des autres couches.

A.3.1 Implémentation

La couche *Implémentation* représente la phase de développement des algorithmes. Cela concerne aussi bien les algorithmes de traitements d'images, de mesures des données que des systèmes d'apprentissage. Les personnes sollicitées pour la mise en place de cette couche sont principalement des mathématiciens et des informaticiens. Nous avons mis en place certains critères à respecter afin de faciliter l'intégration des algorithmes dans cette couche :

- Emplacement standardisé et nom générique des algorithmes.
- Optimisation des algorithmes afin de minimiser l'utilisation des ressources de calculs.
- Code source portable et recompilable à la volé sur des clusters.
- Code commenté afin de pouvoir être réutilisé par les personnes responsables de la couche *Automatisation*.
- Générer des classes réutilisables par d'autres algorithmes.
- Standardisation des entrées et des sorties des algorithmes afin de les emboîter dans la chaîne de traitement.
- Définition standardisée des paramètres permettant de s'adapter à tous types d'images.

De plus, nous avons favorisé le développement des algorithmes dans des langages libres, autonome et puissant. Par exemple, un algorithme développé en Matlab (langage le plus utilisé par nos différents partenaires) est en général très gourmand en ressource, nécessite l'installation d'un logiciel spécifique sur le cluster, et de plus, fonctionne par jeton payant. Cela limite ainsi son utilisation intensive. Afin que les chercheurs n'aient pas à devoir apprendre des langages de programmation plus fastidieux, nous avons développé des bibliothèques communes et des techniciens spécialistes ont réécrit certains algorithmes dans des langages puissant comme C++.

A.3.2 Automatisation

La couche d'*Automatisation* est au centre du workflow. Cette couche intermédiaire fait le lien entre les algorithmes et les utilisateurs. Elle permet d'automatiser l'exécution des algorithmes et de réparer les erreurs standards, de gérer les différentes ressources de calculs, de contrôler l'historique des tâches effectuées, et, enfin, d'archiver le stockage des données. Cette couche est détaillé dans la partie [A.5](#).

A.3.3 Utilisation

La couche *Utilisation* permet aux biologistes de pouvoir exécuter de façon simple (et conviviale) l'ensemble des algorithmes mise en place dans le workflow en masquant la complexité technique de leurs exécutions. Elle est accessible par l'intermédiaire d'une interface Web graphique (l'onglet *workflow* du site de [BioEmergences](#)). Cela permet, sans avoir les connaissances spécifiques à chacun des domaines de traitement des données, d'effectuer la reconstruction d'un embryon tout juste imagé. Nous verrons par la suite différentes captures d'écran de l'interface web du suivi des différents processus de cette couche.

A.4 Ressources

Compte tenu de la grosse quantité de données à gérer et de l'automatisation des processus nous avons mis en place plusieurs types de ressources informatiques. Du *typage* du langage à la puissance de calcul, nous avons standardisé les ressources pour qu'elles soient compatibles avec de nouvelles données et adaptables à tout type de changements.

A.4.1 Puissance de calcul

L'exécution des différents algorithmes du workflow nécessite de grandes ressources de calculs. Nous avons dans le cadre du développement du workflow intégré l'exécution des algorithmes sur plusieurs machines en même temps. Cela permet d'avoir accès en peu de temps à la reconstruction d'un embryon. Il faut par exemple en moyenne 24h pour reconstruire l'arbre de lignage cellulaire d'un embryon sur 1000 processeurs. Nous avons utilisé pour les besoins du projet deux types de centre de calcul différents : le centre de calcul du CRNS, l'*IN2P3* et la grille européenne de calcul, *EGEE*. Ces deux centres ont des principes de fonctionnement différents et ont leur propre mode d'exécution des algorithmes (soumission de *jobs*). De plus nous avons développé le *workflow* de façon à pouvoir intégrer facilement un nouveau centre de calcul ayant ses propres spécificités.

A.4.1.1 IN2P3 - Lyon

Le Centre de calcul de l'*IN2P3* (*Institut national de physique nucléaire et de physique des particules*) à Lyon est un département du CNRS qui a pour mission de promouvoir et de fédérer les activités de recherche dans les domaines de la physique nucléaire et des hautes énergies. Si ces grands thèmes représentent le cœur de l'activité, l'*IN2P3* a aussi vocation à apporter ses compétences à d'autres domaines scientifiques. Le centre de calcul permet ainsi à la communauté des biologistes d'avoir accès à ses différentes ressources. Cela nous a permis de mettre en place des solutions adaptées aussi bien au niveau du stockage que des ressources de

calculs nécessaires. Nous disposons d'une puissance de calcul d'environ 500 processeurs à l'IN2P3-Lyon, ainsi que d'une grande capacité de stockage dont nous n'atteignons pas les limites pour le moment.

A.4.1.2 EGEE

Depuis quelques années, un nouveau concept pour le calcul intensif a vu le jour sous le nom de grille de calcul (*grid computing*). Le principe est de mutualiser les moyens de calcul qu'ils soient de taille modeste (moyens locaux et régionaux) ou de taille plus importante (centres de calcul nationaux et internationaux). Pour cela il a fallu développer les logiciels spécifiques (*middleware*) permettant la mise en commun des données et des programmes de traitement. Le projet Européen EGEE (*Enabling Grids for E-science*) permet la mise en production de la grille de calcul au service de toute la communauté scientifique européenne. Ce projet a en charge le développement du *middleware*, le déploiement et le suivi opérationnel de la grille de calcul. Nous avons accès pour la mise en place de ce workflow, par l'intermédiaire du Laboratoire de l'accélérateur linéaire (LAL), aux ressources de calcul d'EGEE à hauteur de 2000 processeurs.

A.4.2 Format des données

Il existe un très grand nombre de formats de données brutes ou reconstruites pour des images de type biomédicale. La majorité des microscopes industriels génère en sortie des formats propriétaires qu'il est difficile d'utiliser. Afin de simplifier le partage des données entre les différents partenaires du projet mais aussi, à l'échelle de la plateforme européenne, de proposer des standards utilisables par toute la communauté, nous avons choisi un format libre, puissant, générique à l'ensemble des données et supporté par la majorité des logiciels de visualisation et de traitement des images 4D : *VTK*.

ITK (Insight Segmentation and Registration Toolkit) est une librairie open source, multi-plateforme qui fournit aux développeurs une vaste gamme d'outils pour l'analyse d'image. Cette librairie est très utilisée dans le domaine de traitement d'image, elle est supporté par une très grande communauté et à l'avantage d'être régulièrement actualisée. *VTK (Visualization Toolkit)* est la composante de la librairie contenant une vaste gamme d'algorithmes de visualisation des données. Ces librairies sont constituées de classe C++ et supportent plusieurs couches de différents interpréteurs. Outre l'aspect pratique de traitement et de visualisation des données, ces librairies propose des formats génériques. Nous en avons utilisé 3 principaux :

- **VTK-Rawdata** : les images brutes composées de coupe en z. Ces images peuvent être obtenu à partir des algorithmes de filtrage, le format reste identique.
- **VTK-Center** : une collection de centres qui sont représentés par des points en coordonnées cartésiennes (X, Y, Z).

- **VTK-Polydata** : structure contenant la segmentation des cellules. Elle est représentée pour chacune d'entre elle par un maillage de triangles. Nous utilisons ce format pour tout type de segmentation, les cellules, les champs-morphogénétiques et les embryons.

A.4.3 Meta-langage

Nous avons développé un formalisme des connaissances approprié à la gestion de nos données. L'avantage de l'utilisation d'un formalisme générique aboutissant à un meta-langage permet à tout moment de pouvoir ajouter des nouveaux types de données ou d'algorithmes, sans avoir à modifier les autres éléments de la structure globale. Ce meta-langage est utilisé dans toute l'architecture du *workflow* et est modifiable par l'interface web. Il permet d'encapsuler les différents types de données et de définir les paramètres d'entrée/sortie des algorithmes. Ce meta-langage permet d'avoir une grande sûreté sur l'enchaînement des différents processus. De plus la syntaxe compacte s'inspirant des langages fonctionnels reste parfaitement lisible.

Il peut s'expliquer par quelques règles simples :

- $(element1, element2)$ représente l'ensemble comprenant les deux éléments.
- $\{element\}$ représente une liste d'éléments.
- $O : (input)$ signifie que l'objet O est composé de l'élément $input$.
- $F : (input) \rightarrow output$ signifie que la fonction F prend en entrée $input$ et renvoie en sortie $output$.

La description suivante représente, sous une forme simplifiée, les composantes permettant de créer une expérience :

- **Experiment** : (animal, age, microscope, person, montage, staining, comments, name, SchemaType, temperature, date) \rightarrow (TimeLaps, idDB)
- *TimeLaps* : ({ImageStack}, t, deltaT, imgfile)
- *VTK-Rawdata* : (z, deltaZ, {ImageXY})
- *ImageStack* : (z, deltaZ, {ImageXY})
- *ImageXY* : (x, deltaX, y, deltaY)
- *SchemaType* : (name, Schema)
- *Schema* : (Coord3D, Rotation3D, Length3D)
- *Coord3D* : (x, y, z)
- *Length3D* : (L, l, h)
- *Rotation3D* : (Ox, Oy, Oz)

Nous donnons ici les spécifications nécessaires à l'exécution de différents algorithmes :

- **ConvertImage** : (TimeLaps) \rightarrow VTK-Rawdata
- **ImageProcessing** : (VTK-Rawdata, AlgoType, {valueParameters}) \rightarrow VTK-Rawdata
- *AlgoType* : (name, {DefParameters}, location)

- *DefParameters* : (name, format, Bornes)
- *Bornes* : (borneInf, borneSup)
- **ImageSegmentation** : (VTK-Rawdata, AlgoType, {parametres}) → VTK-Polydata
- *VTK-Polydata* : (numberCells, cellMaillage)

La liste suivante représente une structure d'embryon, composée de champs morphogénétiques, eux même composés de cellules.

- *Embryo* : (VTK-Rawdata, VTK-Polydata, MorphogeneticField)
- *MorphogeneticField* : (numberCells, Cell, position, shape)
- *Cell* : (id, volume, mass, surface, cellMaillage, velocity, {Neighbours})
- *Neighbours* : (numberCells, Cell)

A.4.4 Stockage des données

En fonction des besoins de chacun des processus du workflow, nous avons utilisé différents types de stockages de données. Chacun offrant des performances optimales répondant à des besoins spécifiques :

- **AFS- THRONG** (Andrew File System) : espace limité (16 GO) mais très rapide d'accès . Il offre un système de sauvegarde régulier et de gestion d'archivage des algorithmes. Cet espèce de stockage est dédié à la création du code des algorithmes.
- **AFS - GROUP** : (30 GO) Ce système à les mêmes propriétés d'accès que le précédent, mais sans système de sauvegarde. Il contient l'ensemble des bibliothèques compilées nécessaire aux développement des algorithmes (exemple : VTK / ITK).
- **SPS** (Semi-Permanent Storage) : (500 GO) ce système est utilisé pour stocker les données de façon temporaire. Lors du développement des algorithmes, les chercheurs ont un besoin permanent d'accéder à des données d'images. Ce système offre de très bon rapport de quantité de stockage par rapport au temps d'accès.
- **TMPBATCH**, représente l'espace disque utilisé par un noeud (cluster). il est alloué au début du job, et effacé à la fin. Il peut être de taille variable en fonction des besoins.
- **SRB** (Storage Resource Broker) (en centaines de TO) permet de stocker de gros volumes de données. Il contient l'ensemble de données, des images brutes en sortie des microscopes jusqu'à la reconstruction finale. SRB est un système de gestion de fichiers permettant la distribution et l'accès à des données réparties sur différents sites et sur des supports hétérogènes (fichiers sur disque, bases de données, bandes magnétiques). Cet accès se fait de façon transparente du point de vue de l'utilisateur. SRB est porté sur de multiples plateformes, ainsi que sur les projets *grid computing*. Il est en outre facile d'utilisation, et peut-être utilisé par différents langages (C++, Java).
- **Base de donnée Oracle** (100 GO), elle permet la sauvegarde des données re-

construites auxquelles nous voulons accéder très rapidement et façon transversale. Elle permet à l'interface de visualisation d'avoir un accès presque en temps réel des données afin d'obtenir un rendu dynamique et interactif. C'est une base de donnée relationnelle contenant les descriptifs de chacune des meta-données. Un schéma simplifié de la base est représenté sur la figure A.9. Ce système a une granularité qui autorise la modification rapide d'éléments unaires.

A.4.5 Denali

Dans le cadre du développement de l'interface web et du développement du workflow, nous avons collaboré avec un partenaire industriel Denali spécialisé en *Business Process Management* (BPM) et *Service Oriented Architectures* (SOA). Denali a développé *FlexoBPM*, une technologie de BPM orientée *business* qui permet, de façon itérative et collaborative, d'éditer un workflow, ses écrans d'interaction avec les utilisateurs, de générer automatiquement des prototypes, de la documentation, des scénarios de tests et pouvoir réutiliser le code généré.

A.5 Workflow automatique

A.5.1 Définition de l'expérience

La première étape automatisé du workflow est la gestion des données en sortie du microscope. Cette étape se décompose en deux parties. La première concerne l'acquisition des données et ses paramètres, cette partie se fait de façon manuelle par les biologistes. Ils doivent remplir sur l'interface web du workflow les champs descriptifs d'une nouvelle fiche d'expérience (figure A.10). La deuxième partie complètement automatisée concerne l'envoi et la conversion des données vers le serveur de stockage. Elle rassemble plusieurs scripts dont certains s'exécutent en local, sur le lieu d'acquisition et d'autres au centre de calcul. A partir des formats propriétaires des microscopes, nous convertissons les données en format standard compressé afin de minimiser l'utilisation de la bande passante.

A.5.2 Description et exécution des algorithmes

Après une vérification des règles établies par la couche de *Implémentation*, les algorithmes sont ensuite définis dans l'interface web du workflow en spécifiant l'ensemble de leurs critères (figure A.12). Une liste des types d'algorithmes utilisés dans la reconstruction (filtrage, détection des centers, tracking, etc.) est prédéfinie, mais un nouveau type peut-être rajouté en respectant les contraintes du métalangage. L'application web du workflow archive automatiquement les différentes versions des algorithmes.

C'est au responsable de chacun des algorithmes de la couche *Implémentation* de le décrire dans l'interface du workflow. Cela permet ensuite aux *Utilisateurs* de

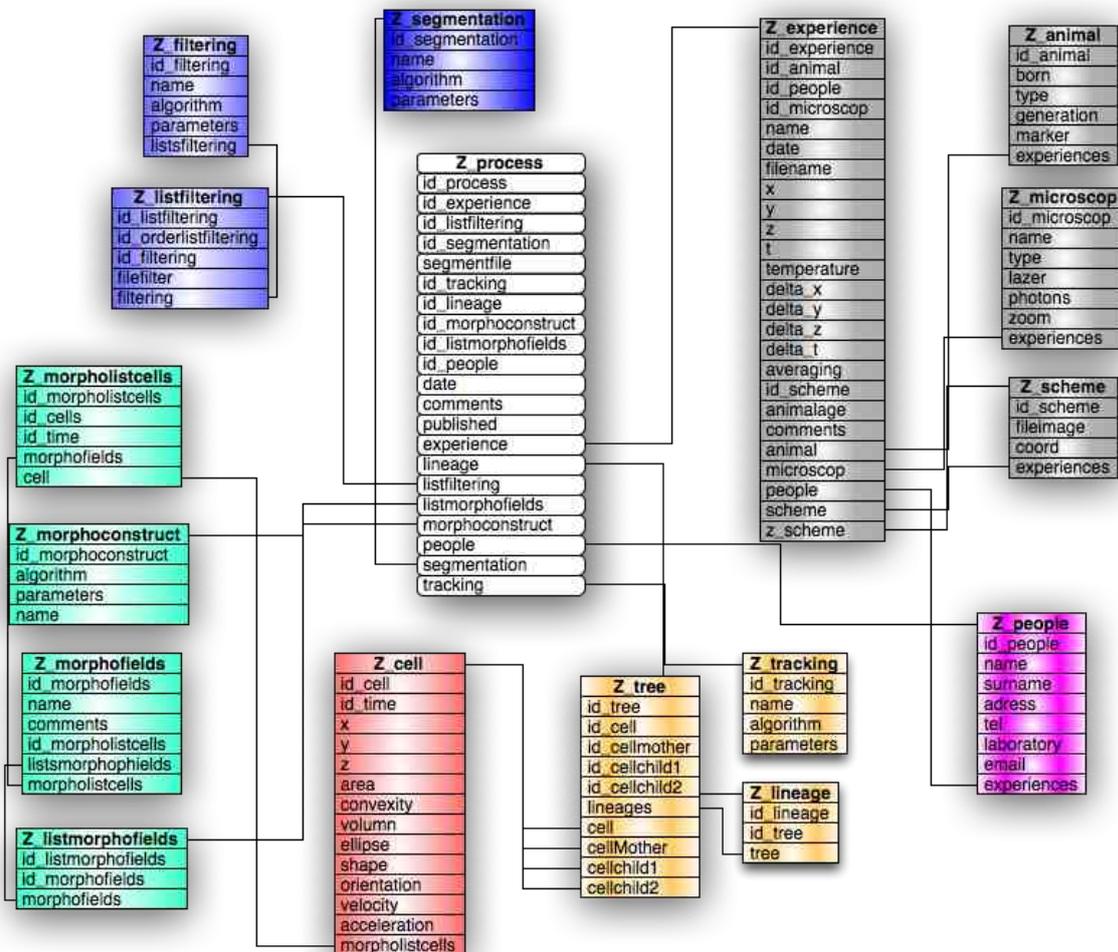


FIG. A.9 – Représentation graphique de la base de données du workflow. Chaque module représente une table de la base. La reconstruction d'un embryon passe par la mise en place d'un processus (*z_process*). Elle se fait en plusieurs étapes qui comporte chacune une table : l'expérience (*z_experience*) est effectuée par un biologiste (*z_people*) ; *z_filtering* contient les liens vers les données filtrées ; la segmentation des noyaux et des membranes des cellules sous forme de maillage triangulaire est directement sauvée dans *z_segmentation* ; de la cellule (*z_cell*) au tracking (*z_tree*), en passant par la reconstrution des champs morphogénétiques (*z_morphofields*).

FIG. A.10 – Les champs à remplir pour la saisie d’une nouvelle expérience. Description des onglets : **EXPERIMENT** : descriptif de l’expérience (date, nom de l’utilisateur, etc...); **ANIMAL** : description de l’animal (espèce, type, sexe, drogue, etc...); **MICROSCOPE** : le microscope avec lequel l’expérience a été faite (type, objectif, zoom, température.); **IMAGING** : description du volume imagé (nombres de coupes, temps imagé, taille des pixels, etc ...); **REPOSITORY** : emplacement actuel des données ; **NOTIFICATION** : permet de signaler par mail la personne qui a fait l’acquisition tout traitement de données qui sera effectué dessus.

FIG. A.11 – Définition d’un algorithme dans l’interface web du workflow. Description des onglets : **ALGORITHM** : description générale de l’algorithme (nom, type, version, mode (production / test)) et description des ressources (temps d’exécution maximum, mémoire vive, système d’exploitation, ferme de calcul parallèle, nombre de processeurs, etc...); **PARAMETERS** : spécification des paramètres (nom, type, valeur par défaut, bornes, description); **ADVANCED** : description plus détaillée de l’exécution de l’algorithme; **ITERATION** : fonctionnement des entrées sorties en cas d’itération.

FIG. A.12 –

les exécuter sur une expérience spécifique. Ils peuvent à tout moment contrôler la bonne exécution des algorithmes et, en cas d'erreurs, obtenir le descriptif détaillé de l'exécution, modifier les paramètres et relancer l'exécution.

Les données ainsi produites par les algorithmes du workflow sont récupérables depuis une interface spécifique à la gestion des données sur SRB et sont directement envoyées à la base de donnée connectée à l'interface de visualisation.

A.5.3 Définition des *pipelines*

L'étude des phénomènes de la morphogenèse nécessite l'exécution de différents algorithmes de façon parallèle ou séquentielle. Dans le but de rendre transparent au maximum la complexité des algorithmes aux utilisateurs du workflow, nous avons développé un module d'encapsulation de différents algorithmes (*pipelines*). Cette emboîtement se fait par une description des noms génériques des algorithmes sous formes de méta-langage, en connectant les entrées et les sorties respectives des différentes exécutions. La figure A.13 décrit un exemple de pipeline permettant depuis les données brutes sortant du microscope d'obtenir un arbre de lignage cellulaire. Ce graphe est généré automatiquement par l'interface web du workflow lors de la définition du pipeline. Le concept de pipeline permet selon les définitions des algorithmes d'ordonner les algorithmes qui s'emboîtent les uns après les autres et de faire tourner en parallèle les algorithmes n'ayant pas de dépendance entre eux. Nous contrôlons les erreurs à chaque étape de l'exécution du pipeline avant de passer à l'étape suivante (cf A.5.4).

Les utilisateurs peuvent, de façon similaire aux algorithmes, exécuter les pipelines depuis l'interface web. La gestion de l'exécution d'un pipeline permet à chaque étape de suivre son déroulement, et de relancer individuellement les algorithmes en cas d'erreurs.

A.5.4 Gestion des *jobs* et des erreurs

L'ensemble du workflow passe par une gestion des processus appelés *jobs*. La figure A.15 en représente une version schématisée. L'interface du workflow est cordonnée par une base de donnée (*Metadata DB*) qui gère l'historique des expériences, des algorithmes et des exécutions. Cette base de donnée ordonne les différentes tâches à exécuter (*job scheduler*). Ces tâches consistent à soumettre les jobs à exécuter sur les différentes grilles de calculs. avec l'ensemble des paramètres nécessaires à l'exécution des algorithmes et des expériences. De plus elle contient un système d'envoi d'email pour signaler la fin de l'exécution d'un pipeline ou d'un algorithme. La grille de calcul (*computers Grid*) envoie les différents jobs sur des clusters qui peuvent être parallèles (*MPI*) ou séquentiel (*SEQ*). Chaque grille de calcul est managée par un monitoring des jobs (*Jobs Monitor*) permettant au workflow de suivre l'exécution des algorithmes.

Le service des gestion des erreurs (*Error Management*, schématisé sur la figure A.15), vérifie à intervalle régulier le bon déroulement des jobs en cours. A la fin de

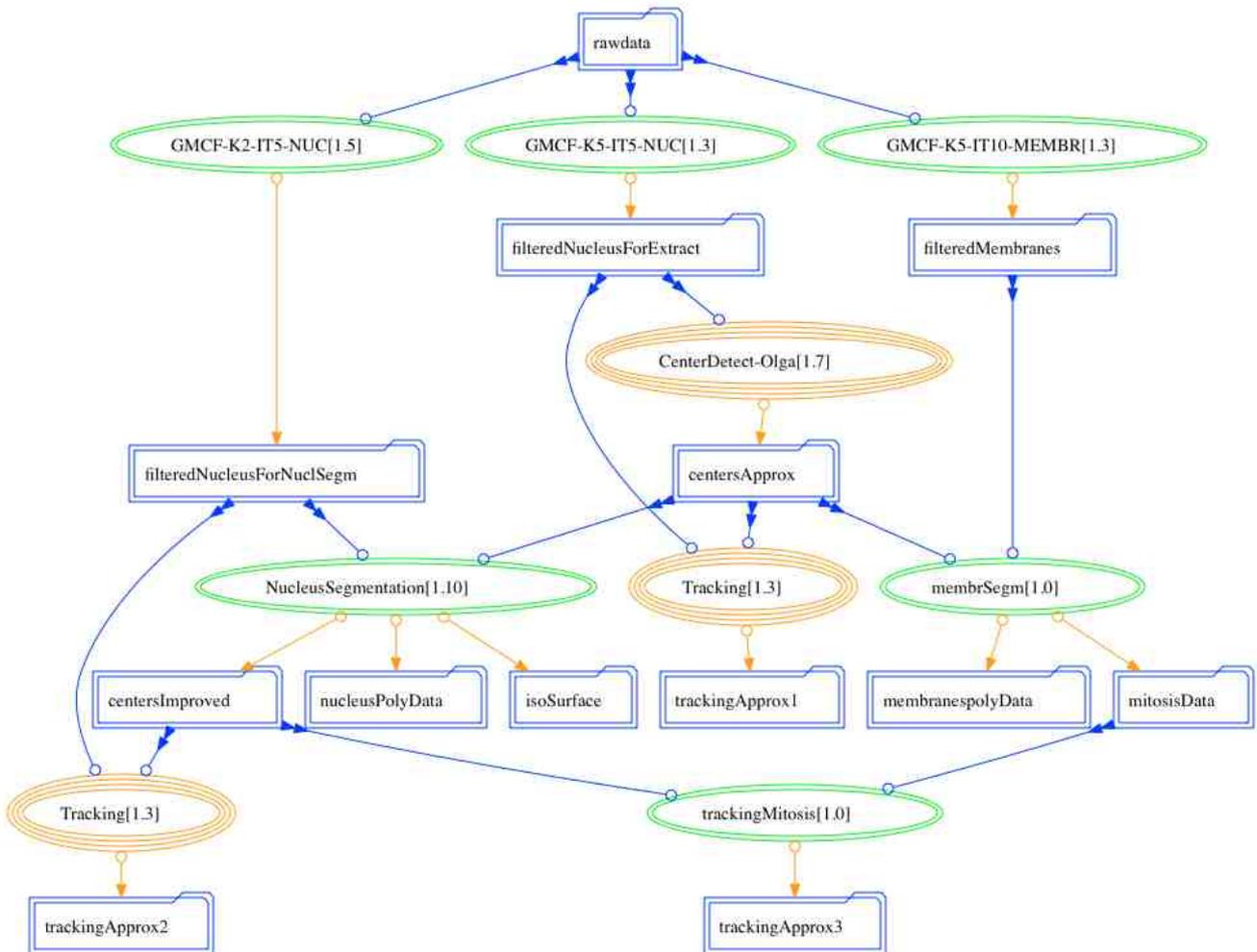


FIG. A.13 – Exemple de pipeline de différents algorithmes : représentation du pipeline de l’algorithme de tracking. Les cercles représentent les différents algorithmes nécessaire au processus global, les boîtes représentent des données intermédiaires ou finales. Des données brutes (*rawdata*) trois méthodes de filtrages sont utilisées (*GMCF...*) produisant chacune leur données en sortie (*filteredNucleus* et *filteredMembrane*). L’algorithme de détection des centres (*CenterDetect*) génère ensuite une collection de centres (*centersApprox*) permettant la segmentation des noyaux (*NucleusSegmentation*) et des membranes (*NucleumembrSegm*). Ensuite un premier tracking est effectué (*Tracking*) et enfin amélioré par la détection des mitoses *trackingMitosis*.

chacun des jobs, nous récupérons un *log* (fichier de sortie), permettant de caractériser les différents types d'erreurs. Nous avons développé des scripts permettant de solutionner les erreurs standards et de relancer automatiquement les jobs. Dans le cas d'une erreur non résoluble par les scripts l'interface web signale une erreur et un mail est envoyé automatiquement à l'utilisateur. De plus, nous gérons la liste des jobs en attente d'exécution, et dans un cas d'utilisation sur-intensive d'un centre de calcul, nous basculons certains jobs de façon automatique sur un autre centre.

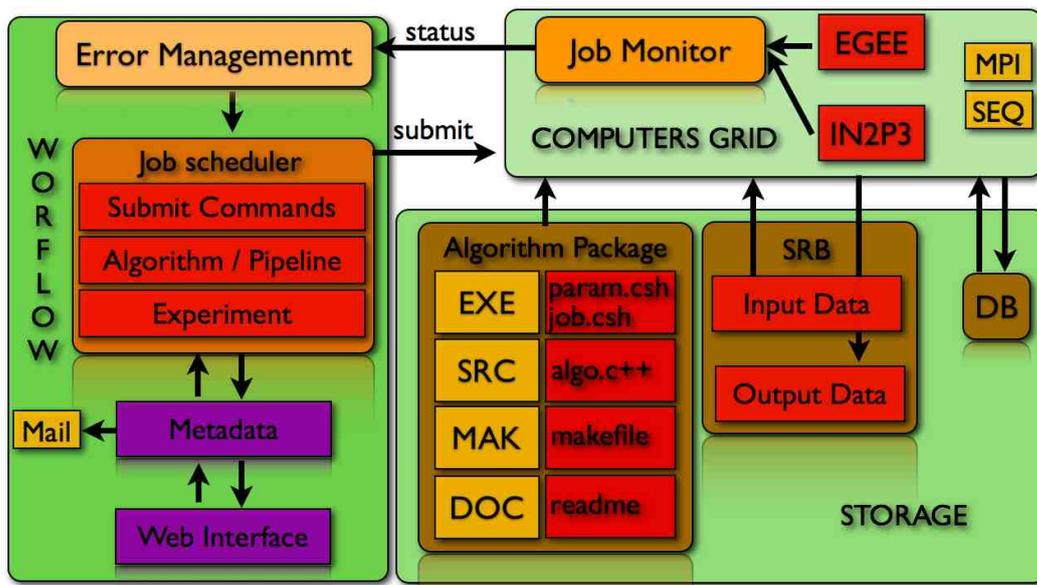


FIG. A.14 – Représentation schématique de la gestion des jobs par le workflow

A.6 Conclusion

La réalisation d'un système de gestion de flux de travail (un *workflow*) a pour but l'amélioration du rendement. L'automatisation de l'ensemble des traitements effectués sur une quantité massive de données permet d'obtenir une précision et une qualité qui ne seraient pas possibles d'atteindre manuellement. La reconstruction phénoménologique de la morphogenèse animale à partir d'imagerie *in toto* d'embryons de poisson zèbre a nécessité la mise en place de méthodes mathématiques et d'outils informatiques adaptés. Nous avons conçu un *workflow* de traitement des données permettant d'obtenir la reconstruction de l'embryogenèse de façon rapide, simple, standard et complètement automatisé. Nous avons pour cela défini des formalismes appropriés, des formats standards et trouver des solutions adaptées au stockage d'important volume de données. Nous avons utilisé des ressources de calculs à grande échelle afin d'obtenir des résultats dans des temps rai-

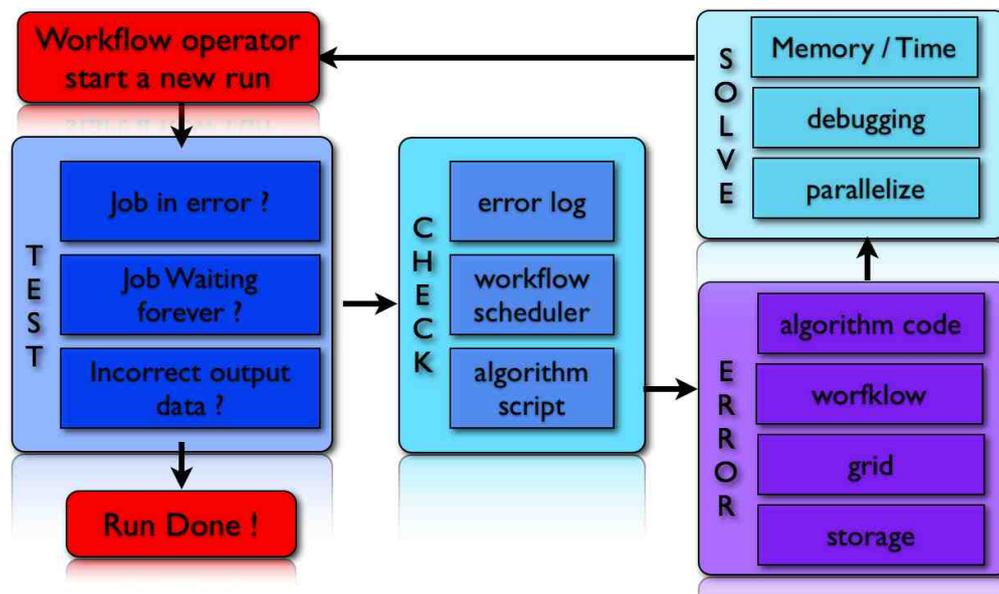


FIG. A.15 – Représentation schématique de la gestion des erreurs par le workflow. A la fin de chaque job, un test est effectué pour vérifier son bon déroulement. En cas d'erreur nous en analysons la cause, la classons selon son type et enfin si le type est référencé et résolvable de façon automatique, nous pouvons relancer le job.

sonnables (de l'ordre d'une journée pour la reconstruction totale d'un embryon). Nous avons développé de nouvelles techniques de traitement d'images et de reconstruction des données. Du filtrage à la reconstruction de l'arbre de lignage cellulaire, en passant par des outils statistiques, chacune des étapes du traitement des données embryonnaires est imbriquée. Par l'intermédiaire d'une interface web conviviale, les biologistes peuvent maintenant accéder à la reconstruction d'un embryon. Nous sommes à même de pouvoir comparer différents embryons d'une même classe sous des conditions génétiques et environnementales similaires. Mais nous pouvons aussi tester l'impact d'une nouvelle drogue contre le cancer sur plusieurs embryons, le reconstruire et le comparer automatiquement à un embryon non-taillé. Nous visons à travers ce workflow à fournir une plateforme européenne de traitement contre le cancer.

Embryomics : Reconstructing the Cell Lineage Tree as the Core of the Embryome

Résumé

La plateforme théorique, méthodologie et expérimental d'Embryomics se consacre à la reconstruction morphodynamique de l'arbre du lignage cellulaire de l'embryogenèse animale. Cette plateforme rassemble un ensemble de stratégies, de méthodes et d'algorithmes permettant de *séquencer* l'arbre de lignage cellulaire comme un processus de branchement spatio-temporel. Notre approche se situe au niveau de l'organisation cellulaire à fin d'obtenir une compréhension intégrée des processus biologiques par le biais de la modélisation de la dynamique des réseaux de gènes et l'analyse de la dynamique multi-échelles. Les morphodynamiques cellulaires sont reconstruites à partir d'images temporelles en 3D, obtenues en microscopie multiphoton à balayage laser au début du développement embryonnaire de la zebrafish. Les noyaux et des membranes de toutes les cellules sont colorées par l'expression d'une combinaison de protéines fluorescentes. Due aux limitations des techniques d'imagerie, l'ensemble de l'arbre de lignage cellulaire est d'abord réalisé par chevauchement de séquences spatio-temporelles. Nous définissons l'"*embryome*" d'un organisme comme la description dynamique multi-échelle des stades de développement qui corrént le génotype et le phénotype. Nous avons conçu des méthodes de traitement d'image, adaptées à la morphologie des cellules, basées sur les équations aux dérivées partielles pour effectuer un filtrage d'image en 4D, détecter les centres des noyaux, et segmenter la forme des membranes et des noyaux ainsi que détecter les mitoses grâce à la plateforme de calcul *Embryome*. Le suivi des cellules a été réalisé par une procédure d'estimation maximisation. Nous avons conçu une plateforme de visualisation permettant de visualiser simultanément les données brutes et reconstruites. Elle permet une exploration systématique de l'origine et de la complexité clonale des organes, ainsi que des mouvements cellulaires permettant la formation de structures locales et des champs morphogénétiques.

Sommaire

B.1 Methods	169
B.1.1 Embryo staining and mounting.	169
B.1.2 Image acquisition.	169
B.1.3 Image filtering.	170
B.1.4 Nuclei centre detection.	170
B.1.5 Segmentation of cellular nuclei and membranes.	170
B.1.6 Cell tracking.	171
B.1.7 Tracking validation.	171
B.1.8 Interactive Embryome Visualisation Platform.	172
B.1.9 Figures	173
B.2 Supplementary material	181
B.2.1 Table S1 Developmental table of the imaged embryo ID 070418a181	
B.2.2 Movies	181

We define the "embryome" of an organism as the multi-scale dynamics description of the early stages of development that correlates genotype and phenotype, integrating both upward and downward causation. In this sense, it is part of the broader concept of *physiome* as specified by the international Physiome Project ([Bassingthwaight \(2009\)](#)). The reconstruction of the embryome requires a paradigm of systematic investigation of the cellular behaviours and reconstruction of the cell lineage tree as a branching process in space and time. Here we show that original mathematical methods and computational tools inspired by the mammalian visual system can be used to achieve the automated reconstruction of embryogenesis, even at rates of over 50,000 cells per 24 hours, as is the case of the zebrafish *Danio rerio* ([Kane \(1999\)](#)). We have imaged the prospective brain of zebrafish embryos by multiphoton microscopy from 4 hours post-fertilisation through gastrulation and early neurulation stages ([Kimmel *et al.* \(1995\)](#)). The imaged cells, whose number grows from 3,000 to more than 11,000, were tracked with less than 2% error for at least 6 hours. Prior to this work, the automated reconstruction of the cell lineage tree attempted for *Caenorhabditis elegans* reached comparable accuracy only for the first 194 cells, despite much more limited cell movements ([Sulston *et al.* \(1983\)](#), [Bao *et al.* \(2006\)](#)). In the zebrafish, lineage studies previously relied at best on semi-automated cell tracking methods ([England *et al.* \(2006\)](#)). Our breakthrough in the reconstruction of the embryo lineage tree with standardised and controlled strategies launches "embryomics" as a new field that will link ontogeny and phylogeny through the reconstruction of the embryome of different species.

Two-photon-excited fluorescence (2PEF) microscopy, which allows *in vivo* imaging of embryos for extended periods of time, has proven to be the most effective technique for deep-tissue fluorescence imaging with sub-cellular resolution

(Squirrell *et al.* (1999), Helmchen & Denk (2005)) (Embryomics and BioEmergences EC consortia, unpublished). However, achieving the automated tracking of cells throughout the whole living embryo from 4D image data sets involves a difficult compromise between many variables, such as signal-to-noise ratio, spatial and temporal resolution, thickness of the explored volume and cell survival, which has not been successfully challenged to date.

We report here about methodologies and tools that combine for the first time imaging and fully automated reconstruction of cell behaviours to build a complete zebrafish virtual embryo. We have designed a multishot strategy based on a spatio-temporal tiling scheme that incorporates views taken from different embryos. These embryos were observed either from the animal pole (AP scheme, Movie B.2.2.1) or the vegetal pole (VP scheme, Movie B.2.2.2), possibly completed by lateral views (L scheme, Movie B.2.2.3) (Figure B.1b). The AP scheme was chosen for the first step of our strategy of phenomenological reconstruction. Embryo labelling was obtained through RNA injection performed at the one-cell stage to obtain ubiquitous expression of H2B/mCherry fusion protein and farnesylated eGFP, which stained nuclei and membranes respectively (Moriyoshi *et al.* (1996), Shaner *et al.* (2005)). While nuclear staining was instrumental to perform cell tracking, membrane staining was essential to assess cell morphology, behaviour and neighbourhood, and to reveal morphological landmarks. Embryo imaging started when sufficient signal was detected (4 hours post-fertilisation hpf at 28°C, i.e., at the sphere stage) and imaging proceeded for the next 10 hours at 26°C (Figure B.1a-h; Table S1, Movie B.2.2.1). Following the image processing steps of the phenomenological reconstruction, the virtual embryo consisting of nucleus centres and individual cell trajectories could be explored with our interactive Embryome Visualisation Platform, which revealed the cells' clonal history and allowed to derive cell population dynamics and behaviours (Figure B.1i-l, Movie B.2.2.4 and B.2.2.5).

Our computational image-processing framework performed nonlinear image multiscale analysis (Alvarez *et al.* (1993)) inspired by theoretical knowledge of mammalian vision. For every low level image-processing task, we designed specific nonlinear partial differential equations (PDEs) that modelled some of the remarkable adaptive properties of the mammalian visual system and were also able to deal with more difficult recognition tasks at later developmental stages.

We filtered the image using so-called "geodesic mean curvature flow" (GMCF) in the level set formulation, a particular form of nonlinear geometrical PDEs that appeared to perform best when processing 2PEF microscopy images (Rizzi *et al.* (2007b)). In GMCF, the contrast-invariant flows achieved nonlinear multiscale analysis on scalar intensity functions in grey-level scale (Figure B.2a-c, Movie B.2.2.6). Filtered images were used in the subsequent algorithmic steps.

Nuclear centre detection was performed using the PDE-based nonlinear multiscale strategy called "flux-based level set centre detection" (FBLSCD). All visible objects in the image could be seen as humps of relatively higher image intensity. The multiscale FBLSCD method made these humps decrease until a certain stopping condition adapted to the rest of the algorithm chain was achieved (Figure B.2d,e, Movie B.2.2.6) (Frolkovic *et al.* (2007)). Approximate nuclear centres were defined as the final position of local maxima.

Cell nuclei and membranes were automatically segmented (Figure B.2f-j, Movie B.2.2.7) using the approximate cell centres produced by the above step as the gaze from which a point-of-view surface was constructed. The evolution of the point-of-view surface, with respect to a metric induced by the image, tended toward a minimal surface in a Riemannian manifold representing the segmentation of the nuclear or membrane shape. This "subjective surface technique" (SST) was used for its ability to fill missing information in the image (Sarti *et al.* (2000)). The computational implementation of SST, GMCF and FBLSCD was based on co-volume methods (Corsaro *et al.* (2007), Mikula *et al.* (2008b)) and parallel computer execution (Mikula & Sarti (2007)). Nuclear segmentation was then used to correct "double centres" that remained from the nuclear centre detection step (Figure B.2h). Finally, membrane segmentation was used to detect mitosis (Figure B.2k-l) and this information was implemented into the tracking algorithm. The number of cells in the imaged volume increased from 3,291 at $t=0$ (sphere stage) to 11,176 at $t=540$ (8-somite stage), representing about 1/3 of the total embryo's size 2. The average cell density obtained from the segmentation of the global imaged volume was plotted as a function of time (Figure B.2m-o). Cell density increased throughout gastrulation and stabilised by early somitogenesis ($t=300-360$). This observation fits with the decrease of cell volume through cell divisions observed during early embryogenesis and provided an estimate of the average proliferation rate during this developmental period (Kane *et al.* (1992)).

Cell tracking in the 4D space of the segmented data was not properly achieved by classical methods using a purely local nearest-neighbour principle. Although we kept temporal resolution as fine as possible (Δt is only 67 seconds in the image data set 070418a), cell displacement during mitosis was a major problem that the tracking algorithm had to solve (Melani *et al.* (2007)). The spatio-temporal lineage tree reconstruction was best achieved with a tracking algorithm relying on the minimisation of a heuristic functional. Strong constraints were used i) *a priori*, i.e., before minimisation, to restrict the unknown tracking function and ii) *a posteriori*, i.e. after minimisation, to identify biological inconsistencies for correction of false negatives and false positives in the nuclear centre and mitosis detection steps. The heuristic functional assigning penalties for high displacements and deformations in cellular feature space was minimised by simulated annealing (Kirkpatrick *et al.* (1983)).

The outcome of the tracking procedure was assessed by using dedicated tools from our Embryome Visualisation Platform and following the trajectories of individual cells. A cell that was properly tracked through consecutive cell divisions is shown in Figure B.3a-c (Movie B.2.2.8). Our visualisation platform allows going back and forth between the 4D rendering and the annotated flat representation of the tree (Figure B.3b). The algorithmic chain provides for each cell at each time step its identification number, nuclear centre 4D coordinates, nuclear and membrane shape, and the times of the previous and next division along the tree. Quantitative parameters describing cell dynamics can then be derived from these data. The tracking error rate was measured with the sampling procedure described in Figure B.3c,d and Methods section. Altogether, over 15,000 links between the t and $t+1$ positions of cells have been manually checked at different time points. For the first 300 time steps (5h34 imaging, end of gastrulation) the total error rate is lower than 2%, i.e., more than 98% of the cells were correctly tracked (Figure B.3d). Tracking accuracy degraded somewhat beyond time step 400 (7h26 imaging, 3-somite stage), but 90% of the links were still correctly found by time step 500 (9h18 minutes imaging, 7-somite stage). False positives in nuclear detection were kept below 0.5% and were not the major source of tracking errors. About half of the tracking errors were identified as false trajectories (the tracking jumped from one cell trajectory at time t to another at time $t+1$) and the other half was false deaths (the tracking did not find the position of the cell at time $t+1$, and cell trajectory ended). All the information provided by error checking and categorisation was used to improve the performances of the tracking procedure. Thus the reconstruction accuracy could be further improved by users' annotations through the embryome visualisation platform.

The exploitation of the virtual embryo opens the way to *in silico* experimental embryology to follow the formation and transformation of 4D morphogenetic fields, backtracking a 3D morphogenetic section towards its earlier 3D fate map. We explored this potential by manually selecting morphogenetic fields in the brain at $t=479$ (6/7-somite stage), based on morphological landmarks (Figure B.4d). Backtracking the selected cells, which were categorised in 12 domains, revealed cellular population dynamics consistent with the order and coherence of the brain fate map (Kimmel (1994), Woo & Fraser (1995)). The latter - as all fate maps in zebrafish - was however determined no earlier than the shield stage (gastrulation onset), and only for the most superficial cell layer of the 4-5 "deep" cell layers of the embryo, leaving the fate of all deeper cells unknown. Our data reveal that this fate map actually holds for the whole depth of the blastoderm, and that it can be backtracked to the earlier sphere stage, with no major loss in coherence (Figure B.4a-c, Movies B.2.2.9, B.2.2.10, B.2.2.11).

Further analysing the collective cell behaviours requires integrating at least three levels of organisation : single cells, cell neighbourhoods and cell populations. We explored such a multiscale approach to investigate cell behaviour in the

presumptive forebrain at the midline (Bao *et al.* (2006), Varga *et al.* (1999), Wilson & Houart (2004)). The speeds of selected cells from the presumptive hypothalamus (PH), presumptive ventral telencephalon (PVT), and prechordal plate (PCP) were averaged over 10 time steps and plotted as a function of time between the onset of gastrulation and early somitogenesis (Figure B.4e right panel). These measurements indicated transitions in cell behaviours correlating with the relative positions of PCP, PH and PVT (Figure B.4e right panel, Movie B.2.2.12). We observed that the speed of PCP decreased as it met PH territory (t123, 60% epiboly). PH, which was moving anti-parallel to PCP, started moving rostrally by 75% epiboly (t200). The three selected cell populations accelerated similarly while keeping different speeds. By the end of gastrulation (t260), PCP slowed down abruptly. For the next hour, PH moved faster than the two others and passed under PVT. These observations, previously inaccessible but readily provided by our *in toto* reconstruction strategy, definitively validate the subduction scenario proposed earlier (Bao *et al.* (2006)), hence revealing coordinated behaviour between PCP and PH cells. Further measurement of the distance between cells of pairs chosen either within the same field or on both sides of the frontier between the two cell populations, indicate that PCP and PH cells do not remain tightly associated, rather they slide on top of each other (Figure B.4f, Movie B.2.2.13). This transient interaction, which is expected to allow proper cell signalling, is part of the biomechanical process driving PH cells rostrally. However, the cell population dynamics revealed by our phenomenological reconstruction indicated that PH displacement might depend on larger scale constraints.

Cell behaviour and clonal history are the basis for understanding morphogenetic processes (Petit *et al.* (2005)). The virtual embryo allows a systematic exploration of the clonal origin and clonal complexity of organs, as well as the contribution of cell proliferation modes and cell movements to the formation of local patterns and morphogenetic fields. By providing essential parameters of cell behaviour and clonal history, this first assembly of the cell lineage tree underlying zebrafish brain early morphogenesis constitutes the core of the embryome, i.e., the phenomenological and theoretical reconstruction of the entire system's multiscale dynamics.

Completing the reconstruction of zebrafish early embryogenesis with our multishot strategy will prompt further investigation and measurement of individual variation. We expect that qualitative descriptions of biologists based on this concept and others manipulated in this study (e.g., *morphogenetic fields*) will be turned into more quantitative, formal descriptions amenable to automated treatment. By entering the field of complex systems science, biologists might become more dependent on interdisciplinary interactions, but will also gain greatly in time and explanatory power. We expect the field of *embryomics* to follow the path of high-energy physics : the strength and usefulness of phenomenological and theoretical reconstructions will come from the efforts of the scientific community to standar-

dise and share their strategies, methods and open tools.

B.1 Methods

B.1.1 Embryo staining and mounting.

Wild-type zebrafish embryos were injected at the one cell stage with 200pg mCherry/H2B RNA and 200pg eGFP-ras prepared from PCS2+ constructs (Moriyoshi *et al.* (1996), Shaner *et al.* (2005)). Although mCherry, unlike eGFP, bleached significantly through imaging, this colour combination was the best compromise allowing proper staining of the cell membranes for further segmentation. Injected embryos were raised at 28.5°C for the next 3 hours. Embryos were mounted in a 3cm Petri dish with a glass coverslip bottom, sealing a hole of 0.5mm at the Petri dish centre where a Teflon tore (ALPHAnov) with a hole of 780 μm received the dechorionated embryo. The embryo was maintained and properly oriented by infiltrating around it 0.5% low melting point agarose (Sigma) in embryo medium (Westerfield (1995)). Temperature control in the room resulted in a temperature of about 26°C under the objective slightly slowing down development with respect to the standard 28.5°C developmental table (Kimmel *et al.* (1995)) (Table S1). After the imaging procedure, embryo morphology was checked under the dissecting binocular and the animal was raised for at least 24h to assess morphological defects. Embryo survival depends on total imaging duration, average laser power, and image acquisition frequency (or time step Δt). 070418a and 070420a correspond to a standardised procedure with 50s < Δt < 70s allowing 75-110 sections per time point for 10 hours with an average laser power of 80 mW delivered to the sample. Lowering the laser power to less than 60 mW and lengthening Δt up to 3.5min allowed imaging embryos for more than 20 hours, then raising them until adulthood. These conditions were used for making up to 320 sections at 400Hz line scan rate (bidirectional scanning), or 200Hz to improve signal-to-noise ratio, as for the image data set 080322a shown in Movie B.2.2.3.

B.1.2 Image acquisition.

Imaging was achieved with a Leica DM6000 upright microscope SP5 MLSM equipped with a 20/0.95NA W dipping lens objective (Olympus). Axial resolution at the sample surface (1.5 μm) was estimated by recording 3D images of 0.1 or 1 μm fluorescent polystyrene beads (Invitrogen) at the surface of an agarose gel. Field size was 700-700 in x, y, 140 μm in z for 070418a and 100 μm in z for 070420a ; voxel size was 1,37 μm^3 . Simultaneous dual wavelength excitation with pulses at two different wavelengths (980 and 1030nm) with pulsed laser beams (50Mhz, 200fs) was provided by a solid-state Ytterbium femtosecond oscillator (T-pulse 20, Amplitude Systèmes) operating at 1030nm. The beam was split by a polarising cube. One of the two beams was launched into a nonlinear photonic crystal fibre (Amplitude Systèmes) that shifted the optical spectrum from 1030 to 980nm. Light beams

(1030nm and 980nm) were then recombined before being injected into the microscope. Raw data visualisation in Figure B.1e-h, Movies B.2.2.1,B.2.2.2,B.2.2.3, was done using Amira software (Mercury Computer Systems).

B.1.3 Image filtering.

The partial differential equation for the GMCF filtering reads as follows (Chen *et al.* (2000)) :

$$u_t = |\nabla u| \nabla \cdot (g(|\nabla G_\sigma * u|) \frac{\nabla u}{|\nabla u|}) \quad (\text{B.1})$$

and is accompanied by the initial condition $u(0, x) = u_N^0(x)$ or $u(0, x) = u_M^0(x)$, depending on which channel is filtered, and taking into account the zero conditions of the image domain $\partial\Omega$. The solution $u(t, x), t > 0$ represents the filtered image. In the model, the mean curvature motion of level sets of function u is determined by the edge indicator function $g(s) = 1/(1 + Ks^2), K \geq 0$, applied to the image gradient pre-smoothed by the Gaussian kernel G_σ with small variance σ .

B.1.4 Nuclei centre detection.

In our method, the normal velocity V of any level set is given by $V = \delta + \mu k$ where δ and μ are constants (model parameters) and k is the mean curvature. Such process of the level set evolution can then be represented by the following equation (Frolovic *et al.* (2007)) :

$$\partial_t u = \delta \frac{\nabla u}{|\nabla u|} \cdot \nabla u + \mu |\nabla u| \nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right) \quad (\text{B.2})$$

applied to the initial condition $u_{N_f}^0$ resulting from the filtering algorithm (eq. B.1). Due to the shrinking and smoothing of all (real and spurious) structures in this evolution process, we observe a decrease in the number of local maxima of the solution u .

B.1.5 Segmentation of cellular nuclei and membranes.

The shape of cellular nuclei or membranes is extracted by evolving initial segmentation functions to solve the generalised subjective surface equation (Sarti *et al.* (2000), Mikula *et al.* (2008b)) :

$$\partial_t u - w_{dif} g(|\nabla G_\sigma * u_{N_f}^0|) \sqrt{\varepsilon^2 + |\nabla u|^2} \nabla \cdot \left(\frac{\nabla u}{\sqrt{\varepsilon^2 + |\nabla u|^2}} \right) - w_{con} \nabla g(|\nabla G_\sigma * u_{N_f}^0|) \nabla u = 0 \quad (\text{B.3})$$

where $\varepsilon > 0, w_{dif} > 0, w_{con} > 0$, are the method parameters. The edge indicator function $g(s)$ is of the form $g(s) = f(1/(1 + Ks^2)), K \geq 0$, where f is an appropriate (most typically a power) function. The edge indicator function $g(s)$ is again applied to the image gradient pre-smoothed by the Gaussian kernel G_σ with small variance σ .

B.1.6 Cell tracking.

Let I be the whole set of cells at all time steps. The unknown tracking graph f on I , representing the lineage tree, was determined in three phases : i) the first phase used strong local constraints (maximum speed, binary tree structure) for restricting the unknown tracking graph ; ii) the second phase minimised a heuristic functional, combining information extracted from images (nuclear positions, large scale vector field, segmentation) with biological regularisation constraints (elastic behaviour) to measure the correctness of cell matching at consecutive time steps. The heuristic functional had the general form :

$$E(f) = \sum_i \|X_{f(i)} - X_i\|_1 + \sum_i \sum_{j \in V_i} \|X_{f(i)} - X_i - X_{f(j)} - X_j\|_2 \quad (\text{B.4})$$

where i and j are indicators of neighbour cells at time t , X_i the vector of cell i features (including its position), V_i is the neighbours of the cell i , $\|\dots\|_1$ a norm in the feature space and $\|\dots\|_2$ a norm in the feature deformation space. The functional minimisation was done by a simulated annealing algorithm (Kirkpatrick *et al.* (1983)), parallelised and distributed on a computer cluster. iii) The third phase's goal was to minimise false-positive and false-negative errors from the nuclear and mitosis detection steps. This was achieved by looking at the whole biological coherence of the lineage tree identifying : a) discontinuities in cell trajectories as false negative nuclear centres and b) cells that lived for only a few times steps as false positive nuclear centres.

B.1.7 Tracking validation.

Tracking accuracy was measured by assessing t to $t+1$ link for selections of cells over 10 time steps from time t_{n0} ($t_{n0} \in \{1, 20, 100, 200, 300, 400, 500\}$). Altogether, 15,686 links were manually checked, 1603 at $t_{n0} = 1$ (i.e. from t_0 to t_1 and so on for other t_{n0}); 1762 at $t_{n0} = 20$; 1828 at $t_{n0} = 100$; 2948 at $t_{n0} = 200$; 3169 at $t_{n0} = 300$; 2931 at $t_{n0} = 400$; 1445 at $t_{n0} = 500$. In addition to tracking links from t to $t + 1$, the checking procedure provided an error measurement on cell centre detection. These measurements were used to estimate the tracking error rate. The total error number (n_{error}) was defined as the number of incorrect t to $t + 1$ links ($n_{badLinks}$) plus the number of false deaths ($n_{falseDeath}$). The $n_{badLinks}$ were subdivided into two categories : errors due to false positive nuclear centres ($n_{falsePositive}$) and errors due to false tracks : the tracking jumped from one cell trajectory at time t to another cell trajectory at time $t+1$, (n_{change}). The total number of errors was given by $n_{error} = n_{falsePositive} + n_{change} + n_{falseDeath}$. The tracking error rate at a given time point t_{n0} was expressed as a percentage from the ratio of the error number (n_{error}) over the total number of tested links ($n_{allLinks}$).

B.1.8 Interactive Embryome Visualisation Platform.

Data set 070418a raw data, segmented nuclear centres and cell tracking are provided in a demonstration version of the interactive Embryome Visualisation Platform. <http://bioemergence.in2p3.fr/embryomics/embryome.html>

B.1.9 Figures

B.1.9.1 Figure B.1 : Reconstructing zebrafish brain early embryogenesis from time lapse optical sectioning

(a, c, d) Bright field views of live zebrafish embryos; (a) sphere stage, arrow-heads point to animal pole (AP) and vegetal pole (VP); (c, d) 8-somite stage, (c) lateral view, dorsal to the right, (d) animal pole view, rostral to the top; (b) schematic and idealised representation of the spatiotemporal tiling strategy with four overlapping volumes, AP scheme in pink, VP scheme in blue, L (lateral view) scheme in yellow and green. (e-h) Image data set ID 070418a, raw data 3D visualisation, animal pole view, same orientation as (d), orthoslice removing upper sections, time points and shown (z projected) sections indicated top right, developmental stage indicated bottom left. (i-l) Virtual embryo created with the Embryome Visualisation Platform, virtual counterpart of raw data (e-h); (i) colour code for cell displacement from yellow (rostralwards) through grey (null or medio-lateral) to blue (caudalwards) indicating cell displacement orientation; (j) inset, magnification showing each nucleus centre representation as a cube with a vector indicating cell trajectory for the next 12 time steps; fb, forebrain; hb, hindbrain; mhb, midbrain-hindbrain boundary; nk, neural keel; opv, optic vesicle; otv, otic vesicle; pcp, prechordal plate; so, somites; y, yolk; ysl, yolk syncytial layer; tb, tail-bud.

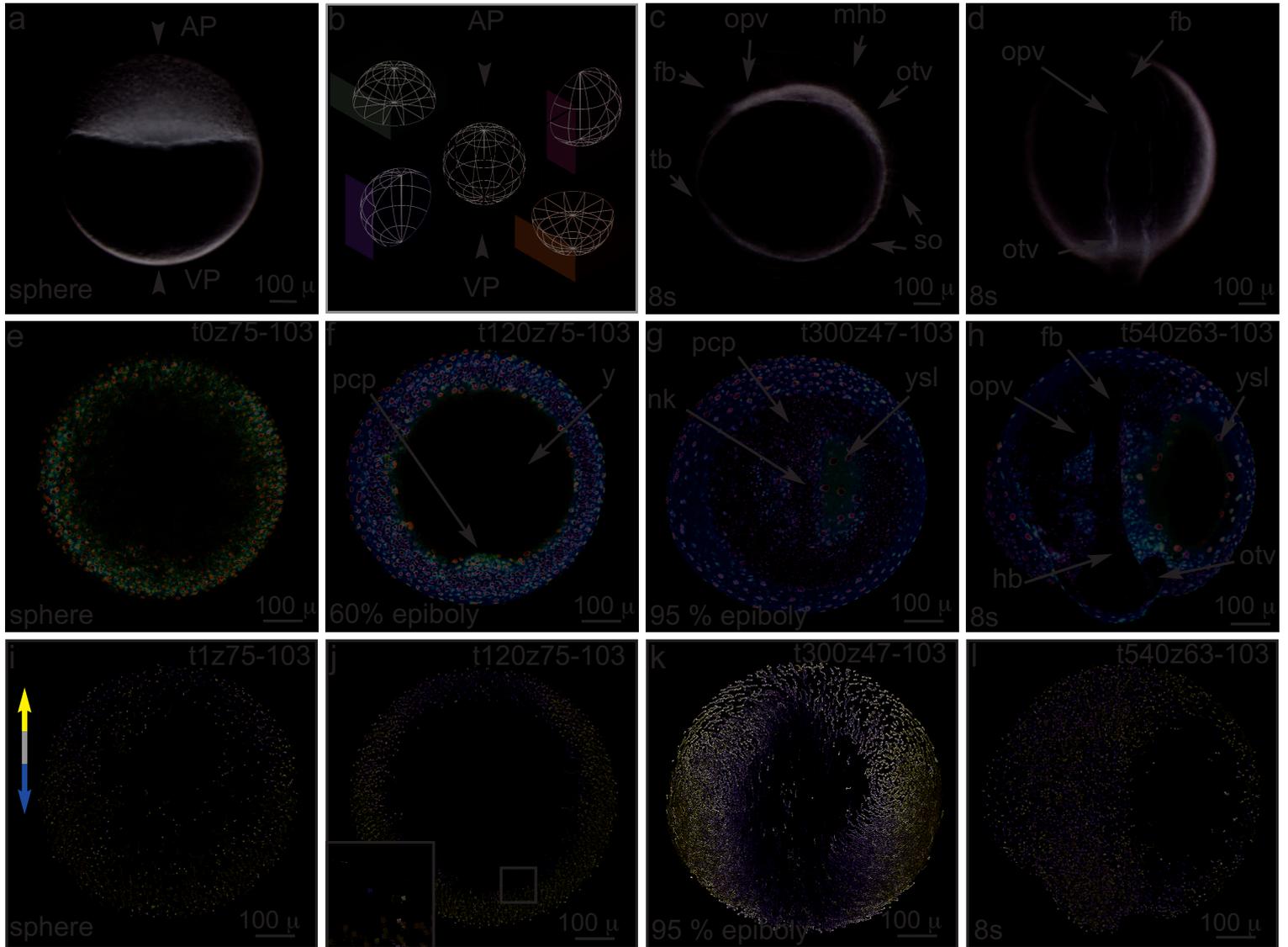


FIG. B.1 – Reconstructing zebrafish brain early embryogenesis from time lapse optical sectioning

B.1.9.2 Figure B.2 : A PDEs-based chain of algorithms for the 3D segmentation of embryonic cells

(a-e) Data set ID 070418a, image filtering and nuclei centres detection at t500, section z70 ; (a) raw data xy plane, same orientation as in Figure B.1e-l ; (b-d) plot of local maxima, magnification of the inset from (a) ; (c) result of the filtering algorithm (five steps) applied to (a) ; (d) result of the multiscale process (15 steps) for detecting nuclear centres applied to (c) ; (e) x,y frame with nuclei centres representation as cubes (blue) ; inset, magnification to show superimposed centres and raw data (grey scale) ; note that nucleus centre may be below or above the raw data section. (f-k) Data set ID 070420a, AP scheme similar to 070418a, f-h Nuclei segmentation, (f) top panels, 2D cut of initial segmentation function (pink) superimposed with raw data ; bottom panels, final segmentation function (orange) ; left panels, cell ID 707/t001 ; right panels, cell ID 780/t001, metaphase nucleus ; (g, h) filtered data shown in grey scale, (g) cells ID at t001, 1628, 1638, 1671 ; top panel, SST applied to nuclei segmentation, orthoslices view ; bottom panel, same as top panel but 3D rendering. (h) Cell ID 1089/t001, correction of double centres by comparing nuclei segmentation functions ; top panel, originally detected centres and corresponding segmentations ; bottom panel, correction by averaging the two centres and merging segmentation functions ; (i,j) membrane segmentation, (i) cell ID 802/t001, superimposed membrane segmentation and raw data ; top panel, two orthoslices ; bottom panel, 3D rendering ; (j) whole volume, 3D rendering of segmented cell shapes ; (k,l) mitosis detection, (k, m) filtered data shown in grey scale, (k) left panel, cell ID 1089/t001, initial segmentation function, two centres ; right panel, final segmentation, the two centres are found inside one cell ; (l) mother cell ID 215/t008, 3D rendering of nucleus and membrane shape throughout mitosis ; (m-o) data set 070418a, (m) shape segmentation of outer and inner embryo surface, time step and section indicated top right ; (n) local cell density in the whole volume at t479, same x,y plane as (m) ; (o) plot of the total cell number (blue curve) and cell density (brown curve) as a function of time ($\Delta t=67s$), arrowhead points to the stabilisation of cell density by early somitogenesis.

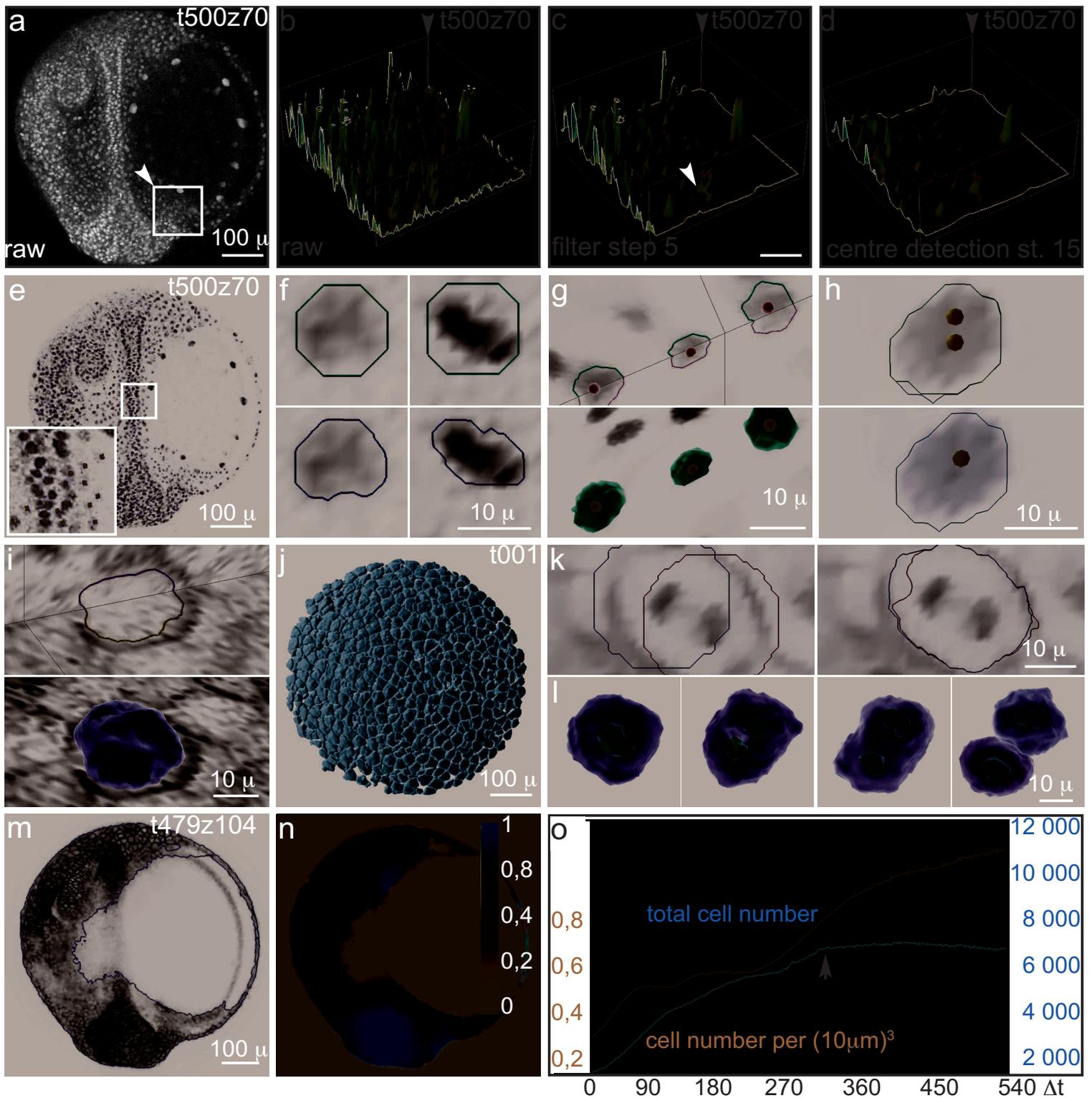


FIG. B.2 – A PDEs-based chain of algorithms for the 3D segmentation of embryonic cells

B.1.9.3 Figure B.3 : The lineage tree reconstruction and validation

Image data set 070418a, visualisation and validation of the individual cell tracking, Embryome Visualisation Platform ; (a) cell centre (blue) and 2 time steps cell trajectory superimposed with raw data membranes (grey scale), time point indicated top right of each panel ; (b) left panel, flat representation of the lineage tree ; the cell shown in (a) is circled in blue, its sub-tree is shown in blue, the rest of the tree is shown in white, note sister cell division at t88 ; right panel, for each cell, annotations include ID and the corresponding x,y,z,t coordinates ; membrane (mb) shape and nucleus (nuc) shape in the form of triangulated surfaces, time of previous (tp) and next division (tn) ; (c) nuclear centre detection checked for a sub-population of cells here at t400 (also done for other time points as described in Methods), correct centres indicated in green, false positive in red (right panel, one red nucleus in the inset), vertical white bars through nucleus centre used to assess position relative to raw data orthoslice ; tracking was manually checked for this population from t400 to t410 ; (d) Plot of tracking errors according to the validation strategy described in Methods.

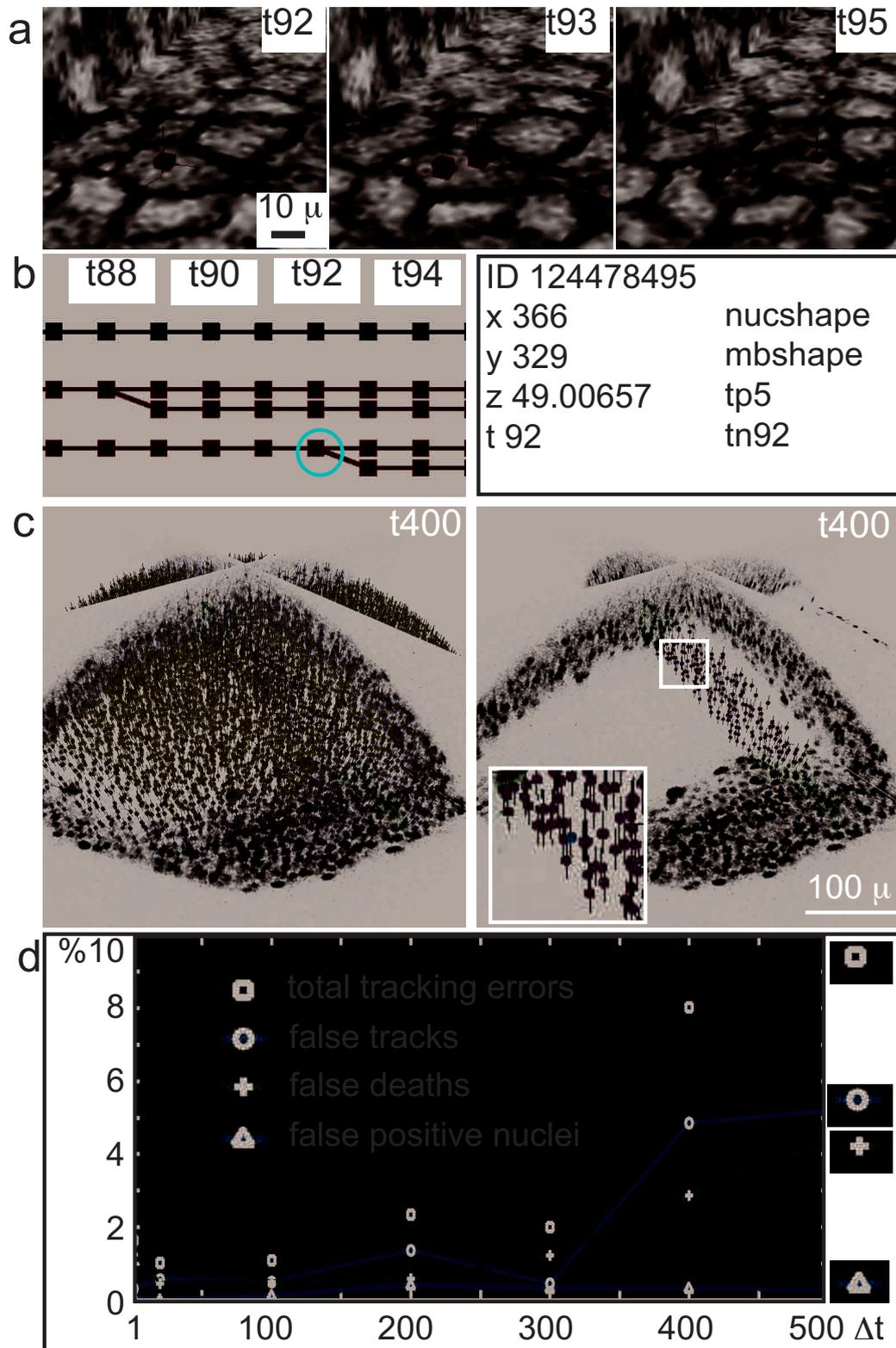


FIG. B.3 – The lineage tree reconstruction and validation

B.1.9.4 Figure B.4 : Multi scale measurements of cell behaviours from their clonal history

(a-f) Data set 070418a ; (a-d) animal pole view from the surface ; (a-c) trajectories of cells selected in presumptive organs at t479 and backtracked, temporal window for the displayed trajectories indicated top right, corresponding developmental periods indicated bottom left ; (d) at t479, 3D rendering of nuclear centres coloured according to the following code : yellow, hypoblast including PCP and notochord (h) ; red, presumptive hypothalamus (ph) ; pink, right eye (e) ; light pink, right optic stalk (os) ; purple, left eye ; light purple, left optic stalk ; green, ventral telencephalon (t) ; gold, dorsal telencephalon, placodes and neural crest ; light blue, ventral midbrain (mb) ; light brown, dorsal midbrain and dorsal diencephalon ; blue, ventral hindbrain (hb) ; dark brown, dorsal hindbrain ; dark blue, ears (ea). (e) Left panels, relative movements of cells selected in the prechordal plate (PCP, yellow), presumptive hypothalamus (PH, red) and presumptive ventral telencephalon (PVT, green), lateral views, rostral to the left, orthoslice of raw data (nuclei) behind selected cells, time step indicated top right of each panel ; right panel, average speed of the selected cells over 10 time steps (in $\mu\text{m}/\text{min}$), positive values indicate anterior movement, arrowheads indicate time points shown in left panels ; (f) Left panel, PCP (yellow) and PH (red) cell populations in which pairs of adjacent cells were chosen at t235 ; view from inside the embryo ; white arrowhead points away from the anterior border of PCP migrating from top to bottom of the image ; inset shows magnification of adjacent cells across PCP-PH frontier (dotted line) ; blue, centre of non selected cell ; (f) right panel, average distance and variance for the selected cell pairs, tracked between t220 and t300, within PCP (yellow, 27 pairs), within PH (red, 37 pairs), or across the PCP-PH frontier (brown, 38 pairs) ; dotted line at t235 when pairs of adjacent cells were chosen, all cell trajectories have been manually checked.

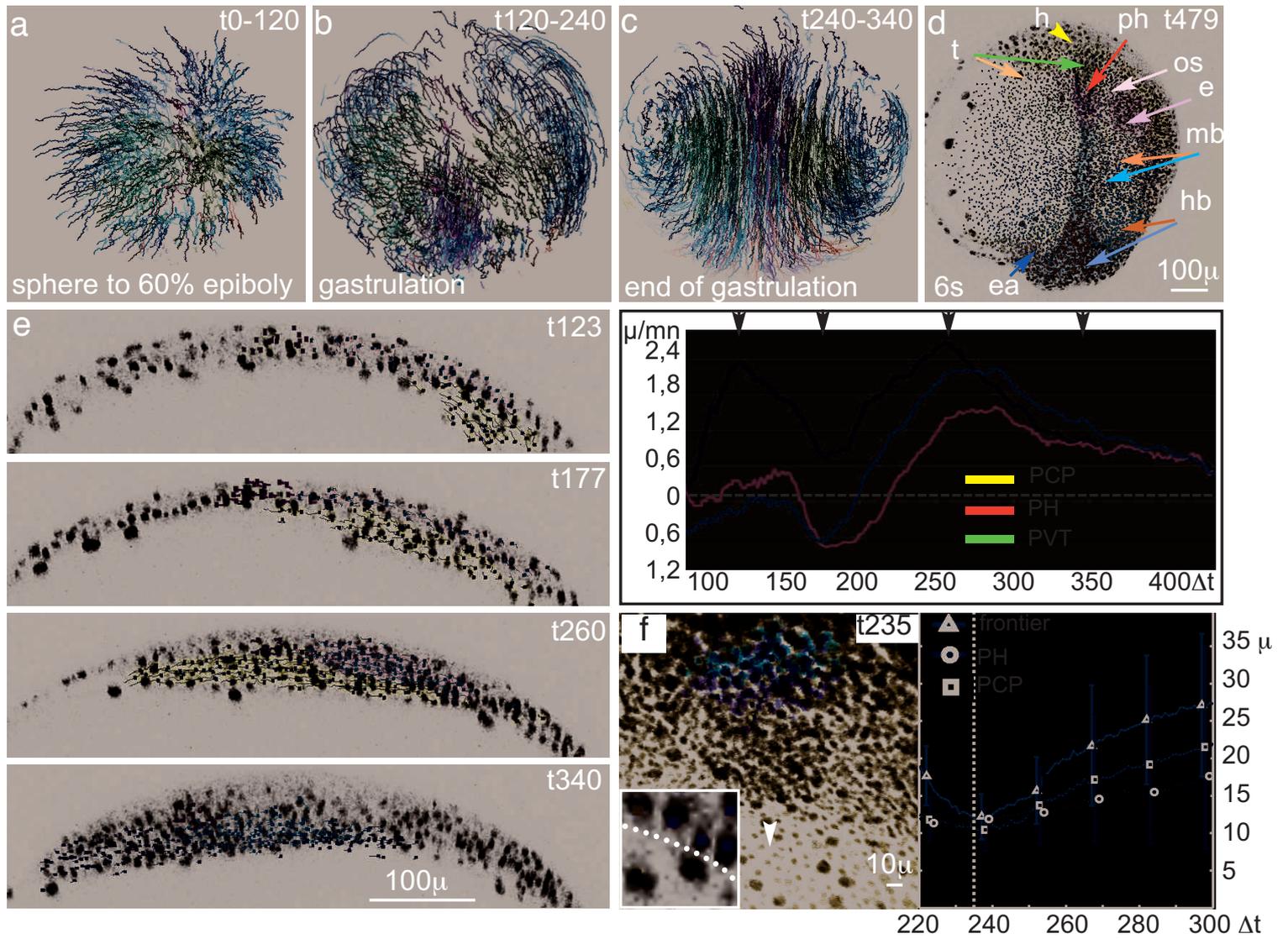


FIG. B.4 – Multi scale measurements of cell behaviours from their clonal history

B.2 Supplementary material

B.2.1 Table S1 Developmental table of the imaged embryo ID 070418a

Time step	Imaging duration	Total development time	Drift from the 28,5°C table	Equivalent at 28,5°C	Developmental stage
t0	0h00	4h10	Non detectable	4h00 and 4h20	Late sphere stage
t120	2h13	6h23	Non detectable	6h00 to 6h30	close to 60% epiboly
t320	5h57	10h07	+ 0h07	10h00	Tail bud
t360	6h39	10h49	+ 0h29	10h20	1 somite
t420	7h46	11h56	+ 0h36	11h20	3-4 somite
t480	8h53	13h03	+ 0h57	12h00	5-6 somite
t540	10h03	14h13	+ 1h13	13h00	7-8 somite

B.2.2 Movies

B.2.2.1 Movie 1-1

Data set 070418a, view from the animal pole at 4 hpf at 28.5°C (AP scheme more precisely designated as AP4 to indicate the onset of the imaged developmental period, imaged volume $700 \times 700 \times 140 \mu\text{m}^3$, $\Delta t = 67\text{s}$, 541 time steps), raw data with orthoslice z63, only the sections below the orthoslice are shown (Amira software visualisation).

B.2.2.2 Movie 1-2

Data set 071227b, view from the vegetal pole at the onset of gastrulation (6 hpf at 28.5°C, VP scheme, designated as VP6 to indicate the onset of the imaged developmental period, imaged volume $700 \times 700 \times 178 \mu\text{m}^3$, $\Delta t = 87\text{s}$, 413 time steps), raw data with orthoslice z45, only the sections below the orthoslice are shown (Amira software visualisation).

B.2.2.3 Movie 1-3

Data set 080322a, lateral imaging scheme, starting at 4 hpf (28.5°C, L scheme designated as L4 to indicate the onset of the imaged developmental period). In the L scheme, the position of the dorsal side identified a posteriori is indicated as 0 for a dorsal view, 6 for a ventral view and from 1 to 5 for intermediate positions, here a L4-4 scheme. The whole volume of raw data is shown (Amira visualisation software). For this data set, the light with a wavelength of 980 nm was provided by a mode-locked Ti:Sapphire laser (Mai Tai from Spectra-Physics). The pulse width was less than 100 fs.

B.2.2.4 Movie 1-4

Virtual embryo 070418a, with orthoslice as in Figure B.1i-l. Colour code from yellow, through grey, to blue indicating cell displacement orientation (projection of the speed vector on the y axis). Yellow indicates migration from caudal to rostral, blue indicates migration from rostral to caudal, grey indicates nul or medio-lateral displacement. Each nucleus centre is represented as a cube with a vector indicating its path for the next 12 time steps. Embryome Visualisation Platform.

B.2.2.5 Movie 1-5

Same as Movie 1-4 but with the whole volume to show the collective displacement of cells.

B.2.2.6 Movie 2-1

Movie 2-1 Animation of the filtering and centre detection process described in Figure B.2a-d. Frame 0, raw data ; frames 1-5, filtering steps ; frames 6-20, nucleus centre detection steps.

B.2.2.7 Movie 2-2

Data set 070420a at t001. 3D rendering of segmented nucleus and membrane shape for selected cells shown during mitosis. The virtual cells are in the environment of the raw data (green membranes, red nuclei). Transparency was introduced around the segmented cells. Note that the inner virtual cells that are under the epithelium (the enveloping layer, EVL) stand below the junction of EVL cells and rapidly move to align with the EVL junctions after cell division. Embryome Visualisation Platform.

B.2.2.8 Movie 3-1

Tracking of cell ID 124516142 (data set 070418a). Superimposed raw data and reconstructed data illustrating the augmented phenomenology concept. Here, nuclear centre of the tracked cell is superimposed with raw data orthoslices (membrane channel). Embryome Visualisation Platform.

B.2.2.9 Movie 4-1

Same as Figure B.4a-c. Cells manually selected in presumptive organs at t479 have been backtracked then allowing forward tracking visualisation, animal pole view. Colour code described in Figure B.4d. Cell trajectories are kept for the past 100 time steps. Embryome Visualisation Platform.

B.2.2.10 Movie 4-2

Same colour code as Movie 4-1, only nuclear centres (cubes) of selected cells are shown. Selected cells are backtracked from t480 to t1; at t1, a cell is tracked forward and shown with superimposed raw data orthoslices (nucleus channel) to assess the correctness of the lineage tree assigning its progeny to the optic vesicle (pink). Embryome Visualisation Platform.

B.2.2.11 Movie 4-3

Same selection of presumptive organs as Movie 4-1 and 4-2; animal pole view from the surface, visualisation of nucleus shape segmentation, cells that are not part of the selections are shown in grey. Embryome Visualisation Platform.

B.2.2.12 Movie 4-4

Relative movements of cells selected in the prechordal plate (PCP, yellow), presumptive hypothalamus (HP, red) and presumptive ventral telencephalon (PVT, green) as shown in Figure B.4e. Lateral view, rostral to the right. Augmented phenomenology visualisation with nuclear centres, cell trajectories and raw data orthoslices (nucleus channel). Embryome Visualisation Platform.

B.2.2.13 Movie 4-5

Same as Figure B.4f left panel. Small populations of PCP (yellow) and PH (red) cells where pairs of adjacent cells were chosen for distance measurements shown in Figure B.4f right panel. Augmented phenomenology visualisation with 3D rendering of raw data (nucleus channel) and nuclear centres throughout gastrulation. Last frames with raw data orthoslice nucleus channel, then membrane channel, to show the position of selected cells relative to morphological landmarks (marked frontier between PCP and PH at t316, end of gastrulation). Embryome Visualisation Platform.

Modèle du développement embryonnaire

C.1 Concept

Nous avons construit ce modèle sur le principe des ressorts en *plaçant* un ressort entre cellules voisines. Nous avons ensuite rajouté quelques règles supplémentaires, comme l'héritage par les cellules de l'information héritée de leur mère, selon les concepts des L-systems et la position de l'information modélisée par la distance des cellules au vitellus. Ces simples règles suffisent à modéliser un embryon virtuel aux mouvements globaux comparables à ceux observés dans l'embryon réel. Dans l'initialisation, une première cellule est positionnée sur le haut du vitellus, elle a pour rayon $350\mu m$. Ensuite elle se divise. Nous décrivons succinctement les quelques règles implémentées dans le modèle.

C.2 Définition d'une cellule

La cellule est assimilée à une sphère S de rayon membranaire r . Son volume correspondant est donc $V = \frac{4}{3}\pi r^3$. Nous associons à chaque cellule, une particule au centre de la sphère.

Soit la particule p_i associée à la cellule i définie par ses coordonnées $(x_i(t), y_i(t), z_i(t))$ à l'instant t .

Chaque particule a une masse propre m_i définie par la relation $m_i = \rho V_i$ où ρ correspond à la masse volumique cellulaire. Dans notre modèle, la gravité n'a pas d'effet sur la masse de la cellule, la masse intervient seulement lorsque des particules s'attirent ou s'opposent.

C.3 Membrane cellulaire

Par confort visuel, nous avons modélisé la membrane de chaque cellule. Les diagrammes de Voronoi peuvent être utilisés à cette fin, mais ils ne fonctionnent pas lorsqu'une cellule n'a pas de voisin sur un côté. Pour palier à cela nous avons approximé la membrane d'une cellule par la sphère S tronquée à ses contacts de voisinage.

C.4 Stochasticité

Nous avons rajouté un facteur de stochasticité δ_s au mouvement cellulaire. Cela consiste à décaler la position de la particule associée à la cellule d'un certain nombre de pixels dans une direction aléatoire. Ce facteur peut aussi être nul. On a donc pour p_i à t les coordonnées $(x_i(t), y_i(t), z_i(t))$ et ainsi à $t + 1$, si aucune force n'intervient dans le calcul de la position de la cellule les coordonnées deviennent $(x_i(t) \pm \delta_x, y_i(t) \pm \delta_y, z_i(t) \pm \delta_z)$ où $\delta_x, \delta_y, \delta_z \in [0, \delta_s]$ tirés de façon aléatoire.

C.5 Voisinage

Nous calculons à chaque pas de temps le voisinage cellulaire à partir des positions des particules par la triangulation de Delaunay. Ainsi il est possible d'avoir la liste des voisins associés à chaque cellule.

Chaque paire de voisin est relié par un ressort s . Ce ressort a une certaine raideur s_r et une taille de repos s_z . Soit c_1 et c_2 deux cellules ayant pour rayon respectif r_1 et r_2 et reliées par un ressort s_{12} . La distance du ressort s_{12} au repos est de $r_1 + r_2$. Ainsi par analogie au contact membranaire entre deux cellules, si aucune force extérieure ou intérieure aux cellules ne s'exerce, les cellules vont se maintenir de telle façon que les extrémités de leur sphère se touchent.

C.6 Vitellus

Le vitellus est assimilé à une sphère (non-déformable) de rayon $700\mu m$. Chaque cellule est reliée au vitellus par un ressort. Nous assimilons cette contrainte, au fait que le vitellus représente la source de nutrition des cellules, et donc plus elles en sont proches plus elles peuvent se nourrir facilement, plus elles s'en éloignent, plus le ressort les reliant au vitellus se tend.

C.7 Taille des cellules

Après chaque division, les deux cellules filles héritent chacune de la moitié du volume de la cellule mère (et donc de la moitié de la masse associée à la particule mère). Nous avons vu dans le chapitre 4 que la taille des cellules stagnait autour de $12\mu m$ après 8 heures de développement. De façon similaire dans notre modèle, quand les cellules filles ont un rayon inférieur à $12\mu m$ alors elles grossissent légèrement jusqu'à l'atteindre.

C.8 Couche cellulaire

Nous avons introduit la notion de couche cellulaire comme étudiée dans le chapitre 4. A chaque instant, en fonction de la position dans l'espace de la cellule

i , nous calculons le nombre de cellules interposées entre elle et le vitellus. Cela représente la couche L_i de la cellule i .

C.9 Mouvement cellulaire

La position de chacune des cellules est enregistré au cours du temps. Il est donc possible à tout moment de calculer la vitesse moyenne de déplacement de chaque cellule ou d'un groupe de cellules. Soit $\vec{V}_i(t, \Delta t)$ la vitesse moyenne de déplacement de la cellule i à l'instant t moyenné sur Δt pas de temps. Nous avons rajouté un paramètre de viscosité, ayant pour conséquence que les cellules entraînent ou freinent leurs voisines dans leur mouvement.

C.10 Division cellulaire

Nous pouvons contrôler à l'aide de ce modèle l'ensemble des forces exercées sur une cellule en division, et ainsi prévoir son orientation. Quand une cellule se divise, elle se transforme en deux cellules filles, qui héritent des informations de la cellule mère. Outre l'information de la masse que l'on a vu dans C.7, les cellules filles héritent dans ce modèle de l'information du temps de vie de la mère. Le temps de vie de chacune des filles et donc l'instant où elles se diviseront, est calculé à chaque instant. Soit τ_i le temps de vie de la cellule i . On a $\tau_j = \tau_i + L_j + \lambda_d$ où j est une fille issue de la division de la cellule i et le paramètre λ_d est une constante qui permet d'allonger le temps de vie des cellules au cours de l'embryogenèse.

A partir de la direction du vecteur de mouvement \vec{V}_i , nous pouvons décider l'orientation du plan de division. Ainsi selon les mesures effectuées au chapitre 4, nous faisons diviser la cellule dans le plan perpendiculaire à son mouvement.

Nomenclature

μ	Micron
3D	Trois Dimensions
ADN	Acide DésoxyriboNucléique
AFS	Andrew File System
AML	Active Machine Learning
AP	Antéro-Postérieur
ARN	Acide RiboNucléique
BPM	Business Process Management
BSP	Bilateral Symetry Plan
CLSM	Confocal Laser Scanning Microscopy
CRNS	Centre National de la Recherche Scientifique
DAPI	Di Aminido Phenyl Indol
DB	Base de Donnée
DV	Dorso-Ventral
EGEE	Enabling Grids for E-science
EMP	Espérance Maximisation
EVL	Epithélium
FITC	Fluoresceine Iso Thio Cyanate
GFP	Green Fluorescent Protein
GO	Giga-Octets
GRC	Cellule Ganglionnaire Rétinienne
HMN	Hidden Markov Models
hpf	Heures Post Fécondation
IN2P3	Institut Dational De Physique nucléaire et de Physique des particules
ITK	Insight Segmentation and Registration Toolkit
KNN	K-Nearest Neighbor algorithm
LAL	Laboratoire de l'Accélérateur Linéaire

MDV	Maximim Distance Value
ML	Médio-Latéral
MLSM	Multiphoton Laser Scanning Microscopy
MPI	Message Passing Interface
PDE	Partial Differential Equation
SC	Systèmes Complexes
SOA	Service Oriented Architectures
SPIM	Microscopie par Sélection du Plan Illuminé
SPS	Semi-Permanent Storage
SRB	Storage Resource Broker
STM	Segment Time of Mitosis
SV	Syncytium Vitellin
SVM	Machine à Vecteurs de Support ou Séparateurs à Vaste Marge
TMB	Transition Mi-Blastuléenne
TO	Tera-Octets
TRITC	Tetra methyl Rhodamine Iso Thio Cyanate
UMPC	Université Pierre et Marie Curie
VTK	Visualizaton Toolkit
WFMC	WorkFow Management Coalition

Bibliographie

- Adams, Richard. 1996. Metaphase Spindles Rotate in the Neuroepithelium of Rat Cerebral Cortex. *Journal of Neuroscience*, **16**(23), 7610. 23, 109
- Ahringer, J. 2003. Control of cell polarity and mitotic spindle positioning in animal cells. *Current Opinion in Cell Biology*, **15**(1), 73–81. 23, 25, 108
- Aizerman, M A, Braverman, E M, & Rozonoer, L I. 1964. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*. Jan. 44
- Al-Kofahi, Omar, Radke, Richard J, Goderie, Susan K, Shen, Qin, Temple, Sally, & Roysam, Badrinath. 2006. Automated cell lineage construction : a rapid method to analyze clonal development established with murine neural progenitor cells. *Cell Cycle*, **5**(3), 327–35. 48, 49
- Alvarez, L, Guichard, F, Lions, P, & Morel, J. 1993. Axioms and fundamental equations of image processing. *Archive for Rational Mechanics and Analysis*, Jan. 165
- Amatruda, J, Shepard, J, Stern, H, & Zon, L. 2002. Zebrafish as a cancer model system. *Cancer Cell*, Jan. 3
- Bahnon, Alfred, Athanassiou, Charalambos, Koebler, Douglas, Qian, Lei, Shun, Tongying, Shields, Donna, Yu, Hui, Wang, Hong, Goff, Julie, Cheng, Tao, Houck, Raymond, & Cowsert, Lex. 2005. Automated measurement of cell motility and proliferation. *BMC Cell Biol*, **6**(1), 19. 48
- Bao, Zhirong, Murray, John I, Boyle, Thomas, Ooi, Siew Loon, Sandel, Matthew J, & Waterston, Robert H. 2006. Automated cell lineage tracing in *Caenorhabditis elegans*. *Proc Natl Acad Sci USA*, **103**(8), 2707–12. 164, 168
- Bassingthwaighte, J. 2009. Physiome Project. <http://www.physiome.org/>. 164
- Belousov, L, Opitz, J, & Gilbert, S. 1997. Life of Alexander G. Gurwitsch and his relevant contribution to the theory of morphogenetic fields. *International Journal of Developmental Biology*, Jan. 17
- Bennett, K, & Demiriz, A. 1999. Semi-supervised support vector machines. *Advances in Neural Information Processing Systems*, Jan. 61
- Boland, M V, & Murphy, R F. 2001. A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of HeLa cells. *Bioinformatics*, **17**(12), 1213–23. 46
- Boland, M V, Markey, M K, & Murphy, R F. 1998. Automated recognition of patterns characteristic of subcellular structures in fluorescence microscopy images. *Cytometry*, **33**(3), 366–75. 46, 55

- Bolker, J. 2000. Modularity in Development and Why It Matters to Evo-Devo 1. *Integrative and Comparative Biology*, Jan. 17
- Boser, B, Guyon, I, & Vapnik, V. 1992. A training algorithm for optimal margin classifiers. *Proceedings of the fifth annual workshop on Computational learning theory*, Jan. 43, 44
- Boutet, J., Calvez, S Le, & Ulfendahl, M. 2001. Image restoration for confocal microscopy : improving the limits of deconvolution, with application to the visualization of the mammalian hearing organ. *Biophys J*, 80(5), 2455–70. 142
- Bowring, James, Rehg, James, & Harrold, Mary. 2004. Active learning for automatic classification of software behavior. *ISSTA '04 : Proceedings of the 2004 ACM SIGSOFT international symposium on Software testing and analysis*, Jul. 56
- Box, G. 1987. Author Empirical model-building and response surfaces/George EP Box, Norman R. Draper. *New York : Wiley*, Jan. 42
- Bunyak, F, Palaniappan, K, Nath, S, & Baskin, T. 2006. Quantitative cell motility for in vitro wound healing using level set-based active contour tracking. *3rd IEEE International Symposium on Biomedical Imaging : Nano to Macro*, Jan. 50
- Cedilnik, A, Baumes, J, Ibanez, L, Megason, S, & Wylie, B. 2007. Integration of information and volume visualization for analysis of cell lineage and gene expression during Embryogenesis. *Proc. SPIE*, Jan. 48
- Chalfie, M, Tu, Y, Euskirchen, G, Ward, W, & Prasher, D. 1994. Green fluorescent protein as a marker for gene expression. *Science*, Jan. 26
- Chan, J, Bayliss, P, Wood, J, & Roberts, T. 2002. Dissection of angiogenic signaling in zebrafish using a chemical genetic approach. *Cancer Cell*, Jan. 3
- Chen, Shann-Ching, & Murphy, Robert F. 2006. A graphical model approach to automated classification of protein subcellular location patterns in multi-cell images. *BMC Bioinformatics*, 7(Jan), 90. 46, 55
- Chen, Xiaowei, Zhou, Xiaobo, & Wong, Stephen T C. 2006. Automated segmentation, classification, and tracking of cancer cell nuclei in time-lapse microscopy. *IEEE transactions on bio-medical engineering*, 53(4), 762–6. 55
- Chen, Y, Vemuri, B, & Wang, L. 2000. Image denoising and segmentation via non-linear diffusion. *Computers and Mathematics with applications*, Jan. 170
- Cheng, Y. 1995. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Jan. 49
- Chuang, T, Moshir, S, Garini, Y, & Chuang, A. 2004. The three-dimensional organization of telomeres in the nucleus of mammalian cells. *BMC Biol*, Jan. 142

- Chudakov, D, Lukyanov, S, & Lukyanov, K. 2005. Fluorescent proteins as a toolkit for in vivo imaging. *Trends in Biotechnology*, Jan. 26
- Coalition, Workflow Management. 1995. The Workflow Reference Model. 39
- Cohn, D, Ghahramani, Z, & Jordan, M. 1996. Active learning with statistical models. *Arxiv preprint cs.AI*, Jan. 42
- Cointet, J, Faure, E, & Roth, C. 2007. Intertemporal topic correlations in online media. *Proc. ICWSM Intl Conf Weblogs Social Media. Boulder*, Jan. 133
- Concha, M L, & Adams, R J. 1998. Oriented cell divisions and cellular morphogenesis in the zebrafish gastrula and neurula : a time-lapse analysis. *Development*, 125(6), 983–94. 24, 109
- Cooper, M, D'Amico, L, & Henry, C. 1999. Confocal microscopic analysis of morphogenetic movements. *Methods in cell biology*, Jan. 2
- Corsaro, S, Mikula, K, Sarti, A, & Sgallari, F. 2007. Semi-implicit covolume method in 3D image segmentation. *SIAM Journal on Scientific Computing*, Jan. 166
- Costello, D. 1961. On the orientation of centrioles in dividing cells, and its significance : a new contribution to spindle mechanics. *The Biological Bulletin*, Jan. 23
- Dagan, I, & Engelson, S. 1995. Committee-based sampling for training probabilistic classifiers. *Machine Learning -Internation Workshop Then Conference-*, Jan. 42
- Davidson, E H. 1993. Later embryogenesis : regulatory circuitry in morphogenetic fields. *Development*, 118(3), 665–90. 17, 57
- Debeir, O, Ham, P Van, Kiss, R, & Decaestecker, C. 2005. Tracking of migrating cells under phase-contrast video microscopy with combined mean-shift processes. *IEEE transactions on medical imaging*, 24(6), 697–711. 49
- Delaunay, B. 1934. Sur la sphère vide : *Izvestia Akademii Nauk SSSR. Otdelenie Matematicheskikh i Estestvennykh Nauk*, 7, 793–800.
- Dempster, A, Laird, N, & Rubin, D. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society.*, Jan. 52, 76
- Denk, W, & Svoboda, K. 1997. Photon upmanship : why multiphoton imaging is more than a gimmick. *Neuron*, 18(3), 351–7. 30
- Denk, W, Strickler, J, & Webb, W. 1990. Two-photon laser scanning fluorescence microscopy. *Science*, Jan. 30
- Dooley, K, & Zon, LI. 2000. Zebrafish : a model system for the study of human disease. *Current opinion in genetics & development*, 10(3), 252–256. 7

- Doucet, A, & Ristic, B. 2002. Recursive state estimation for multiple switching models with unknown transition probabilities. *IEEE Transactions on Aerospace and Electronic Systems*, Jan. 49
- Drblikova, O. 2005. Finite Volume Scheme for Tensor Anisotropic Diffusion in Image Processing. *Proc. MAGIA*, Jan. 57, 142
- Drblikova, O. 2008. Existence and uniqueness of the solution to diamond-cell finite volume scheme for 3D nonlinear tensor diffusion. *math.sk.* 142
- Driever, W, & Nüsslein-Volhard, C. 1988. The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. *Cell*, 54(1), 95–104. 38
- Driever, W, Stemple, D, Schier, A, & Solnica-Krezel, L. 1999. Zebrafish : genetic tools for studying vertebrate development. *Cell*, Jan. 6
- Duda, R, & Hart, P. 1972. Use of the Hough transformation to detect lines and curves in pictures. *Communications of the ACM*, Jan. 57, 142
- Dufour, A, Shinin, V, Tajbakhsh, S, & Guillen-Aghion, N. 2005. Segmenting and tracking fluorescent cells in dynamic 3-D microscopy with coupled active surfaces. *IEEE Transactions on Image Processing*, Jan. 49, 50
- Dumais, S, Platt, J, Heckerman, D, & Sahami, M. 1998. Inductive learning algorithms and representations for text categorization. *Proceedings of the seventh international conference on Information and knowledge management*, Jan. 42, 43
- Edelsbrunner, H, Kirkpatrick, D, & Seidel, R. 1983. On the shape of a set of points in the plane. *IEEE Transactions on Information Theory*, Jan. 135
- England, Samantha J, Blanchard, Guy B, Mahadevan, L, & Adams, Richard J. 2006. A dynamic fate map of the forebrain shows how vertebrate eyes form and explains two causes of cyclopia. *Development*, 133(23), 4613–7. 14, 109, 164
- Ephrussi, Anne, & Johnston, Daniel St. 2004. Seeing is believing : the bicoid morphogen gradient matures. *Cell*, 116(2), 143–52. 38
- Faure, E., Lombardot, B., Luengo-Oroz, M., Campana, M., Peyri eras, N., Doursat, R., & Bourguine, P. 2007. Active machine learning for embryogenesis. *4th European Conference on Complex Systems, Technische Universit at Dresden, Germany*, October. v
- Fern andez-Gonz alez, R, & Mu oz-Barrutia, A. 2006. Quantitative in vivo microscopy : the return from the ‘omics’. *Current opinion in biotechnology*, Jan. 2
- Flemming, W. 1878. Zur Kenntnis der Zelle und ihrer Teilung-Erscheinungen. *Schr Naturwiss Ver Schlesw-Holst.* 19

- Freund, Y, Seung, H, Shamir, E, & Tishby, N. 1997. Selective sampling using the query by committee algorithm. *Machine Learning*, Jan. 42
- Fricker, M, Runions, J, & Moore, I. 2006. Quantitative fluorescence microscopy : from art to science. *Annual Reviews*, Jan. 28
- Frolkovic, P., Mikula, K., Peyrieras, N., & Sarti, A. 2007. Counting Number of Cells and Cell Segmentation Using Advection-Diffusion Equations. *KYBERNETIKA-PRAHA*. 142, 166, 170
- Gallardo, G, Yang, F, Ianzini, F, Mackey, M, & Sonka, M. 2004. Mitotic cell recognition with hidden Markov models. *Proceedings of SPIE*, Jan. 46, 55
- Gardner, R L. 1985. Single Cell Marking and Cell Lineage in Animal Development. Jan. 18
- Geldmacher-Voss, Benedikt, Reugels, Alexander M, Pauls, Stefan, & Campos-Ortega, José A. 2003. A 90-degree rotation of the mitotic spindle changes the orientation of mitoses of zebrafish neuroepithelial cells. *Development*, 130(16), 3767–80. 25, 109
- Gilbert, S, & Raunio, A. 1997a. Embryology : Constructing the organism. *Sunderland*, Jan. 18
- Gilbert, S F, Opitz, J M, & Raff, R A. 1996. Resynthesizing evolutionary and developmental biology. *Developmental Biology*, 173(2), 357–72. 17
- Gilbert, Scott F., & Raunio, Anne M. 1997b. Embryology : Constructing the Organism . Jan, 537. 15
- Godinez, W J, Lampe, M, Wörz, S, Müller, B, Eils, R, & Rohr, K. 2009. Deterministic and probabilistic approaches for tracking virus particles in time-lapse fluorescence microscopy image sequences. *Medical image analysis*, 13(2), 325–42. 49
- Goessling, W, North, TE, & Zon, LI. 2007. New waves of discovery : modeling cancer in zebrafish. *Journal of Clinical Oncology*, 25(17), 2473. 7
- Goldberg, A, Allis, C, & Bernstein, E. 2007. Epigenetics : a landscape takes shape. *Cell*, Jan. 3
- Goldsmith, P. 2004. Zebrafish as a pharmacological tool : the how, why and when. *Current Opinion in Pharmacology*, 4(5), 504–512. 4
- Goldstein, A M, & Fishman, M C. 1998. Notochord regulates cardiac lineage in zebrafish embryos. *Developmental Biology*, 201(2), 247–52. 50
- Gong, Ying, Mo, Chunhui, & Fraser, Scott E. 2004. Planar cell polarity signaling controls cell division orientation during zebrafish gastrulation. *Nature*, 430(7000), 689–93. 23

- Goppert-Mayer, M. 1931. Uber Elementarakte mit zwei quantensprungen. *Annalen der Physik*, Jan. 30
- Gregoire, T R. 2001. Leaves and fishes : who's in charge ? *Journal of dental technology : the peer-reviewed publication of the National Association of Dental Laboratories*, 18(2), 7–8. 34
- Harder, Nathalie, Mora-Bermúdez, Felipe, Godinez, William J, Ellenberg, Jan, Eils, Roland, & Rohr, Karl. 2006. Automated analysis of the mitotic phases of human cells in 3D fluorescence microscopy image sequences. *Medical image computing and computer-assisted intervention : MICCAI International Conference on Medical Image Computing and Computer-Assisted Intervention*, 9(Pt 1), 840–8. 46
- Hatta, Kohei, Tsujii, Hitomi, & Omura, Tomomi. 2006. Cell tracking using a photoconvertible fluorescent protein. *Nat Protoc*, 1(2), 960–7. 50
- Helmchen, Fritjof, & Denk, Winfried. 2005. Deep tissue two-photon microscopy. *Nat Methods*, 2(12), 932–40. 165
- Hertzler, P. 1992. Cleavage and gastrulation in the shrimp *Sicyonia ingentis* : invagination is accompanied by oriented cell division. *Development*, Jan. 23
- Ho, R. 1992. Cell movements and cell fate during zebrafish gastrulation. *Development*, Jan. 10
- Horn, B, & Schunck, B. 1981. Determining Optical Flow. *Artificial Intelligence*, Jan. 57
- Horvitz, E, & Rutledge, G. *Proceedings of the Seventh Conference on Uncertainty*. 42
- Huisken, Jan, Swoger, Jim, Bene, Filippo Del, Wittbrodt, Joachim, & Stelzer, Ernst H K. 2004. Optical sectioning deep inside live embryos by selective plane illumination microscopy. *Science*, 305(5686), 1007–9. 50
- Hunter, P, Smith, N, Fernandez, J, & Tawhai, M. 2005. Integration from proteins to organs : the IUPS Physiome Project. *Mechanisms of ageing and development*, Jan. 2
- Jaeger, Johannes, & Reinitz, John. 2006. On the dynamic nature of positional information. *Bioessays*, 28(11), 1102–11. 38
- Joachims, T. 1999. Text categorization with support vector machines. *eldorado.tu-dortmund.de*, Jan. 43
- Joachims, T, Nedellec, C, & Rouveirol, C. 1998. Text categorization with support vector machines : learning with many relevant. *Machine Learning : ECML-98 10th European Conference on Machine Learning*, Jan. 42
- Kaelbling, L, Littman, M, & Moore, A. 1996. Reinforcement learning : A survey. *Arxiv preprint cs.AI*, Jan. 42

- Kane, D, Warga, R, & Kimmel, C. 1992. Mitotic domains in the early embryo of the zebrafish. *Nature*, Jan. 166
- Kane, D A. 1999. Cell cycles and development in the embryonic zebrafish. *Methods in cell biology*, 59(Jan), 11–26. 108, 164
- Kane, D A, & Kimmel, C B. 1993. The zebrafish midblastula transition. *Development*, 119(2), 447–56. 9
- Karlstrom, R. 1996. A flipbook of zebrafish embryogenesis. *Development*, Jan. 7
- Kearns, M, & Koller, D. 1999. Efficient reinforcement learning in factored MDPs. *International Joint Conference on Artificial Intelligence*, Jan. 43
- Keller, P, Schmidt, A, Wittbrodt, J, & Stelzer, E. 2008. Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. *Science*, Jan. 50, 51
- Keller, Ray. 2002. Shaping the vertebrate body plan by polarized embryonic cell movements. *Science*, 298(5600), 1950–4. 18
- Kimmel, C. 1994. Cell cycles and clonal strings during formation of the zebrafish central nervous system. *Development*, Jan. 23, 167
- Kimmel, C B, Ballard, W W, Kimmel, S R, Ullmann, B, & Schilling, T F. 1995. Stages of embryonic development of the zebrafish. *Dev Dyn*, 203(3), 253–310. 7, 8, 9, 11, 12, 13, 23, 57, 109, 164, 169
- Kimmel, CB, Warga, RM, & Schilling, TF. 1990. Origin and organization of the zebrafish fate map. *Development*, 108(4), 581. 10, 14
- Kirkpatrick, S, Gelatt, C, & Vecchi, M. 1983. Optimization by simulated annealing. *Science*, Jan. 166, 171
- Kirubarajan, T, Bar-Shalom, Y, & Pattipati, K. 2001. Multiassignment for tracking a large number of overlapping objects. *IEEE Transactions on Aerospace and Electronic Systems*, Jan. 48
- Kitano, Hiroaki. 2002. Computational systems biology. *Nature*, 420(6912), 206–10. 37
- Klonis, N. 2002. Fluorescence photobleaching analysis for the study of cellular dynamics. *European Biophysics Journal*, Jan. 27
- Latombe, J. 1991. Robot motion planning. *Kluwer International Series in Engineering and Computer Science*, Jan. 42
- Lecuit, T. 2008. Developmental mechanics : cellular patterns controlled by adhesion, cortical tension and cell division. *link.aip.org*, Jan. 109

- LeCun, Y, Jackel, L, Bottou, L, Brunot, A, & Cortes, C. 1995. Comparison of learning algorithms for handwritten digit recognition. *International Conference on Artificial Neural Networks*, Jan. 43
- Lewis, D, & Catlett, J. 1994. Heterogeneous uncertainty sampling for supervised learning. *Proceedings of the Eleventh International Conference on Machine Learning*, Jan. 42
- Lewis, D, & Gale, W. 1994. A sequential algorithm for training text classifiers. *Proceedings of the 17th annual international ACM SIGIR*, Jan. 41, 42
- Li, Kang, Miller, Eric D, Chen, Mei, Kanade, Takeo, Weiss, Lee E, & Campbell, Phil G. 2008. Cell population tracking and lineage construction with spatiotemporal context. *Medical image analysis*, 12(5), 546–66. 50
- Li, Ming, & Vitányi, P M. B. 1997. An introduction to Kolmogorov complexity and its applications. Jan, 637. 33
- Lichtman, Jeff W, & Conchello, José-Angel. 2005. Fluorescence microscopy. *Nat Methods*, 2(12), 910–9. 28
- Lin, G, Chawla, M, Olson, K, Guzowski, J, & Barnes, C. 2005. Hierarchical, model-based merging of multiple fragments for improved three-dimensional segmentation. *Cytometry Part A*, Jan. 145
- Lindenmayer, A. 1968. Mathematical models for cellular interactions in development. I. Filaments with one-sided inputs. *Journal of Theoretical Biology*, Jan. 37
- Lippincott-Schwartz, J, & Patterson, G. 2003. Development and use of fluorescent protein markers in living cells. *Science*, Jan. 26
- Liu, T, Li, G, Nie, J, Tarokh, A, Zhou, X, & Guo, L. 2008. An automated method for cell detection in zebrafish. *Neuroinformatics*, Jan. 142
- Lombardot, B., Luengo-Oroz, M.A., Melani, C., Faure, E., Santos, A., Peyriéras, N., Ledesma-Carbayo, M., & Bourgine, P. 2008. Evaluation of four 3D non rigid registration methods applied to early zebrafish development sequences. *Proceedings MICAI08 : Microscopic image analysis with application in biology*, September. 145
- Luengo-Oroz, M, Faure, E, Lombardot, B, & Sance, R. 2007. Twister Segment Morphological Filtering. A New Method for Live Zebrafish Embryos Confocal Images. *IEEE ICIP*, Jan. 142
- Luengo-Oroz, M, Duloquin, L, Castro, C, Savy, T, Faure, E, Lombardot, B, Bourgine, P, Peyriéras, N, & Santos, A. 2008. Can voronoi diagram model cell geometries in early sea-urchin embryogenesis? *Biomedical Imaging : From Nano to Macro*. 146

- Malpica, N, de Solorzano, C, Vaquero, J, & Santos, A. 1997. Applying watershed algorithms to the segmentation of clustered nuclei. *Cytometry*, Jan. 142
- Malumbres, M, & Barbacid, M. 2001. To cycle or not to cycle : a critical decision in cancer. *Nature Reviews Cancer*, 1(3), 222–31. 19
- Mathis, L, & Nicolas, J F. 2000. Different clonal dispersion in the rostral and caudal mouse central nervous system. *Development*, 127(6), 1277–90. 19
- Mathis, L, Sieur, J, Voiculescu, O, Charnay, P, & Nicolas, J F. 1999. Successive patterns of clonal cell dispersion in relation to neuromeric subdivision in the mouse neuroepithelium. *Development*, 126(18), 4095–106. 19
- McCallum, A. 1998. Employing EM in pool-based active learning for text classification. *Proceedings of ICML-98*, Jan. 42
- McLachlan, G, Krishnan, T, & InterScience, W. 1997. The EM algorithm and extensions. *lavoisier.fr*, Jan. 76
- Megason, S, & Fraser, S. 2003. Digitizing life at the level of the cell : high-performance laser-scanning microscopy and image analysis for in toto imaging of development. *Mech Dev*, Jan. 2, 40, 54
- Meijer, Laurent, Jézéquel, Armelle, & Roberge, Michel. 2003. Progress in Cell Cycle Research. Cell Cycle Regulators as Therapeutic Targers. Jan. 19
- Meijering, E, Jacob, M, Sarria, J, & Steiner, P. 2004. Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images. *Cytometry Part A*, Jan. 145
- Meijering, E, Smal, I, & Danuser, G. 2006. Tracking in molecular bioimaging. *IEEE signal processing magazine*, Jan. 48
- Melani, C, Peyrieras, N, Mikula, K, & Zanella, C. 2007. Cells tracking in a live zebrafish embryo. *Engineering in Medicine and Biology Society*, Jan. 51, 57, 109, 121, 149, 166
- Metscher, B, & Ahlberg, P. 1999. Zebrafish in context : uses of a laboratory model in comparative studies. *Developmental Biology*, Jan. 6
- Meyer, F. 2001. An overview of morphological segmentation. *International Journal of Pattern Recognition and Artificial Intelligence*, Jan. 145
- Mikula, K., & Sarti, A. 2007. Parametric and Geometric Deformable Models : An application in Biomaterials and Medical Imagery. *Springer Publishers*, ed. Farag, E. J. S. S. a. A., 123–160. 166

- Mikula, K, Peyri ras, N, Reme řkov, M, & Sarti, A. 2008a. 3D embryogenesis image segmentation by the generalized subjective surface method using the finite volume. *Proceedings of FVCA5–5th International Symposium on Finite Volume*, Jan. 60
- Mikula, K., Peyrieras, N., Remesikova, M., & Sarti, A. 2008b. Finite Volumes for Complex Applications V. ed. J.M.Herard, R. E. a., **Wiley**, 585–592. 166, 170
- Minsky, M. 1961. MINSKY. *Pattent*. 28
- Miyawaki, A, Sawano, A, & Kogure, T. 2003. Lighting up cells : labelling proteins with fluorophores. *IMAGING*, Jan. 27
- Moore, A. W., Schneider, J. G., Boyan, J. A., & Lee, M. S. 1998. Q2 : Memory-based active learning for optimizing noisy continuous functions. *Proceedings of the Fifteenth International Conference on Machine Learning. Morgan Kaufmann*. 42
- Moriyoshi, K, Richards, L J, Akazawa, C, O’Leary, D D, & Nakanishi, S. 1996. Labeling neural cells using adenoviral gene transfer of membrane-targeted GFP. *Neuron*, 16(2), 255–60. 165, 169
- Mukherjee, Dipti Prasad, Ray, Nilanjan, & Acton, Scott T. 2004. Level set analysis for leukocyte detection and tracking. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 13(4), 562–72. 49, 50
- Murphy, R, Velliste, M, & Porreca, G. 2003. Robust numerical features for description and classification of subcellular location patterns in fluorescence microscope images. *The Journal of VLSI Signal Processing*, Jan. 46
- Myers, Dina C, Sepich, Diane S, & Solnica-Krezel, Lilianna. 2002. Convergence and extension in vertebrate gastrulae : cell movements according to or in search of identity? *Trends Genet*, 18(9), 447–55. 10, 18
- Nakajima, C., Itoh, N., Pontil, M., & Poggio, T. 2000. Object Recognition and Detection by a Combination of Support Vector Machine and Rotation Invariant Phase Only Correlation. *Pattern Recognition, International Conference on*, 4, 4787. 43
- Nicolas, J F, Mathis, L, Bonnerot, C, & Saurin, W. 1996. Evidence in the mouse for self-renewing stem cells in the formation of a segmented longitudinal structure, the myotome. *Development*, 122(9), 2933–46. 19
- Nusslein-Volhard, C, & Wieschaus, E. 1980. Mutations affecting segment number and polarity in *Drosophila*. *Nature*, 287(5785), 795–801. 38
- Olivo-Marin, J. 2002. Extraction of spots in biological images using multiscale products. *Pattern Recognition*, Jan. 142
- Osher, S, & Sethian, J. 1988. Fronts propagating with curvature dependent speed : algorithms based on Hamilton-Jacobi formulations. *Journal of computational physics*, Jan. 50

- Otsu, N. 1975. A Threshold Selection Method from Gray-Level Histograms. *Automatica*, Jan. 142
- Padfield, D, Rittscher, J, Sebastian, T, & Thomas, N. 2006. Spatio-temporal cell cycle analysis using 3D level set segmentation of unstained nuclei in line scan confocal fluorescence images. *3rd IEEE International Symposium on Biomedical Imaging : Nano to Macro*, Jan. 49, 50
- Padfield, D, Rittscher, J, & Roysam, B. 2008. Spatio-temporal cell segmentation and tracking for automated screening. *5th IEEE International Symposium on Biomedical Imaging : From Nano to Macro*, Jan. 49, 50
- Padfield, D, Rittscher, J, Thomas, N, & Roysam, B. 2009. Spatio-temporal cell cycle phase analysis using level sets and fast marching methods. *Medical image analysis*, Jan. 49
- Papan, C, & Campos-Ortega, J. 1994. On the formation of the neural keel and neural tube in the zebrafish *Danio (Brachydanio) rerio*. *Development Genes and Evolution*, Jan. 24
- Parasassi, T, Yu, W, Durbin, D, Kuriashkina, L, Gratton, E, Maeda, N, & Ursini, F. 2000. Two-photon microscopy of aorta fibers shows proteolysis induced by LDL hydroperoxides. *Free Radic Biol Med*, 28(11), 1589–97. 30
- Patterson, G, & Lippincott-Schwartz, J. 2002. A photoactivatable GFP for selective photolabeling of proteins and cells. *Science*, Jan. 27
- Pearl, J. 2000. Causality : Models, reasoning, and inference. Jan. 41
- Peng, H. 2008. Bioimage informatics : a new area of engineering biology. *Bioinformatics*, 24(17), 1827. 40
- Periasamy, A, Skoglund, P, Noakes, C, & Keller, R. 1999. An evaluation of two-photon excitation versus confocal and digital deconvolution fluorescence microscopy imaging in *Xenopus* morphogenesis. *Microscopy research and technique*, Jan. 30
- Perona, P, & Malik, J. 1990. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Learning*, Jan. 57, 140
- Petit, A, Legué, E, & Nicolas, J. 2005. Methods in clonal analysis and applications. *Reprod. Nutr. Dev*, Jan. 168
- Pickart, Michael A, Sivasubbu, Sridhar, Nielsen, Aubrey L, Shriram, Sabitha, King, Richard A, & Ekker, Stephen C. 2004. Functional genomics tools for the analysis of zebrafish pigment. *Pigment Cell Res*, 17(5), 461–70. 7
- Poggi, Lucia, Vitorino, Marta, Masai, Ichiro, & Harris, William A. 2005. Influences on neural lineage and mode of division in the zebrafish retina in vivo. *The Journal of Cell Biology*, 171(6), 991–9. 25, 109

- Potapova, Tamara A, Daum, John R, Pittman, Bradley D, Hudson, Joanna R, Jones, Tara N, Satinover, David L, Stukenberg, P Todd, & Gorbsky, Gary J. 2006. The reversibility of mitotic exit in vertebrate cells. *Nature*, **440**(7086), 954–8. 20
- Prasher, D, Eckenrode, V, Wade, W, & Prendergast, F. 1992. Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene(Amsterdam)*, Jan. 26
- Psychoyos, D, & Stern, C D. 1996. Fates and migratory routes of primitive streak cells in the chick embryo. *Development*, **122**(5), 1523–34. 18
- Raff, Rudolf A. 1996. The Shape of Life : genes,development,& the evolution of animal form. Jan, 520. 18
- Rizzi, B, Campana, M, Zanella, C, & Melani, C. 2007a. 3-D Zebrafish Embryo Image Filtering by Nonlinear Partial Differential Equations. *Engineering in Medicine and Biology Society*, Jan. 142
- Rizzi, B, Campana, M, Zanella, C, Melani, C, Bourguine, P, Mikula, K, Peyrieras, N, & Sarti, A. 2007b. 3-D Zebrafish Embryo Image Filtering by Nonlinear Partial Differential Equations. *math.sk.* 146, 165
- Rizzi, B, Melani, C, Campana, M, & Bourguine, P. 2007c. Segmentation of Cells from 3-D Confocal Images of Live Zebrafish Embryo. *Engineering in Medicine and Biology Society*, Jan. 146
- Robertis, E M De, Morita, E A, & Cho, K W. 1991. Gradient fields and homeobox genes. *Development*, **112**(3), 669–78. 17
- Roth, S, Stein, D, & Nüsslein-Volhard, C. 1989. A gradient of nuclear localization of the dorsal protein determines dorsoventral pattern in the *Drosophila* embryo. *Cell*, **59**(6), 1189–1202. 38
- Saga, Yumiko, & Takeda, Hiroyuki. 2001. The making of the somite : molecular events in vertebrate segmentation. *Nature Reviews Genetics*, **2**(11), 835. 10
- Sarti, A, Malladi, R, & Sethian, J. 2000. Subjective surfaces : A method for completing missing boundaries. *Proceedings of the National Academy of Sciences*, Jan. 57, 146, 166, 170
- Sausedo, R A, Smith, J L, & Schoenwolf, G C. 1997. Role of nonrandomly oriented cell division in shaping and bending of the neural plate. *J Comp Neurol*, **381**(4), 473–88. 23, 109
- Schier, Alexander F, & Talbot, William S. 2005. Molecular genetics of axis formation in zebrafish. *Annual review of genetics.* 8, 16
- Schoenwolf, G C, & Alvarez, I S. 1989. Roles of neuroepithelial cell rearrangement and division in shaping of the avian neural plate. *Development*, **106**(3), 427–39. 23, 109

- Scholkopf, B, Platt, J, Shawe-Taylor, J, & Smola, A. 2001. Estimating the Support of a High-Dimensional Distribution. *Neural Computation*, Jan. 61
- Sethian, J. 1999. Level Set Methods and Fast Marching Methods : Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, Computer-Aided Design, Optimal Control and Material Sciences. [BOOK], Jan. 49, 57
- Seung, H, Opper, M, & Sompolinsky, H. 1992. Query by committee. *Proceedings of the fifth annual workshop on Computational learning theory*, Jan. 42
- Shafizadeh, Ebrahim, & Paw, Barry H. 2004. Zebrafish as a model of human hematologic disorders. *Curr Opin Hematol*, 11(4), 255–61. 6
- Shalizi, C, & Shalizi, K. 2004. Blind construction of optimal nonlinear recursive predictors for discrete sequences. *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, Jan. 52
- Shaner, Nathan C, Steinbach, Paul A, & Tsien, Roger Y. 2005. A guide to choosing fluorescent proteins. *Nat Methods*, 2(12), 905–9. 165, 169
- Shih, J, & Fraser, S E. 1996. Characterizing the zebrafish organizer : microsurgical analysis at the early-shield stage. *Development*, 122(4), 1313–22. 15
- Shimomura, O, & Johnson, F H. 1972. Structure of the light-emitting moiety of aequorin. *Biochemistry*, 11(9), 1602–8. 26
- Shitong, W, & Min, W. 2006. A new detection algorithm (NDA) based on fuzzy cellular neural networks for white blood cell detection. *IEEE Transactions on information technology in biomedicine*, Jan. 145
- Simpson, P. 1979. Parameters of cell competition in the compartments of the wing disc of *Drosophila*. *Developmental Biology*, 69(1), 182–93. 19
- Smal, I, Niessen, W, & Meijering, E. 2006. Bayesian tracking for fluorescence microscopic imaging. *Proc. IEEE Int. Symp. on Biomedical Imaging*, Jan. 49
- Smal, Ihor, Draegestein, Katharina, Galjart, Niels, Niessen, Wiro, & Meijering, Erik. 2007. Rao-Blackwellized marginal particle filtering for multiple object tracking in molecular bioimaging. *Information processing in medical imaging : proceedings of the conference*, 20(Jan), 110–21. 49
- Solnica-Krezel, L. 2006. Gastrulation in zebrafish—all just about adhesion? *Current opinion in genetics & development*, Jan. 10
- Squirrell, J, Wokosin, D, White, J, & Bavister, B. 1999. Long-term two-photon fluorescence imaging of mammalian embryos without compromising viability. *Nature Biotechnology*, Jan. 165

- Stearns, T. 1995. Green fluorescent protein : The green revolution. *Current Biology*, 5(3), 262–264. 26
- Stern, C. 2004. Gastrulation : from cells to embryo. *Hardcover*, Jan. 18
- Stern, C D, & Fraser, S E. 2001. Tracing the lineage of tracing cell lineages. *Nat Cell Biol*, 3(9), E216–8. 47
- Stern, Howard M, & Zon, Leonard I. 2003. Cancer genetics and drug discovery in the zebrafish. *Nature Reviews Cancer*, 3(7), 533. 7
- Stewart, Zoe A, Westfall, Matthew D, & Pietenpol, Jennifer A. 2003. Cell-cycle dysregulation and anticancer therapy. *Trends Pharmacol Sci*, 24(3), 139–45. 23
- Stickney, HL, Barresi, MJF, & Devoto, SH. 2000. Somite development in zebrafish. *Developmental Dynamics*, 219(3). 10
- Strasburger, E. 1880. [BOOK] Zellbildung und Zellteilung. *G. Fischer*, Jan. 19
- Strehlow, D, Heinrich, G, & Gilbert, W. 1994. The fates of the blastomeres of the 16-cell zebrafish embryo. *Development*, 120(7), 1791. 14
- Streisinger, G, Walker, C, Dower, N, Knauber, D, & Singer, F. 1981. Production of clones of homozygous diploid zebra fish (*Brachydanio rerio*). *Nature*, 291(5813), 293–6. 6
- Strickler, J, & Webb, W. 1991. Two-photon excitation in laser scanning fluorescence microscopy. *Proceedings of SPIE*, Jan. 30
- Strome, S. 1993. Determination of cleavage planes. *Cell(Cambridge)*, Jan. 23
- Sulston, J E, Schierenberg, E, White, J G, & Thomson, J N. 1983. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Developmental Biology*, 100(1), 64–119. 47, 164
- Sutton, R. 1988. Learning to predict by the methods of temporal differences. *Machine Learning*, Jan. 43
- Tao, Charles Y, Hoyt, Jonathan, & Feng, Yan. 2007. A support vector machine classifier for recognizing mitotic subphases using high-content screening data. *Journal of biomolecular screening : the official journal of the Society for Biomolecular Screening*, 12(4), 490–6. 46
- Thery, M, & Bornens, M. 2006. Cell shape and cell division. *Current Opinion in Cell Biology*, 18(6), 648–657. 23, 109
- Thom, R. 1976. Structural stability and morphogenesis. *Pattern Recognition*, Jan. 37
- Thompson, D, Wentworth, A, & Bonner, J. 1992. On growth and form. *orton.catie.ac.cr*, Jan. 37

- Tong, S, & Chang, E. 2001. Support vector machine active learning for image retrieval. *Proceedings of the ninth ACM international conference on Multimedia*, Jan. 44
- Tong, S, & Koller, D. 2002. Support vector machine active learning with applications to text classification. *The Journal of Machine Learning Research*, Jan. 44
- Turing, Alan Mathison, Ince, D, Britton, J L, & Saunders, P T. 1992. Collected Works of A. M. Turing. Jan. 37
- Uttenweiler, Dietmar, Weber, Cornelia, Jähne, Bernd, Fink, Rainer H A, & Scharr, Hanno. 2003. Spatiotemporal anisotropic diffusion filtering to improve signal-to-noise ratios and object restoration in fluorescence microscopic image sequences. *Journal of biomedical optics*, 8(1), 40–7. 142
- VanderAalst, W., & VanHee, K. 2000. Workflow Management , Models, methods and systems. *Eindhoven*. 39
- Vapnik, Vladimir Naumovich. 1982. Estimation of Dependences Based on Empirical Data. Jan. 43, 44, 61
- Vapnik, Vladimir Naumovich. 2000. The Nature of Statistical Learning Theory. Jan, 314. 43, 44
- Varela, F. 1979. [BOOK] Principles of biological autonomy. *North Holland New York*, Jan. 37
- Varga, ZM, Wegner, J, & Westerfield, M. 1999. Anterior movement of ventral diencephalic precursors separates the primordial eye field in the neural plate and requires cyclops. *Development*, 126(24), 5533. 168
- Vermot, J, Fraser, S, & Liebling, M. 2008. Fast fluorescence microscopy for imaging the dynamics of embryonic development. *HFSP Journal*, Jan. 28
- Vincent, L, & Soille, P. 1991. Watersheds in digital spaces : an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Learning*, Jan. 48, 146
- Vonesch, C, Aguet, F, Vonesch, J, & Unser, M. 2006. The colored revolution of bioimaging. *IEEE signal processing magazine*, Jan. 28
- Voronoi, G. 1904. Sur une fonction transcendante et ses applications à la sommation de quelques séries. *numdam.org*, Jan. 82
- Wahlby, C, Karlsson, P, Henriksson, S, & Larsson, C. 2008. Finding cells, finding molecules, finding patterns. *International Journal of Signal and Imaging Systems Engineering*, Jan. 142
- Waldeyer, W. 1888. Ueber Karyokinese und ihre Beziehungen zu den Befruchtungsvorgängen. *Archiv für mikroskopische Anatomie*, Jan. 19

- Wang, M, Zhou, X, Li, F, Huckins, J, King, R, & Wong, S. 2008. Novel cell segmentation and online SVM for cell cycle phase identification in automated microscopy. *Bioinformatics*, Jan. 46, 55
- Wang, Meng, Zhou, Xiaobo, King, Randy W, & Wong, Stephen T C. 2007. Context based mixture model for cell phase identification in automated fluorescence microscopy. *BMC Bioinformatics*, 8(Jan), 32. 46, 55
- Ward, Andrea B, Warga, Rachel M, & Prince, Victoria E. 2007. Origin of the zebrafish endocrine and exocrine pancreas. *Dev Dyn*, 236(6), 1558–69. 15
- Warga, R M, & Kimmel, C B. 1990. Cell movements during epiboly and gastrulation in zebrafish. *Development*, 108(4), 569–80. 10
- Watkins, C, & Dayan, P. 1992. Q-learning. *Machine Learning*, Jan. 43
- Wei, Y, & Mikawa, T. 2000. Formation of the avian primitive streak from spatially restricted blastoderm : evidence for polarized cell division in the elongating streak. *Development*, 127(1), 87–96. 23, 109
- Weiss, P. 1939. [BOOK] Principles of development. *Henry Holt New York*, Jan. 17
- Westerfield, Monte. 1995. The zebrafish book : A Guide for the Laboratory Use of Zebrafish (*Danio Rerio*). Jan. 169
- White, RM, Sessa, A, Burke, C, Bowman, T, LeBlanc, J, Ceol, C, Bourque, C, Dovey, M, Goessling, W, & Burns, CE. 2008. Transparent adult zebrafish as a tool for in vivo transplantation analysis. *Cell Stem Cell*, 2(2), 183–189. 7
- Wilson, E. 1925. The cell in development and heredity. *Electronic Scholarly Publishing*, Jan. 23
- Wilson, Stephen W, & Houart, Corinne. 2004. Early steps in the development of the forebrain. *Dev Cell*, 6(2), 167–81. 168
- Wolpert, L. 1994. Positional information and pattern formation in development. *Developmental Genetics*, Jan. 38
- Woo, K, & Fraser, SE. 1995. Order and coherence in the fate map of the zebrafish nervous system. *Development*, 121(8), 2595. 14, 15, 24, 109, 167
- Xu, C, Zipfel, W, Shear, J, Williams, R, & Webb, W. 1996. Multiphoton fluorescence excitation : new spectral windows for biological nonlinear microscopy. *Proceedings of the National Academy of Sciences*, Jan. 30
- Yan, J., Zhou, X., Yang, Q., Liu, N., Cheng, Q., & Wong, STC. 2006. An efficient system for optical microscopy cell image segmentation, tracking and cell phase identification. *IEEE International symposium on Image Processing*. 55

- Yang, F, Mackey, M, Ianzini, F, Gallardo, G, & Sonka, M. 2005. Cell segmentation, tracking, and mitosis detection using temporal context. *Lecture Notes in Computer Science*, Jan. 49, 50
- Zhang, B, Zimmer, C, & Olivo-Marin, J. 2004. Tracking fluorescent cells with coupled geometric active contours. *IEEE International Symposium on Biomedical Imaging : Nano to Macro*, Jan. 50
- Zhou, X, & Wong, S. 2005. Automated segmentation and tracking of cells in time-lapse microscopy using watershed and mean shift. *Intelligent Signal Processing and Communication Systems*, Jan. 49
- Zhou, X, & Wong, S. 2006. Informatics challenges of high-throughput microscopy. *IEEE signal processing magazine*, Jan. 47, 55
- Zhou, X, Wong, S, Sch, H, & Boston, M. 2006. High content cellular imaging for drug development. *IEEE signal processing magazine*, Jan. 55
- Zimmer, C, Labruyere, E, Meas-Yedid, V, & Guillen, N. 2002a. Segmentation and tracking of migrating cells in videomicroscopy with parametric active contours. *IEEE transactions on medical imaging*, Jan. 145
- Zimmer, C, Zhang, B, Dufour, A, & Thebaud, A. 2006. On the digital trail of mobile cells. *IEEE signal processing magazine*, Jan. 48
- Zimmer, Christophe, & Olivo-Marin, J C. 2005. Coupled parametric active contours. *IEEE transactions on pattern analysis and machine intelligence*, 27(11), 1838–42. 49
- Zimmer, Christophe, Labruyère, Elisabeth, Meas-Yedid, Vannary, Guillén, Nancy, & Olivo-Marin, Jean-Christophe. 2002b. Segmentation and tracking of migrating cells in videomicroscopy with parametric active contours : a tool for cell-based drug testing. *IEEE transactions on medical imaging*, 21(10), 1212–21. 49