SSBD Database
System s Science of Biological Dynamics

第41回日本分子生物学会年会　MBSJ 2018
ワークショップ　IPW2-09 生命科学のデータベース活用法2018

SSBD: 細胞・発生画像情報と生命動態情報の統合データベース

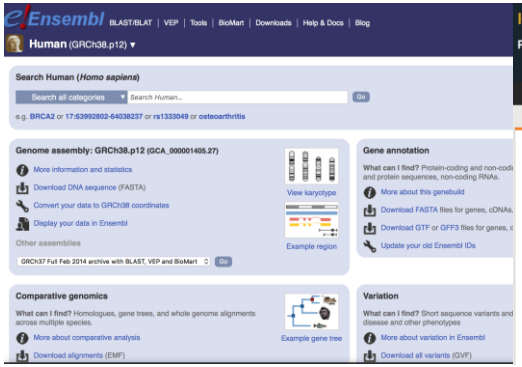
SSBD: an open public integrated database of microscopy images and quantitative data of biological dynamics

Ho, K.H.L.[1], Tohsato, Y.[1,2], Kyoda, K.[1], Itoga, H.[1], Onami, S.[1]
[1] Laboratory for Developmental Dynamics, RIKEN BDR
[2] Osaka Electro-Communication University

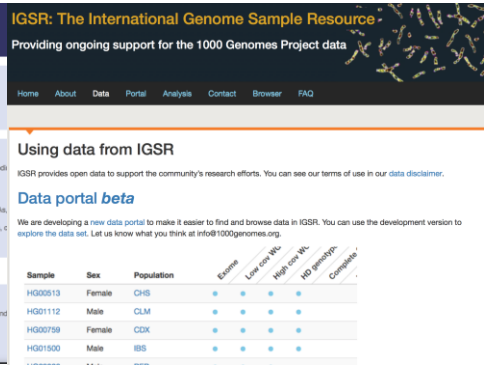
ホー・ケネス[1], 遠里 由佳子[1,2], 京田 耕司[1], 糸賀 裕弥[1], 大浪 修一[1]
[1]理化学研究所生命機能科学研究センター発生動態研究チーム
[2]大阪電気通信大学 情報通信工学部 情報工学科

Need for public repository

- There is a need for public repositories for retaining important knowledge and reusing experimental data



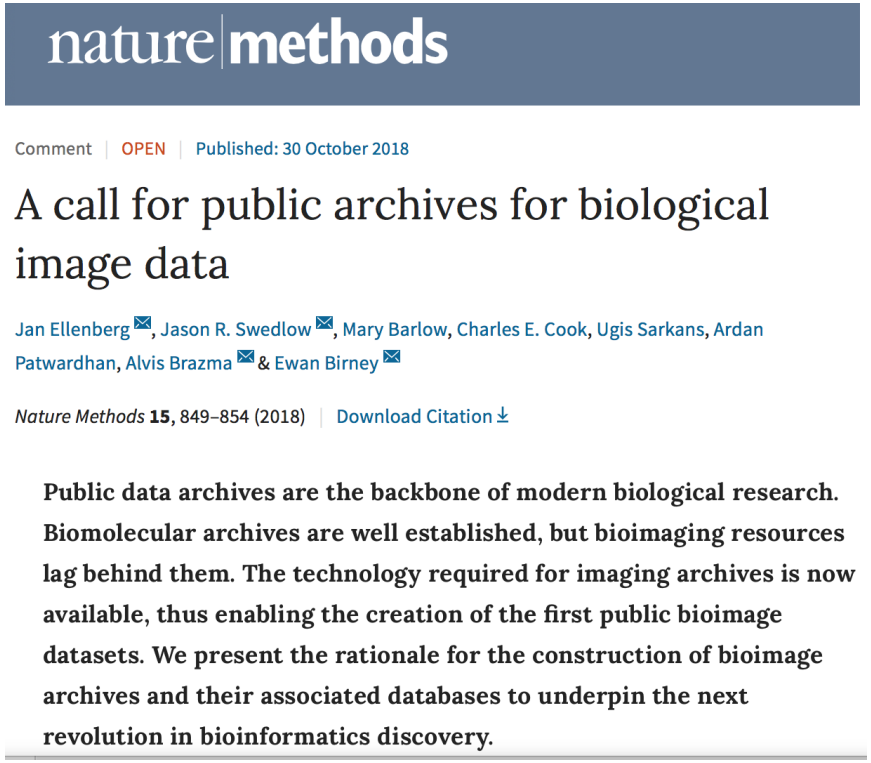
Ensembl



IGSR

KEGG/KEGG MEDICUS

PDBJ



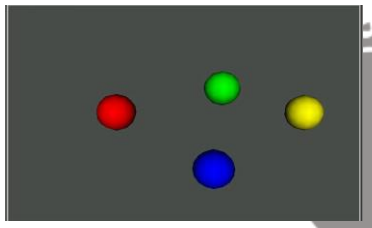
(Ellenberg et al, 2018)

Problem: difficult to reuse

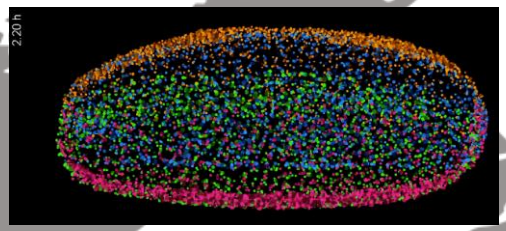
SSBD Database
Systems Science of Biological Dynamics

- Images and Datasets are scattered over the Internet
  - Difficult to locate and search
- Most quantitative data use different formats
  - Time consuming/difficult to reuse these datasets

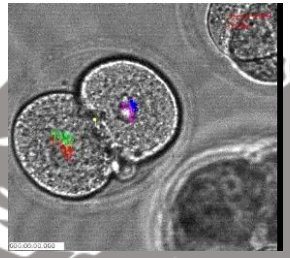
(Bao et al., 2006)



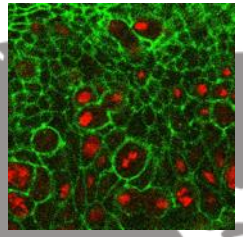
(Keller et al., 2010)



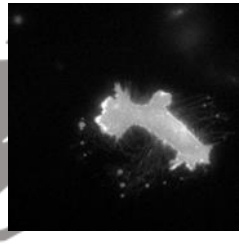
(Kurotaki et al., 2007)

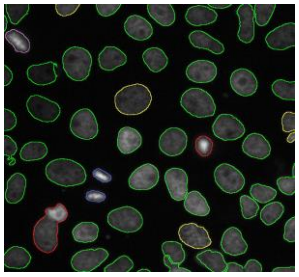


(Kondo & Hayashi 2013)

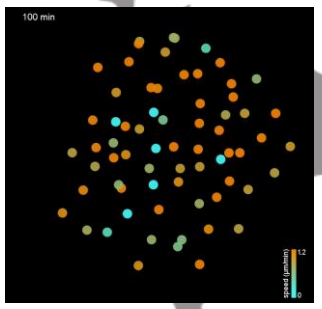


(Kunida et al. 2012)



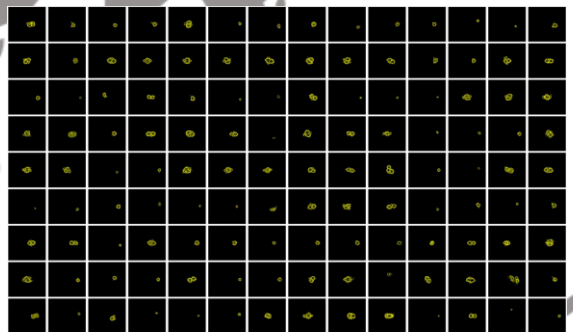


(Heldet et al., 2010)



(Keller et al., 2008)

(Yemini et al., 2013)



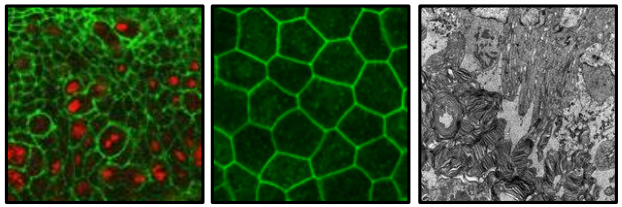
(Kyoda et al., 2013)

SSBD Database

SSBD Database
System s Science of Biological Dynamics

(Tohsato et al, Bioinformatics, 2016)

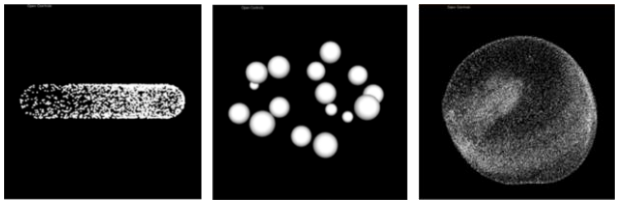
画像データ　Image data



(Kondo & Hayashi 2013)　(Inomata et al. 2013)　(Shirai et al. 2016)

Original format

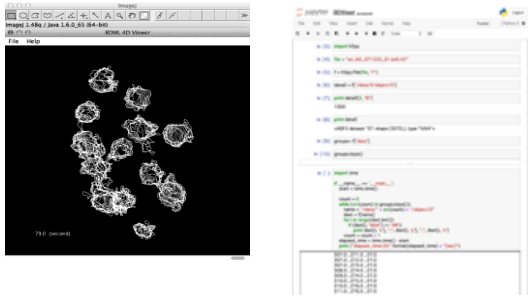
定量データ　Quantitative data



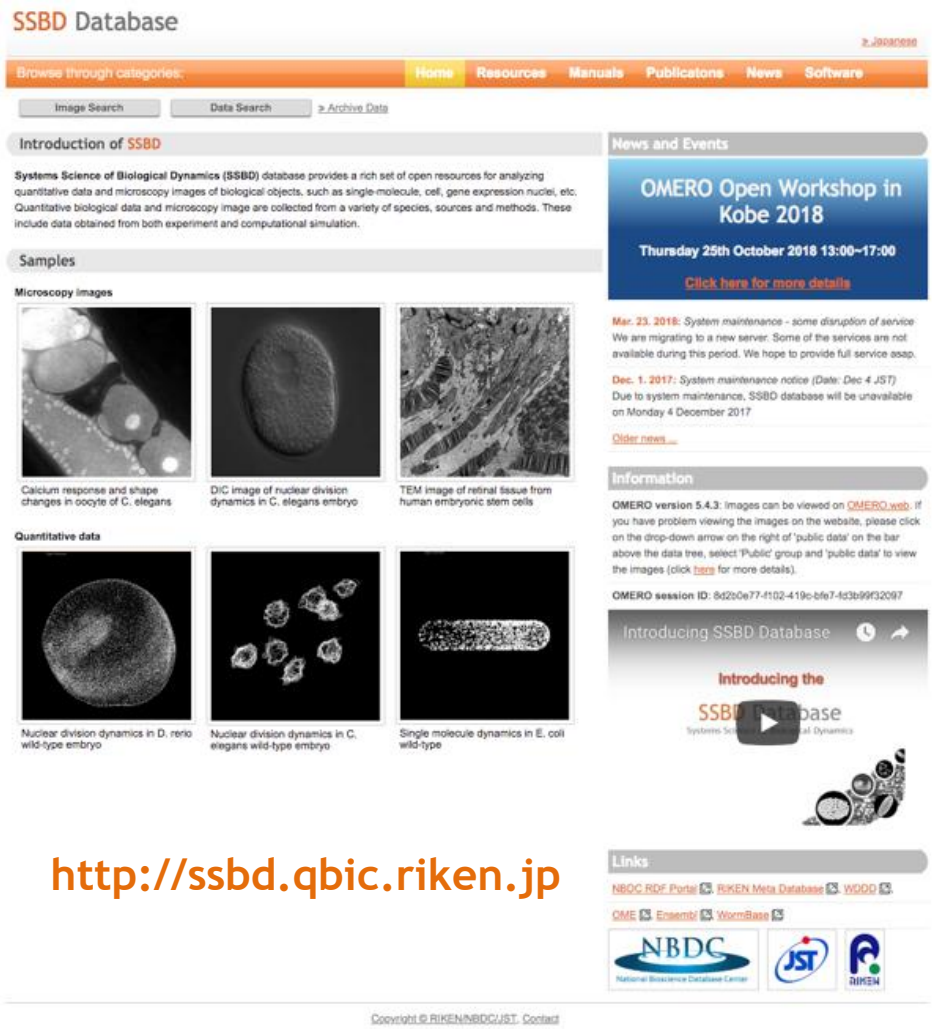
(Arjunan & Tomita 2010)　(Bao et al. 2006)　(Keller et al. 2008)

BDML/BD5 形式

ソフトウェアツール　Software tools



BDML/BD5 形式に対応



http://ssbd.qbic.riken.jp

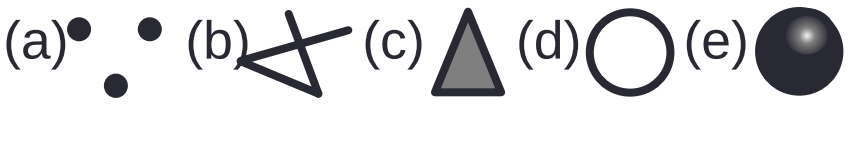
BDML/BD5 – an open unified file format Biological Dynamics Markup Language

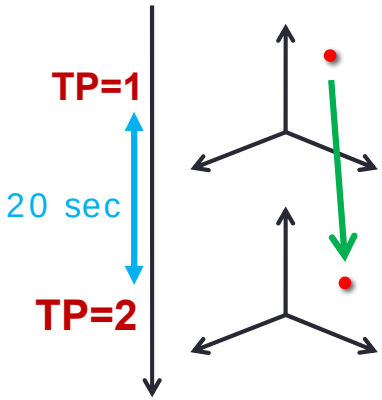
SSBD Database
System s Science of Biological Dynamics

(Kyoda et al., Bioinformatics, 2015)

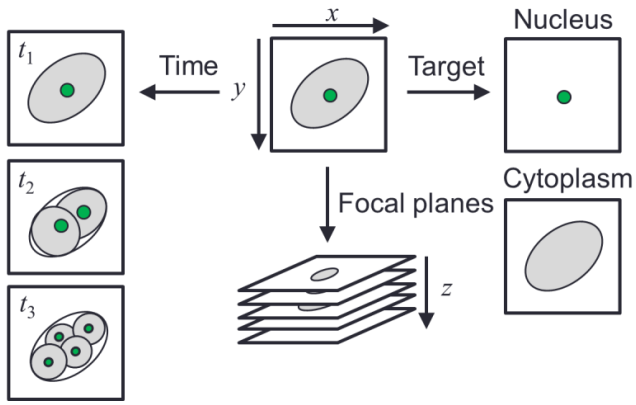
- Meta information in XML

Quantitative data in HDF

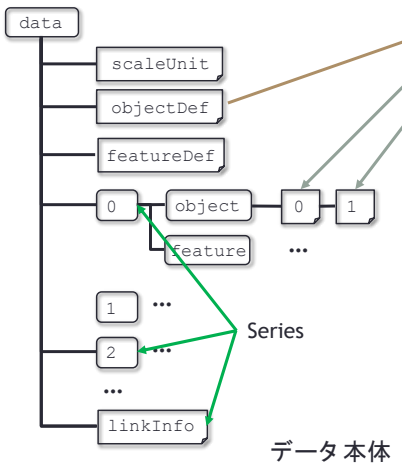





```
<scaleUnit>
 <tScale>20</tScale>
 <tUnit>second</tUnit>
</scaleUnit>
<component>
 <componentID>100</componentID>
 <time>1</time>
 <measurement>
  <point><xyz><x>1.3</x><y>3.4</y><z>1.2</z></xyz></point>
 </measurement>
</component>
<component>
 <componentID>101</componentID>
 <time>2</time>
 <prevID>100</prevID>
 <measurement>
  <point><xyz><x>0.9</x><y>3.2</y><z>1.1</z></xyz></point>
 </measurement>
</component>
```



Modified from (Goldberg et al, Genome Biology 2005)



oID	name
0	male_pronucleus
1	microtubule

0

ID	t	entity	x	y	z	radius
N001	0	sphere	189	251	17.2	3.96

1

ID	t	entity	sID	x	y	z
N002	0	line	0	387	153	16.6
N002	0	line	0	388	154	16.7
N002	0	line	1	365	245	15.0
N002	0	line	1	364	246	15.3
…						

SSBD Big Data resource

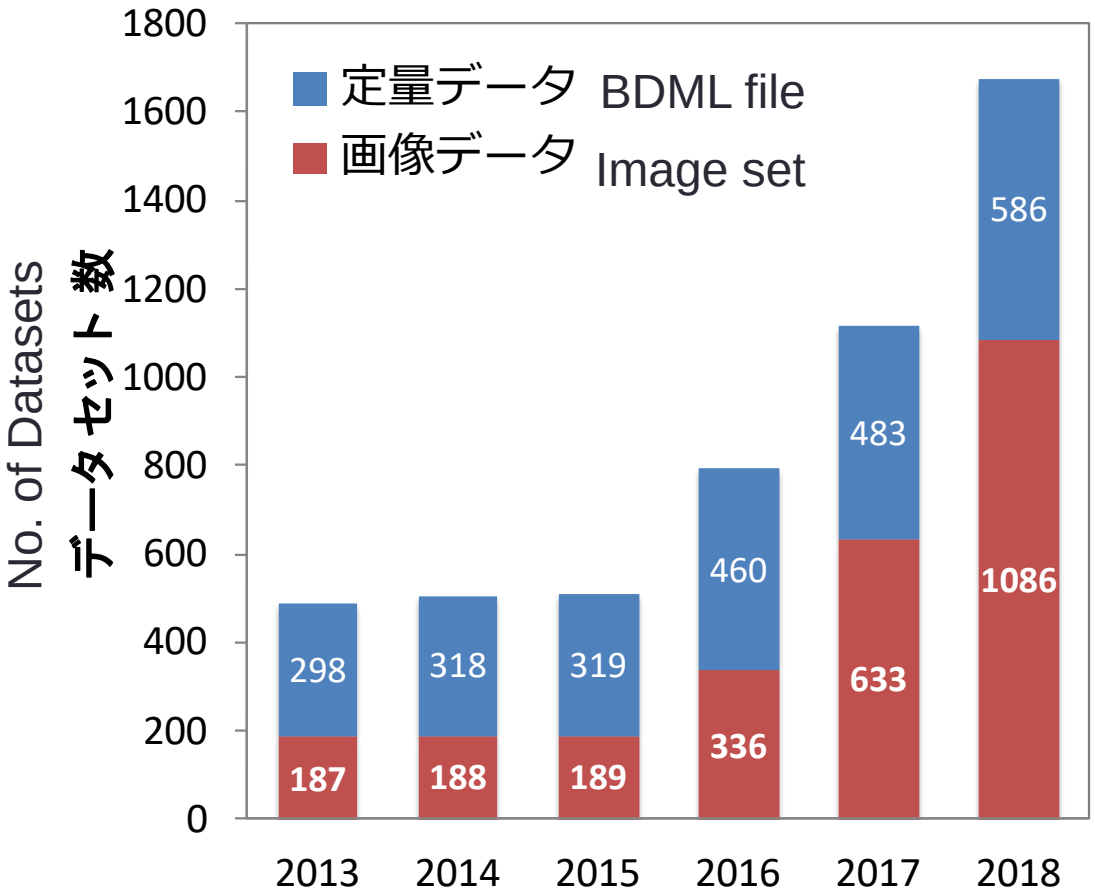
586 BDML-files, 64.7M entities
1086 image sets, 2.6M slices

Resources

As of Nov. 15, 2018 there are 586 BDML-files and 1086 image sets.

Organism	Type	Basedon	Paper	Number of BDML-files	Number of images in SSBD	Released date	Update date
C. elegans	Nucleus	Measurement	Bao et al. 2006	1	0	2013/10/03	**2018/11/15**
C. elegans	Nucleus	Measurement	Kyoda et al. 2013	186	2,221,560	2013/09/02	**2018/11/15**
C. elegans	Nucleus	Simulation	Kimura & Onami 2005	100	0	2013/10/03	**2018/11/15**
C. elegans	Behavior	Measurement	Cronin et al. 2005	11	0	2014/10/03	**2018/11/15**
C. elegans	Nucleus	Measurement	Toyoshima et al. 2016	14	41,380	2016/05/20	**2018/11/15**
C. elegans	Cell	Measurement	Takayama & Onami 2016	119	77,306	2016/10/03	**2018/11/15**
C. elegans	Cell	Measurement	Azuma & Onami 2016	1	144	2017/03/01	**2018/11/15**
C. elegans	Molecule	Measurement	Arata et a. 2016	3	3,000	2017/10/03	**2018/11/15**
D. cf. damesi	Individual	Measurement	Inoue & Kondo 2016	0	6,096	2017/10/03	**2018/11/15**
D. discoideum	Molecule	Measurement	Komatsuzaki et al. 2015	1	5,400	2015/10/03	**2018/11/15**
D. discoideum	Molecule	Measurement	Yasui et al. 2014	0	2,421	2017/10/03	**2018/11/15**
D. discoideum	Cell	Measurement	Shibata et al. 2012	2	1,440	2017/10/03	**2018/11/15**
D. discoideum AX2	Cell	Measurement	Kamimura et al. 2016	0	324	**2018/11/14**	-
D. discoideum, Cultured cell (rat; PC-12), NA	Molecule	Measurement/Simulation	Watabe et al. 2015	4	61	2016/01/18	**2018/11/15**
D. melanogaster	Nucleus	Measurement	Keller et al. 2010	2	0	2013/10/03	**2018/11/15**
D. melanogaster	Cell	Measurement	Kondo & Hayashi 2013	0	85,928	2017/10/03	**2018/11/15**
D. melanogaster	Cell	Measurement	Yasugi et al. 2017	0	12	**2018/11/14**	-
D. rerio	Nucleus	Measurement	Keller et al. 2008	7	0	2013/10/03	**2018/11/15**
E. coli	Molecule	Simulation	Arjunan & Tomita 2010	1	0	2013/09/02	**2018/11/15**
H. sapiens	Cell	Measurement	Takagi et al. 2017	0	32	**2018/11/14**	-
H. sapiens (ES)	Tissue	Measurement	Shirai et al. 2016	0	2	2017/10/03	**2018/11/15**
H. sapiens (iPS)	Cell	Measurement	Kanemura et al. 2013	0	48	2017/10/03	**2018/11/15**
H. sapiens, *M. musculus*	Cell	Measurement	Nojima et al. 2017	0	342	**2018/11/14**	-
M. musculus	Nucleus	Measurement	Bashar et al. 2012	1	2,800	2014/10/03	**2018/11/15**
M. musculus	Gene expression	Measurement	Masumoto et al. 2010	(OmicsBDML) 8	0	2014/10/03	2017/10/03
M. musculus	Cell/Organ	Measurement	Ke et al. 2013	0	21,995	2017/10/03	**2018/11/15**
M. musculus	Molecule/Cell	Measurement	Sakakibara et al. 2015	1	43,044	2017/10/03	**2018/11/15**
M. musculus	Molecule	Measurement	Hirata et al. 2016	11	0	2017/10/03	**2018/11/15**
M. musculus	Cell	Measurement	Noguchi et al. 2015	0	28,245	**2018/11/14**	-
M. musculus	Organ	Measurement	Susaki et al. 2014	0	12812	2017/10/03	**2018/11/15**
M. musculus	Organ	Measurement	Susaki et al. 2015	0	512	2017/10/03	**2018/11/15**
M. musculus	Organ	Measurement	Tainaka et al. 2014	0	543	2017/10/03	**2018/11/15**
M. musculus	Cell	Measurement	Herawati et al. 2016	0	5002	**2018/11/14**	-
M. musculus	Cell	Measurement	Minegishi et al. 2017	0	12	**2018/11/14**	-
M. musculus	Cell	Measurement	Ikeda et al. 2018	0	27	**2018/11/14**	-
M. musculus	Cell	Measurement	Fumoto et al. 2017	0	201	**2018/11/14**	-
M. musculus	Cell	Measurement	Kosodo et al. 2017	0	2738	**2018/11/14**	-
M. musculus	Cell	Measurement	Koeberle et al. 2017	0	677	**2018/11/14**	-
M. musculus	Cell	Measurement	Isshiki et al. 2014	0	5210	**2018/11/14**	-
M. musculus (implanted with KPL-4: *H. sapiens*)	Cell	Measurement	Jin et al. 2016	0	24	**2018/11/14**	-
M. musculus (implanted with KPL-4: *H. sapiens*)	Cell	Measurement	Tsuboi et al. 2017	0	39	**2018/11/14**	-
M. musculus (iPS)	Organ	Measurement	Kinoshita et al. 2016	0	384	2017/10/03	**2018/11/15**
M. musculus (iPS)	Cell	Measurement	Iwasaki et al. 2016	0	39	2017/10/03	**2018/11/15**
M. musculus (iPS)	Tissue	Measurement	Assawachananont et al. 2014	0	8	2017/10/03	**2018/11/15**
M. musculus (iPS), D. melanogaster, Cultured cell (HEK293T: human)	Organ	Measurement	Ke et al. 2016	0	14,656	2017/10/03	**2018/11/15**
M. musculus (iPS)	Cell/Particle	Measurement	Tanaka & Fujita 2015	2	14,573 (.tif)	2016/10/03	**2018/11/15**
M. sieboldi	Cell	Measurement	Tanaka et al. 2017	2	0	**2018/11/15**	-
X. laevis	Molecule/Cell	Measurement	Inomata et al. 2013	0	9,000	2017/10/03	**2018/11/15**
X. laevis	Cell	Measurement	Suzuki et al. 2016	0	1	2017/10/03	**2018/11/15**
X. laevis, Cultured cell (Hela: *H. sapiens*)	Cell	Measurement	Ishiwata et al. 2017	0	78	**2018/11/14**	-
Cultured cell (CHO: hamster + MDCK: dog)	Molecule/Cell	Measurement	Matsuda et al. 2012	0	867	2017/10/03	**2018/11/15**
Cultured cell (CHO: hamster)	Molecule/Cell	Measurement	Matsuda et al. 2015	0	764	2017/10/03	**2018/11/15**
Cultured cell (DO11.10: mouse)	Cell	Measurement	Ichimura et al. 2016	0	48	2017/10/03	**2018/11/15**
Cultured cell (Eph4: *M. musculus*)	Cell	Measurement	Yano et al. 2013	0	3	**2018/11/14**	-
Cultured cell (ES: *M. musculus*)	Cell	Measurement	Maekawa et al. 2015	0	45	**2018/11/14**	-
Cultured cell (Hela: *H. sapiens*)	Cell	Measurement	Inomata et al. 2017	0	24	**2018/11/14**	-

SSBD 2018 release datasets

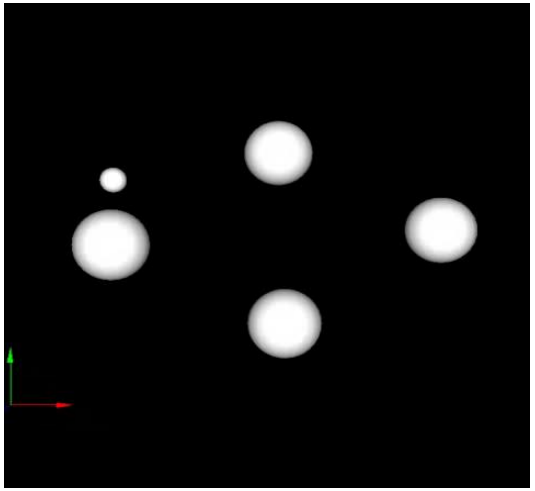


Increases in 2018

- BDML files: 103
  - 85,540 entities
- Image sets: 453
  - 58,919 slices

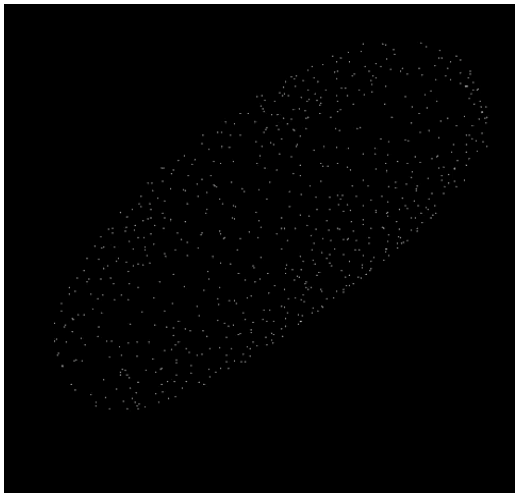
SSBD – examples of quantitative data

C. elegans embryogenesis



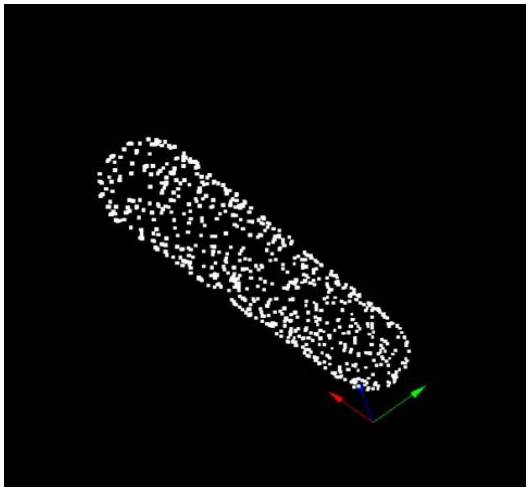
(Bao et al., 2006)

D. melanogaster embryogenesis



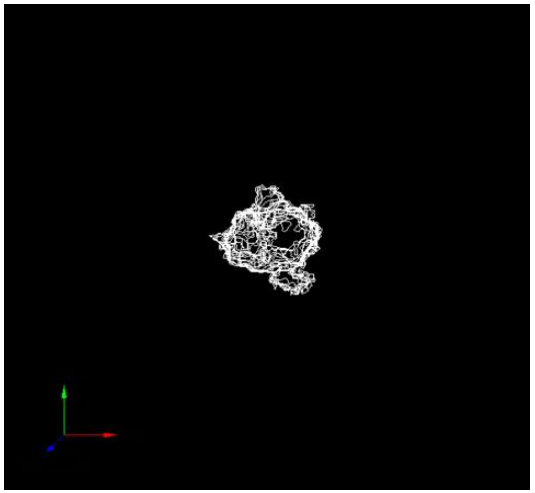
(Keller et al., 2010)

E. coli single-molecules



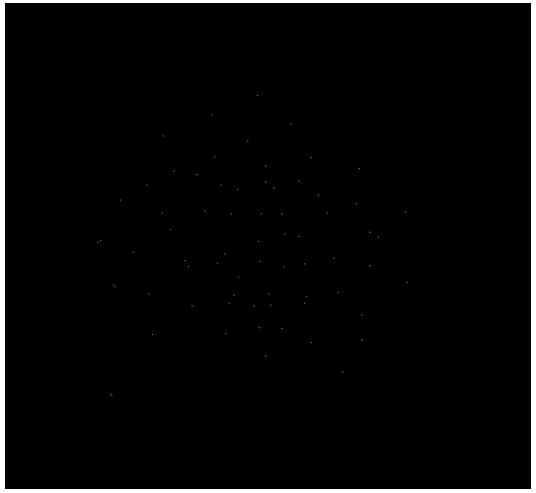
(Arjunan & Tomita, 2010)

C. elegans embryogenesis



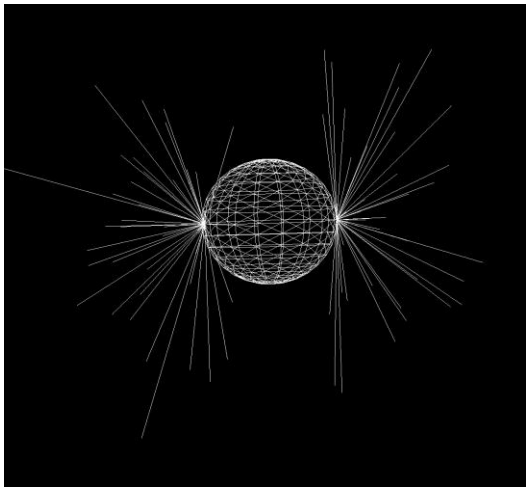
(Kyoda et al., 2013)

zebrafish embryogenesis



(Keller et al., 2008)

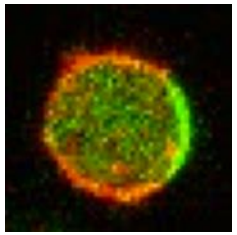
C. elegans pronuclear migration



(Kimura & Onami, 2005)

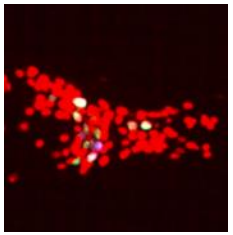
SSBD – examples of image data

Dictyostelium cell



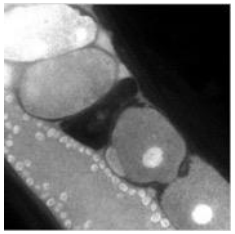
(Watabe et al. 2015)

C. elegans neuronal cell



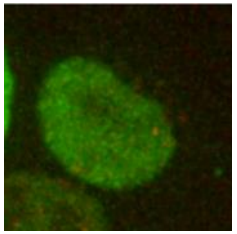
(Toyoshima et al. 2016)

C. elegans oocyte



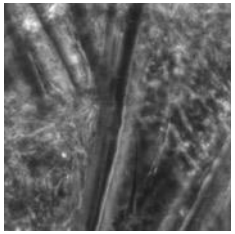
(Takayama & Onami 2016)

Mouse ES cell



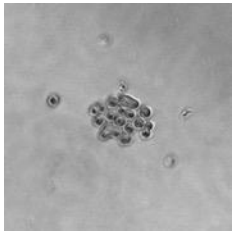
(Ochiai et al. 2015)

Mouse iPS cell



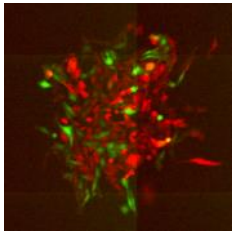
(Tanaka & Fujita 2015)

Mouse cultured cell



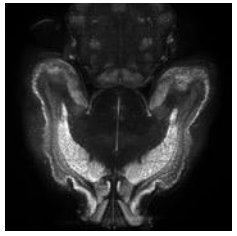
(Yonemura 2014)

Hamster cultured cell



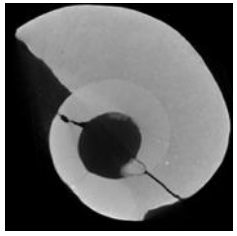
(Matsuda et al. 2015)

Mouse



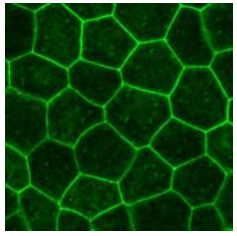
(Susaki et al. 2014)

D. cf. damesi



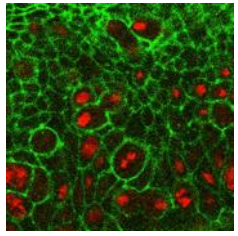
(Inoue & Kondo 2016)

X. laevis



(Inomata et al. 2013)

D. melanogaster

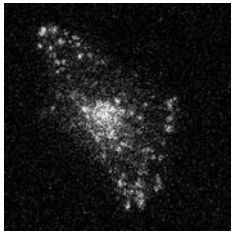


(Kondo & Hayashi 2013)

Rat

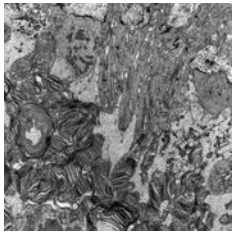
(Aoki et al. 2013)

Dog



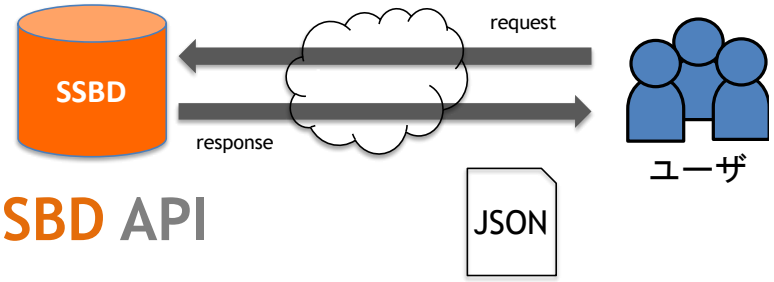
(Takai et al. 2015)

Human



(Shirai et al. 2016)

Analysis — REST API



SSBD API

Data can be accessed directly from SSBD REST API

Analysis - plotting the proliferation curve

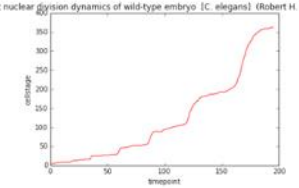
In [24]:
```
no_of_nucleus = []
timept = []
tp = 1
resultdata = retrieve_coord_tp('d15115', tp, display='off')
nn =resultdata['meta']['total_count']
while nn > 0:
    no_of_nucleus.append(nn)
    timept.append(tp)
    # print "tp="+str(tp)+" nn="+str(nn)
    tp=tp+1
    resultdata = retrieve_coord_tp('d15115', tp, display='off')
    nn =resultdata['meta']['total_count']
```

Get the title, organism and contact name of the dataset

In [25]:
```
q.set_display('off')
resultmetadata = q.data(field='bdmlUUID', search='d15115')
for i in resultmetadata['objects']:
    title= i['meta_data']['title']
    name= i['meta_data']['name']
    pmid = i['meta_data']['PMID']
    organism = i['meta_data']['organism']
```

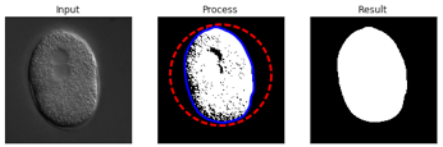
Plotting the curve

In [26]:
```
fig=plt.figure()
plt.plot(timept, no_of_nucleus, 'r')
ax = fig.add_subplot(1,1,1)
ax.set_ylabel('cellstage')
ax.set_xlabel('timepoint')
plottitle = title+' ['+organism+'] ('+name+') PMID: '+str(pmid)+' Cell divsion over time '
plt.title(plottitle);
```

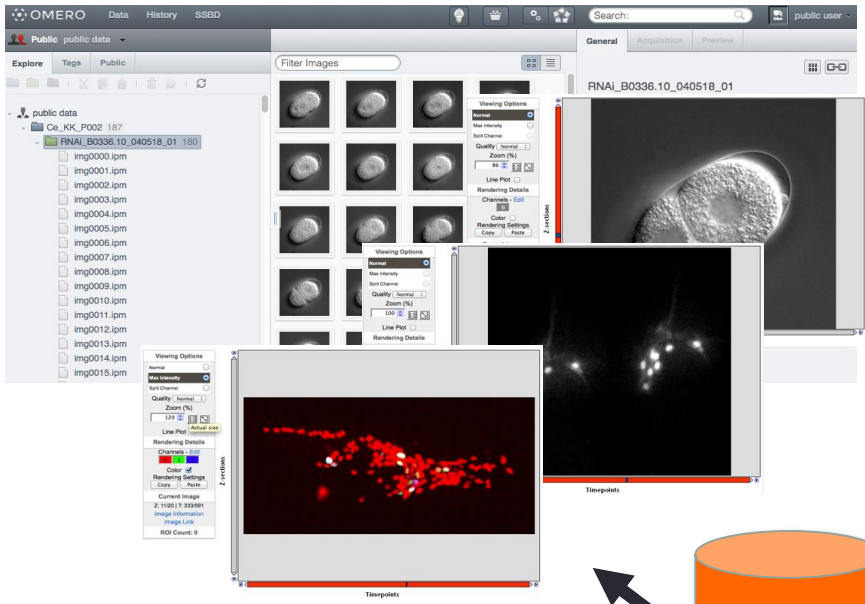


7. Apply active contour model (snake)

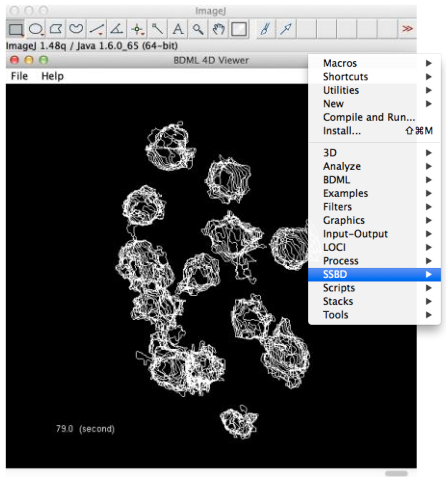
In [257]:
```
(x, y), radius = cv2.minEnclosingCircle(contours[max_i])
center = (int(x), int(y))
radius = int(radius) + 10 # make larger
s = np.linspace(0, 2*np.pi, 300)
x = center[0]+radius*np.cos(s)
y = center[1]+radius*np.sin(s)
init = np.array([x, y]).T

# alpha: length shape parameter. Higher values makes contract faster.
# beta: smoothness shape parameter. Higher values makes smoother.
# gannma: explicit time stepping parameter.
snake = active_contour(cv_img3, init, alpha=0.015, beta=200, gamma=0.001)

c = []
for s in snake:
    c.append([[s[0], s[1]]])
contour = np.array(c, dtype=np.int) # float -> int
#print ("snake.shape", snake.shape)
#print ("contour.shape", contour.shape)

cv_img4 = cv2.fillConvexPoly(np.zeros(size, dtype=np.uint8), contour, (255,255,255))

fig, (ax0, ax1, ax2) = plt.subplots(ncols=3, figsize=(10, 3))

ax0.imshow(cv_img1, cmap=plt.cm.gray)
ax0.set_xticks([])
ax0.set_yticks([])
ax0.set_title("Input")
ax0.axis([0, cv_img1.shape[1], cv_img1.shape[0], 0])

ax1.imshow(cv_img3, cmap=plt.cm.gray)
ax1.plot(init[:, 0], init[:, 1], '--r', lw=3)
ax1.plot(snake[:, 0], snake[:, 1], '-b', lw=3)
ax1.set_xticks([])
ax1.set_yticks([])
ax1.set_title("Process")
ax1.axis([0, cv_img3.shape[1], cv_img3.shape[0], 0])

ax2.imshow(cv_img4, cmap=plt.cm.gray)
ax2.set_xticks([])
ax2.set_yticks([])
ax2.set_title("Result")
ax2.axis([0, cv_img4.shape[1], cv_img4.shape[0], 0])

plt.savefig('result3.tiff')

vec = np.array(contour, dtype=np.int).reshape((contour.shape[0], contour.shape[2]))
#print (vec)
np.savetxt('output.csv',vec,delimiter=',', fmt='%d', header="x,y")
```



SSBD - Visualization

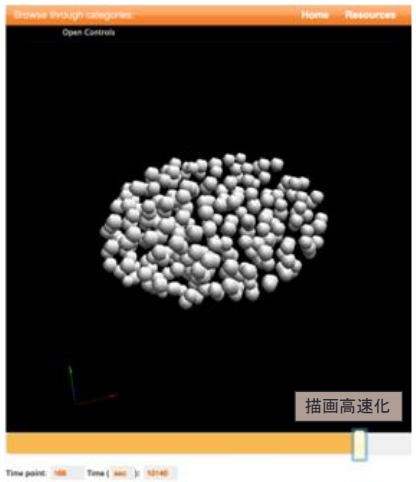


BDML対応ツール

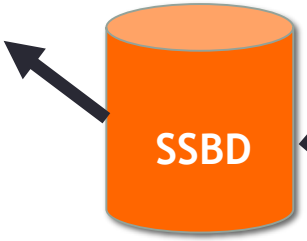


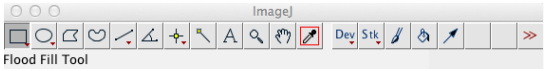
ImageJプラグイン

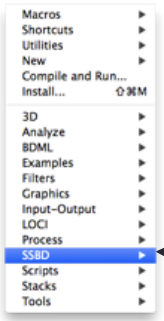
SSBD DBツール



JavaScriptアプリケーション



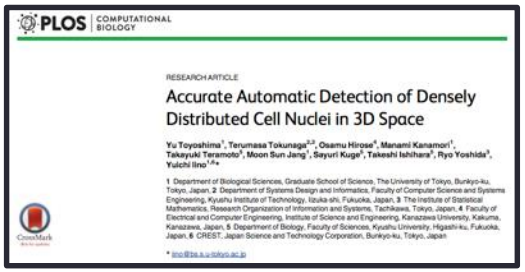




http://ssbd.qbic.riken.jp/image/publicKey/

SSBD repository service

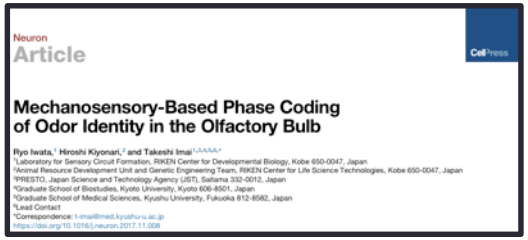
- Submitting publications to journals with image and numerical data
- Reviewer has access to datasets before publications
- After publications, SSBD will include them in our release.


PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Accurate Automatic Detection of Densely Distributed Cell Nuclei in 3D Space

(Toyoshima et al. 2016)


Neuron
Article

Mechanosensory-Based Phase Coding of Odor Identity in the Olfactory Bulb

(Iwata et al. 2017)


BMC Bioinformatics

METHODOLOGY ARTICLE

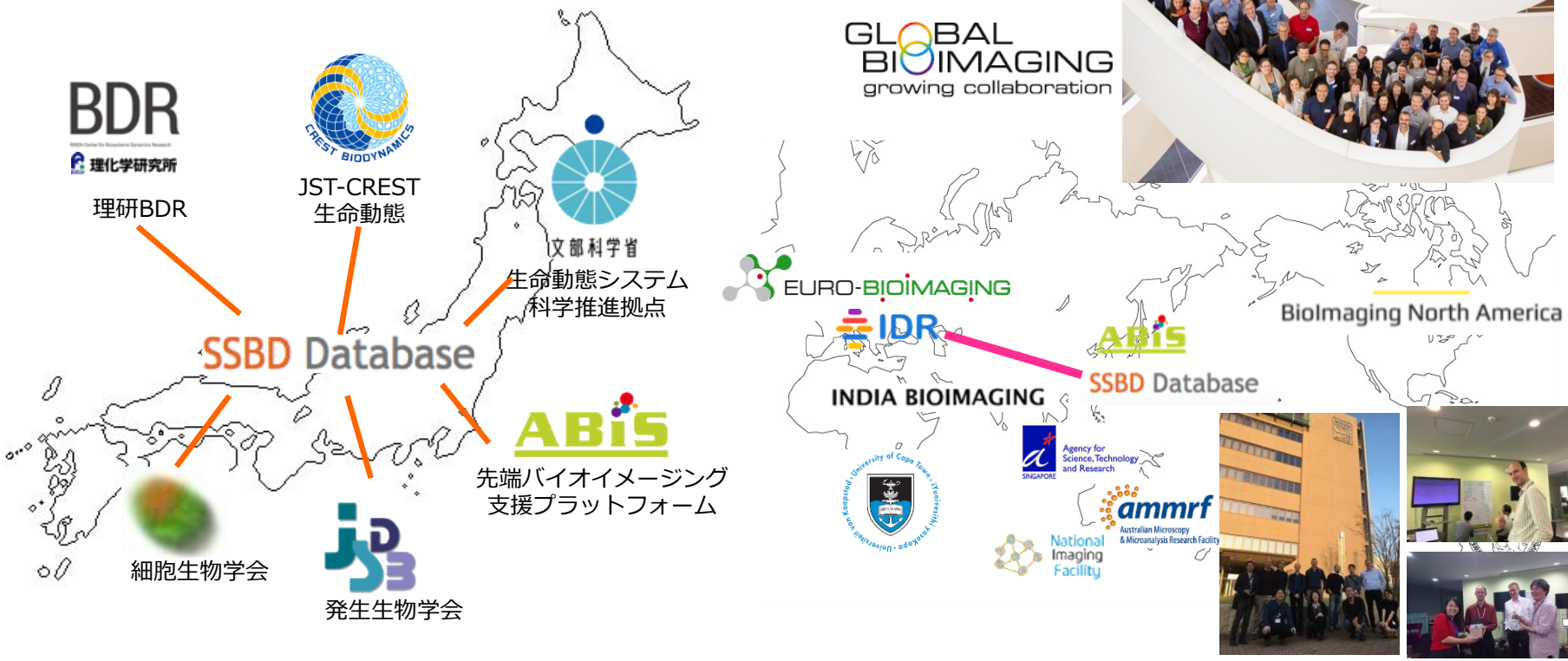
Open Access

Biologically constrained optimization based cell membrane segmentation in *C. elegans* embryos

(Azuma & Onami 2017)

Collaborations in Japan & Global partners

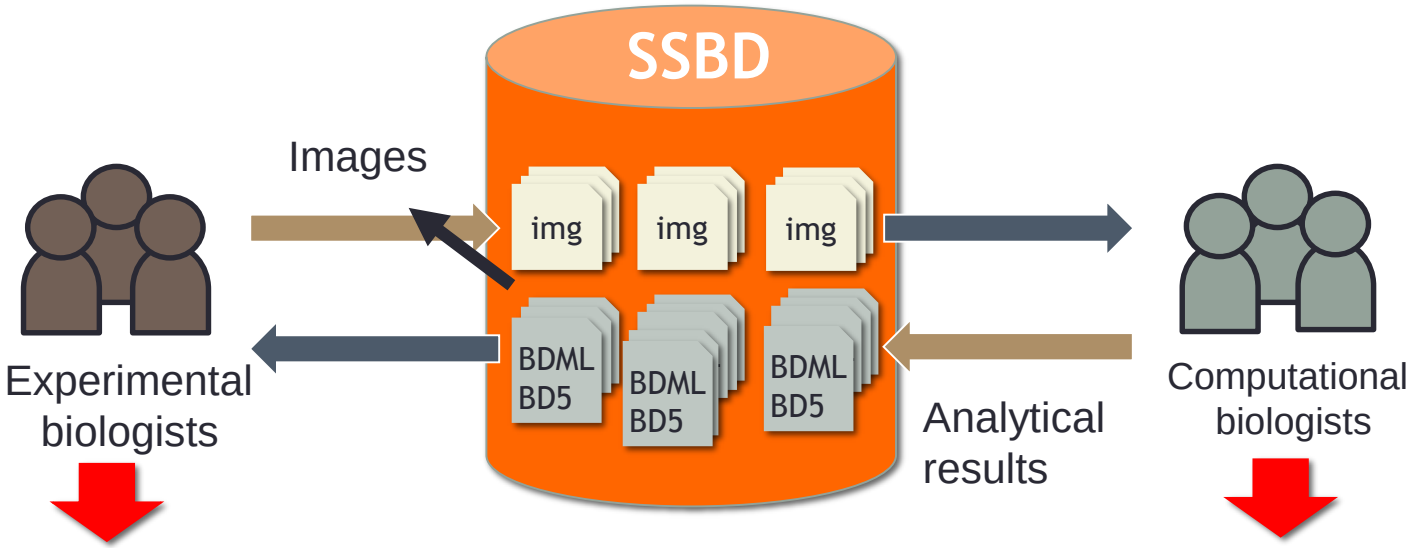
Working with various Japanese societies, institutes and research centers.



SSBD

Benefits of Open Science

- Public repository for sharing and reuse of data by biologists
- Open collaboration globally in developing computational tools/mathematical models



Understanding of biological phenomena

Improvement in image processing, computer simulation and mathematical modeling.

Acknowledgements

- **Project group**
  Yuichi Iino（Univ. Tokyo）
  Sigeo Ihara（Univ. Tokyo）
  Hiroki Ueda（RIKEN）
  Masahiro Ueda（RIKEN）
  Ryoichiro Kageyama（Kyoto Univ.）
  Kunihiko Kaneko（Univ. Tokyo）
  Sinya Kuroda（Univ. Tokyo）
  Minoru Ko（Keio Univ.）
  Shigeru Kondo（Osaka Univ.）
  Makoto Taiji（RIKEN）
  Shinichi Tate（Hiroshima Univ.）
  Michiyuki Matsuda（Kyoto Univ.）

- **Database development**
  OMERO team
  DBCLS
  Hiroshi Masuya（RIKEN）
  Norio Kobayashi (RIKEN)
  Kai Lenz（RIKEN）

- **Onami Lab members (RIKEN)**

SSBD Database
Systems Science of Biological Dynamics

NBDC National Bioscience Database Center
JST
CREST
RIKEN

Email: ssbd@riken.jp