The Subcellular Location Image Finder (SLIF)

### http://slif.cbi.cmu.edu



## Ultimate Goal of the field

- Machine understanding of biological journal articles (text and image)
- Criteria for success: Turing test have machine be able to answer questions about an article as well as a human scientist



### **Intermediate Goal**

- Extract information from combination of text and any kind of image in biological journal article
- Criteria for success: Achieve high precision and recall for extracted assertions (compared to expert scientist)



# Immediate Goal (SLIF)

- Extract information about subcellular location from captions and figures containing fluorescence microscope images in biological journal articles
- Criteria for success: Achieve high precision and recall for extracted assertions (compared to expert scientist)



# State of art: Bio Journal Information Extraction

- A number of systems to index literature via extracted terms
- A few systems to index image content in literature
- A few systems for document classification



## **SLIF** Pipeline components



### **Overview: Image processing tasks**

Segment into "panels"

Detect & remove **annotations** 

Classify panels

Find scale bars





### **Overview: Text processing** tasks

• Find *entity names* in text, and *panel labels* in text and the image.

• *Match* panels labels in text to panel labels on the image.

• Associate entity names to textual panel labels using *scoping* rules.

Figure 1. (A) Single confocal optical section of BY-2 cells expressing U2B 9 GFP. double labeled with GFP (left panel) and autoantibody against p80 coilin (right pane). Three nuclei are shown, and the bright GFP spots colocalize with bright foci of anticoilin labeling. There is some labeling of the cytoplasm by anti-p80 coilin. (B) Single confocal optical section of BY-2 cells expressing U2B 0 -GFP, double labeled with GFP (left panel) and 4G3 antibody (right panel). Three nuclei are shown. Most coiled bodies are in the nucleoplasm, but occasionally are seen in the nucleolus (arrows). All coiled bodies that contain U2B 0 also express the U2B 0-GFP fusion. Bars, 5 m m. Movement of Coiled Bodies Vol. 10, July 1999 2299 **Carnegie Mellon** 

#### http://slif.cbi.cmu.edu Finder

### WELCOME TO

Ð

(96%) Sun 10:18 AM

Q



### **Subcellular Location Image Finder**

SLIF (Subcellular Location Image Finder) automatically extracts information about protein subcellular locations from figure-caption pairs in biological literature. SLIF separates figures into panels and decides which panels contain fluorescence microscope images (FMI). It applies image processing methods to analyze the FMI and extract a quantitative description of the localization patterns they display. The associated captions are also processed to identify which portions of the caption refer to which panels and to identify the names of proteins contained in the captions. The results of this analysis are stored in the SLIF database.

Our long-term goal is to develop a large library of annotated and analyzed fluorescence microscope images, in order to support data-mining.

#### PNAS, version 3.0

The current version of the database contains records for 15180 papers from volumes 94-99 of the Proceedings of the National Academy of Sciences (USA), generously made available by the Academy for demonstration purposes.

Pubmed Central, version 1.1

This database includes all open access articles in Pubmed Central (approximately 42,000 articles through 30 April 2007).

A service of the Robert F. Murphy laboratory Departments of Biological Sciences, Biomedical Engineering, and Machine Learning and Center for Bioimage Informatics Carnegie Mellon University, Pittsburgh, Pennsylvania, U.S.A.

11.

http://slif.cbi.cmu.edu/PMC/

#### Subcellular Location Image Finder

#### Home

Search by words in caption

Search by pixel resolution

Search by protein name

Search for FMI figures/panels

Search by location inferred from GO terms

Search by locations inferred from Images

Browse by combination of locations inferred from GO Terms and Images

Advanced Search

About SLIF

Help

Murphy Lab SLIF Service

SLIF Web Application Version 2.1 Database: pmc

Note: If your browser manages only one cookie per all the windows connecting to the same server.

SLIF (Subcellular Location Image Finder) automatically extracts information about protein subcellular locations from figure-caption pairs in biological literature. SLIF separates figures into panels and decides which panels contain fluorescence microscope images (FMI). It applies image processing methods to analyze the FMI and extract a quantitative description of the localization patterns they display. The associated captions are also processed to identify which portions of the caption refer to which panels and to identify the names of proteins contained in the captions. The results of this analysis are stored in the SLIF database, which can be queried either interactively (using the links at the left) or via external program-generated links.

Our long-term goal is to develop a large library of annotated and analyzed fluorescence microscope images, in order to support data-mining and integration with other databases.

Acknowledgments People Reference

5	Safari File Edit View Histo	ry Bookmarks Window Help 🚯 🔂 📕 🛜 🐠 💷 (96%)	Sun 10:18 AM 🛞 🔍
n	00	http://slif.cbi.cmu.edu/PMC/db_desc.do	
	Subcellular Location Image Finder		
	Home	SLIF Database pmc	
	Search by words in caption	Source	Pubmed Central
2	Search by pixel resolution	Number Of Articles	41981
	bouldin by pixer resolution	Total Captions	81614
	Search by protein name	Total Figures	81617
	Search for FMI figures/panels	Total Panels	264752
	Search by location inferred from	FMI Panels	37026
	GO terms	Total Proteins	45956
	Search by locations inferred from	Total Cell Types	0
	Images	Number of Subcellular Locations Inferred from GO Terms	21
9	Browse by combination of	Number of Subcellular Locations Inferred from Images	10
	locations inferred from GO		
	Terms and Images		
	Advanced Search		
]	About SLIF		
	Help		

A .

Murphy Lab SLIF Service

SLIF Web Application Version 2.1 Database: <u>pmc</u>

Note: If your browser manages only one cookie per all the windows connectine to the same server.

Ś	Safari	File	Edit	View	History	Bookmarks	Window	Help		Ð	C 🔳	(î)	<b>(96%</b> )	Sun 10:19 AM	*	Q
. 19	000	)				htt	p://slif.cbi.	cmu.edu,	/PMC/captio	on_sear	ch.do					
	SI		F	i.												
• 🍋 • 🎯	Subcellu	ılar Lo	cation	Image	Finder											
	Home				Se	arch by words	in caption									
	Search	<u>by wo</u>	ords in o	<u>caption</u>	C						]					
× 18	Search	<u>ı by pix</u>	cel reso	lution	N	umber of reco	rds display	ed per p	age: 10	\$						
?	Search	i by pro	<u>AI figu</u>	<u>ime</u> res/pane	ls	earch										
• 🛅 X 4	Search GO te	1 by loc r <u>ms</u>	ation in	nferred f	f <u>rom</u>											
?	Search Image	<u>ı by loc</u> <u>s</u>	ations	inferred	from											
	Brows locatio Terms	e by co ns infe and In	ombinat rred fro nages	<u>tion of</u> om GO												
	<u>Advar</u> <u>About</u>	SLIF	arch													
? ?	<u>Help</u>															
·?·@	Murphy Lo	ıb SLIF S	ervice													
	SLIF Web Database:	Application <u>pmc</u>	on Versio	n 2.1												
	Note: If yo per all the	ur browse windows o	er manage connecting	es only one g to the sar	cookie ne server,											<b>•</b>

- <b>É</b>	Safari	File	Edit	View	Histo	ry Boo	kmarks	Windo	w Help		Ð	0	ŝ		<b>(95%)</b>	Sun 10:21	AM 🖇	Q
	00		_	_	_	_	http	://slif.c	bi.cmu.edu	/PMC/capt	ion_sea	arch.do	_	_	_	_		
	Subcellu	llar Loc	cation	Image l	Finder													
	Home				S	earched f	or:											
	Search	by wor	ds in c	aption	C	aptions	with the w	ord(s) a	ctin.									
N.	Search	by pixe	el resol	ution	R	esults 1 -	10 of 365	50										
20.	Search	i by prot	tein na	me	0	Next 10 Re	sults											
? .⊠ X	Search Search from C	tor FM by loca O term	<u>II figur</u> ation in <u>s</u>	es/panel	s	Ca	ption		Figure	Pan	el	Protein ↓	n †	Cell Type	Location Inferred from	Location Inferred from GO	Resolutio	n 1)
2	Search from I	<u>by loca</u> mages	ations i	nferred										•		Terms ↑ ↓	•	
🚱 🖂 🎲 <mark>🖵 </mark> 🐮 🧲	Brows locatio Terms Advan About	e by con ns infen and Ima aced Sea SLIF	mbinat red fro ages urch	<u>ion of</u> m GO	a	Preven poptosis cells by l Bcl-XL mutan restr subce localiza	tion of in MCF-7 Bcl-2 and and by ts with icted ellular tion. A)		k to view in		view in	Bcl2	2	N/A	Endosome	N/A	0.057576	5
?·?·?·@	Help Murphy La	b SLIF Se	rvice		I	containing protein fr control transfecte cells exp	g 20 μg of om vector (neo) or d MCF-7 pressing		ew window	<u>a new w</u>	<u>indow.</u>							
© 1	SLIF Web Database: Note: If yo cookie per	Application <u>pmc</u> ur browser all the win	n Version manage dows con	2.1 s only one enecting to	the	Morpho Pax6 exp cultured epithelial	logy and ression of corneal cells. (A)											

Ś	Safa	ri	File	Edit	View	History	Bookmarks	Wind	low Help	Ð	0 🔳 🤋	( <b>♦</b>	<b>(94%)</b>	Sun 10:22	AM 🖇	Q
	0	0	_	_	_	_	htt	p://slif	f.cbi.cmu.edu/P	MC/caption_sear	ch.do	_	_	_	_	
	Subc	ellul	ar Lo	cation	Image	Finder										
<u></u>	Ho	me				s	earched for:									
5	Sea	urch	by wo	rds in o	<u>caption</u>	C	Captions with th	e word	d(s) <b>tubulin</b> .							
	Sea	urch l	by pix	el reso	lution	F	tesuits 1 10 of	16940								
	Sea	irch	<del>by pre</del>	tein na	me	(	Next 10 Results									
?	<u>Sea</u> <u>Sea</u> <u>GC</u>	urch i urch i ) terr	for FN by loc ns	<u>4I figu</u> ation ir	res/pane	<u>ls</u> from	Caption		Figure	Panel	Protein 1	Cell Type ↑↓	Location Inferred from Image †	Location Inferred from GO Terms †	Resolution (µm/pixel) ↑↓	1
?	Sea Ima	urch l ages	by loc	ations i	inferred	from	TNE does not						•	↓.		
	Bro loc Ter	owse ation ms a	by co s infer and Im	mbinat rred fro lages	<u>tion of</u> om GO		interfere with neural stem cel differentiation. (A)	1		β-III-tubulin						
	<u>Ad</u> <u>Ab</u>	vanc out S	ed Sea	arch			Immunostainin of neuronally differentiated neurosphere-	g e			<u>bFGF</u>	N/A	Endosome	N/A	0.09	
?	He	lp					derived NSCs Cells were		Click to view in	Click to view in a new window.						
90.	Murph	y Lab	SLIF S	ervice			D-lysin/laminin coated dishes in	1- 🔺 🧉	a new window.							
	SLIF V Datab	Veb Aj ase: <u>pi</u>	pplicatio <u>mc</u>	on Version	n 2.1		TNF does not interfere with neural stem cel									4
	Note: 1 per all	lf you the wi	r browse indows c	r manage onnecting	s only one to the sai	cookie me server,	differentiation.			β-III-tubulin						•

Ś	Safari	File	Edit	View	History	Bookmarks	Window	Help		Ð	C 🔳	<b>€</b> ♦	<b>(94%)</b>	Sun 10:22 A	M ∦	Q
. 🕲	0 C	)	-	_	_	htt	p://slif.cbi.o	cmu.edu/	PMC/scaleba	ır_sear	rch.do	_	_	_	_	
<ul> <li>3</li> <li>5</li> <li>5</li> <li>5</li> <li>6</li> <li>7</li> <li>7</li></ul>	S] Subcellu	llar Lo	cation	Image	Finder											
	Home				) Ca	Search by pix aptions with the	el resolution e word(s) tr	n, <i>within</i> <b>ubulin</b> .								
	Search	<u>ı by wo</u>	ords in o	<u>caption</u>	0	Search by pix	el resolution	n								
	Search Search	<u>i by pix</u> i by pro	tel resol	lution me		e min resolutio	on is 12 Mic	crons/Pixe	el and the ma	x is 0.	.05 Micr	ons/Pixel				-
۰É	Search	for FN	AI figu	res/pane	<u>ls</u> 💿	High: 0.05 ~ 0	).311									
×	<u>Search</u> GO te	<u>n by loc</u> rms	ation in	nferred f	rom O	Medium: 0.31 Low: 1.931 ~	1 ~ 1.931 12									
?	Search Image	<u>ı by loc</u> <u>s</u>	ations	inferred	from	Other:	~									
- <b>- - -</b>	Brows location Terms	e by co ons infe and In	ombinat rred fro nages	<u>tion of</u> o <u>m GO</u>	N	umber of reco	rds display	ed per pa	age: 10 🛟	)						
	Advar	nced Se	arch													
2	About	SLIF														
?	<u>Help</u>															
? 2 0	Murphy Lo	ab SLIF S	ervice													
) 🛄	SLIF Web Database:	Applicati <u>pmc</u>	on Versio	n 2.1												1
	Note: If yo	ur browse	er manage	es only one	cookie											

<u></u>	Safari	File	Edit	View	History	Bookmarks	Window	Help		Ð	C 🔳	(î)	) (	<b>•</b> (94%)	Sun 10:2	4 AM	*	Q,
	000	)		_	_	htt	p://slif.cbi.	.cmu.edu	/PMC/protein	_sear	ch.do		_	_	_			
	SI] Subcella	llar Lo	cation	Image	Finder													
2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	<u>Home</u> <u>Search</u>	<u>ı by wc</u>	ords in o	<u>caption</u>	Se Im C	arch by proteir ages with <b>pixe</b> aptions with th	n name, <i>with</i> I <b>resolutior</b> ne word(s) t	hin 1 betweer tubulin.	n <b>0.05</b> and <b>0.</b> 3	31072	3250595	538597	micro	ons/pixel	, within			
<b>1</b>	Search	<u>1 by pix</u> 1 by pro	<u>tel resol</u>	lution ime	o E	Protein names	extracted b	y CRFs										
?	Search	n for FN	AI figu	res/pane		Protein names	extracted b	y Sprot n	natcher									
	GO te	<u>rms</u> n by loc	ation ir	inferred i	from	Please select- Protein names	extracted b	y Uniref	100 matcher									
? •	Image	<u>s</u> e by co	mhina	tion of	6	Please select-	🗘											_
	locatio Terms	ons infe and In	rred fro	om GO	Se	arch by proteir Protein names	n name extracted b	v CRFs										
	Advar	st IF	arch		Ē	Please select-		,	•									
<b>ම</b> ?	Help	<u>, olur</u>			0	Protein names	extracted b	y Sprot n	natcher									
?·?·@	Murphy Lo	ab SLIF S	ervice		0	Please select- Protein names	extracted b	y Uniref	100 matcher									
0	SLIF Web Database:	Application <u>pmc</u>	on Versio	n 2.1	- Ni	Please select-	 rds displav	ed per p	<b>age:</b> 10 <b>•</b>	1								
1	Note: If yo per all the	ur browse windows o	er manage connecting	es only one g to the sar	cookie ne server, 🕝			<b>- P</b>		,								•

- <b>É</b>	Safari	File	Edit	View	History	Bookmarks	Window	Help		-9 (		<b>?</b>	) 🔳 (94%	Sun 10:24 AM	*	Q
	00				_	htt	p://slif.cbi.	.cmu.edu	I/PMC/proteir	n_searcl	h.do		_			
	Subcellu	lar Lo	cation	Image	Finder											
000	<u>Home</u> <u>Search</u>	by wo	ords in c	caption	Se Im C	arch by protein ages with <b>pixe</b> <b>aptions</b> with t	n name, with I resolution he word(s) t	hin 1 betwee tubulin.	n <b>0.05</b> and <b>0.</b>	310723	250595	38597 1	microns/pixe	el, within		
N IN	Search	by pix	el resol	lution	0	Protein names	extracted b	y CRFs								
•	Search Search	by pro	otein na AI figur	<u>me</u> res/pane	ls	bFGF Dissociated NSCs EGF GFAP	acted b	y Sprot 1	matcher							
×	Search GO ter	by loc ms	ation ir	nferred f	rom	TNF Protein names	extracted b	y Uniref	100 matcher							
?	Search Images	by loc	ations i	inferred	from	Please select-	🛟	-								
	Brows location Terms	<u>e by co</u> ns infe and Irr	ombinat rred fro nages	<u>tion of</u> m GO	Se	arch by protein	n name									
- <mark></mark> -	Advan	ced Se	arch		•	Protein names	extracted b	y CRFs	•							
- E	<u>About</u>	SLIF			0	Protein names	extracted b	y Sprot 1	matcher							
??	neip				-	Please select-			¢							
0	Murphy La	b SLIF S	ervice		0	Protein names	extracted b	y Uniref	100 matcher							
) <b>  </b>	SLIF Web A Database: j	Application <u> pmc</u>	on Version	n 2.1	L- Ni	umber of reco	 rds display	ed per p	• age: 10 •							
	Note: If you per all the y	ur browse windows c	r manage connecting	s only one to the sar	cookie ne server, 📿											▼ //.

<u></u>	Safari	File	Edit	View	History	Bookmarks	Window	Help		Ð	C 🔳	<b>?</b> ●	<b>(94%)</b>	Sun 10:24 A	M *	Q
	00		_	_		htt	p://slif.cbi.	cmu.edu/	/PMC/protein	_sear	ch.do					
•	SI Subcellu	lar Lo	cation	Image 1	Finder											
	Home Search Search Search Search GO ter Search Images	by wo by pix by pro for FN by loc ms by loc	ords in c cel resol otein na AII figur cation ir cations i	caption lution ume res/pane nferred f	Se Im C Im C Im C Im C Im C Im C Im C Im	arch by proteir ages with <b>pixe</b> <b>aptions</b> with the Protein names Please select Protein names Please select	n name, with I resolution ne word(s) t extracted b extracted b extracted b acted b	hin 1 between tu <b>bulin</b> . y CRFs y Sprot m y Uniref1	a <b>0.05</b> and <b>0.3</b> hatcher	31072	3250595	3 <b>8597</b> m	icrons/pixel	, within		
≥ ? ? 🚱 🖂 🗞 🧲	Brows locatio Terms Advan About Help	e by cc ns infe and Im ced Se SLIF	ombinat rred fro ages arch	tion of om GO	Se • • •	arch by proteir Protein names Please select- Protein names Please select-	extracted b extracted b extracted b	y CRFs ; y Sprot m	atcher							
? 2 0	Murphy La SLIF Web / Database: p Note: If you per all they	b SLIF S Application <u>pmc</u> ur browsee windows of	ervice on Version or manage	n 2.1 es only one e to the son	O I Nu cookie	Protein names	extracted b  rds display	y Uniref1	age: 10 ¢	)						

Ś	Safari	File	Edit	View	History	Bookmarks	Window	Help		Ð	C 🔳	<b>?</b> ₽	<b>(93%</b> )	Sun 10:24 A	M ∦	Q
	00	)	_	_		htt	p://slif.cbi.	.cmu.edu/	PMC/protein	n_sear	ch.do				_	
	Subcellu	llar Lo	cation	Image	Finder											
	Home Search Search Search Search GO ter Search GO ter	a by wo by pix by pix by pro for FN by loc rms by loc s	ords in o cel reso otein na vII figun cation in cations i	caption lution me res/pane nferred f	se Im C Im C C C C C C C C C C C C C C C C	arch by proteir ages with <b>pixe</b> <b>aptions</b> with th Protein names Please select Protein names Please select CF CFAP	a name, with I resolution he word(s) t extracted b extracted b extracted b	hin n between tubulin. by CRFs by Sprot m by Uniref1	a <b>0.05</b> and <b>0.3</b> natcher	31072	3250595	5 <b>38597</b> m	iicrons/pixel	, within		
≥ ? 🚱 🗟 🆓 🔽 😵	Brows locatio Terms Advan <u>About</u> <u>Help</u>	e by cc ns infe and Im nced Se SLIF	ombinat rred fro nages arch	<u>tion of</u> om GO	Se 0 0 0	arch by proteir Protein names Please select- Protein names Please select-	extracted b extracted b extracted b	y CRFs ; y Sprot m	atcher							
in الاق الاق	Murphy La SLIF Web J Database: J	th SLIF S Application <u>pmc</u>	ervice on Version	n 2.1	0 [- N1	Protein names Please select-	extracted b  rds display	y Uniref1	00 matcher	)						
	per all the	windows o	connecting	g to the sar	ne server,	·····										

<b></b>	Safari	File	Edit	View	History	Bookma	rks Window	Help	Ð	S		<b>(93%</b> )	Sun 10:20	5 AM 🛞	Q,
. 74	000	)			_	_	http://slif.cbi	.cmu.edu/PMC/pro	otein_sea	rch.do	_	_	_	_	
	Subcell	ular Lo	ocation	Image 1	Finder										
	Home	2			Searche Protein	d for: name(s) (	GFAP, within								
1 × 🚱	Searc	<u>h by wo</u>	cel reso	lution	Images Captio	with pixel ons with th	l <b>resolution</b> bet he word(s) <b>tubu</b>	ween 0.05 and 0.3 lin.	31072325	059538	3597 micron	s/pixel, wii	thin		
	Searc	<u>h by pro</u>	otein na	ime	Results	1 - 7 of 7									_
· 🔚 X	Searc figure	h for FN s/panels	<u>VII</u> <u>s</u>		Ca	ation	Figure	Papal	Protein	Cell	Location Inferred	Location Inferred from	Resolution	FMI and Protein	
	Searc from	h by loc GO terr	cation ir ns	nferred		ption	rigure	1 anei	†↓	t ↓	Image ↑	GO Terms ↑ ↓	(µIII/pixel) ↑↓	Confidence	e
• 👜	Searc from	h by loc Images	cations	inferred	TNF not int	does erfere									
	Brow location Term	se by co ons infe s and In	ombinat rred fro nages	<u>tion of</u> o <u>m GO</u>	differer (A	cell ntiation									
	<u>Adva</u>	nced Se	arch		Immun of neur differer	ostaini onally ntiated			<u>GFAP</u>	N/A	Endosome	N/A	0.1	0.9823	
? . ? . ?	Help				neuros deri NSCs.	phere- ved Cells	Click to view in a new window	Click to view in a new window.							
0	Murphy L	ab SLIF S	Service				server when w								
	SLIF Web Database	Applicati pmc	on Version	n 2.I	not int with n stem	erfere eural cell									<b>A</b>
	Note: H .	aur brauer		w only one	Idifferer	ntiation	- 69 S - 69 S		1		I		1	1	



	Safari	File Edi	t View	History	Bookmarks	Window	Help		90		<b>?</b> ●	<b>(91%</b> )	Sun 10:30 AM	*	Q
70	00	_	_	_	_	http://slif.c	bi.cmu.ed	u/PMC/go_s	earch.do	)		_			
	Sarah Subcellu Home Search Search Search Search	by words i by pixel re by protein for FMI fig	on Image	Finder So G Su Els	earch by location O terms found ubcellular Loc umber of reco	http://slif.c	from GO to in names e -Please select yed per pa	u/PMC/go_s erms extracted by et ÷	earch.do earch.do CRFs Sprot Unire	ease se	lect			P	
× 🔆 📓 ? 🔰 💃 🏊 🖊 🎨 🔣 🧐 ? . ? . ? 🕑 🎯	Search GO ter Search Images Browse location Terms Advane About Help	by location ms by location e by combins inferred and Images ced Search SLIF	nation of from GO	from from	Search										
	Database: E	<u>omc</u>													4
	Note: If you	ir hrowser man	aees only on	cookie											1

Ű.	Safari	File Edit	View	History	Bookmarks	Window	Help		00		(91%)	Sun 10:30 AM	*	Q
	00					http://slif.c	cbi.cmu.edu/PM	C/go_se	arch.do					
	Safari Safari Solution	File Edit	View	Finder Finder S S S S S S S S S S S S S S S S S S S	Bookmarks earch by location O terms found ubcellular Loc umber of reconserved	Window http://slif.c	Help cbi.cmu.edu/PM from GO terms in names extract Please select promosome toplasm toplasmic Vesicle toskeleton tosol dosome tracellular embrane tochondrion uclear Membrane ucleoplasm ucleus ther Component bosome decidable	C/go_se	earch.do		(91%)	Sun 10:30 AM		Q
* ? ?	<u>About</u> <u>Help</u>	<u>SLIF</u>												
?	Murphy La SLIF Web /	b SLIF Service Application Versi	ion 2.1											
	Note: If you	ur browser mana	ees only one	cookie										
														1

Ś	Safari	File	Edit	View	History	Bookmarks	Window	Help		90	(îr	۲	(90%)	) Su	n 10:3	2 AM	*	Q
	$\odot \odot$	_	_	_			http://slif.c	bi.cmu.e	du/PMC/go_s	earch.do			_		_	_		
	<b>N</b> T	-																
5	<b>N</b>	┛┛																
	Subcellu	lar Lo	cation	Image	Finder													
	Home				Se	arched for:												
					Рі	otein name ex	tractor(s) S	Sprot, wi	ithin									
	Search	<u>by wo</u>	rds in c	caption	L	ocations inferr	ed from the	GO terr	n <mark>Golgi appa</mark> r	ratus.								
≪	Search	by pix	el resol	lution	R	esults 1 - 41 of	41											
	Search	by pro	tein na	me							 							٦
?	Search	for FM	II figur	ec/none			Link to Pl	R Quer	y			Li	nk to SL	IF Qu	iery			$\parallel$
	Seaten	101 111	II IIgui	es/pane			AN	K				Link	to SLIF	query:	ANK			
X	Search	by loc	ation ir	ferred f	rom		BS	<u>A</u>				Link	to SLIF	query	: <u>BSA</u>			
	<u>GO ter</u>	ms					<u>CD</u>	<u>K1</u>			 ]	Link	to SLIF o	uery:	CDK1			
	Search	by loc	ations i	nferred	from		<u>CO</u>	<u>X</u>				Link	to SLIF	query	: COX			
<i>.</i>	Images						<u>CO</u> 2	<u>X-2</u>			Ī	.ink t	o <mark>SLIF q</mark>	uery:	COX-2	2		
	Browse	e by co	mbinat	<u>ion of</u>			<u>C</u> ]	<u>P</u>				Lin	k to SLIF	quer	<u>y: CP</u>			
	location	ns infer	red fro	m GO			D	<u>P</u>				Lin	k to SLIF	quer	<u>y: DP</u>			
L.	<u>Terms</u>		ages				EG	θF				Link	to SLIF	query	: EGF			
2	Advan	ced Sea	arch				<u>F</u>	1				Lin	k to SLI	7 quer	<u>y: F1</u>			
	About	SLIF					<u>F</u>	<u>2</u>				Lin	k to SLI	7 quer	<u>y: F2</u>			
2	Help						<u>FG</u> I	<u>F-1</u>			]	Link	to SLIF o	uery:	FGF-1			
2	110.10						GAP	<u>'DH</u>			L	ink to	o <mark>SLIF q</mark> u	iery: (	GAPDI	H		
?							HE	K				Link	to SLIF	query	: <u>HEK</u>			
B	Murphy Lai	b SLIF Se	ervice				HS	<u>ss</u>				<u>Link</u>	to SLIF	query	: <u>HSS</u>			
9	SLIF Web A	Applicatio	n Versior	12.1			HSu	<u>lf-2</u>			Ī	ink t	o SLIF q	uery:	HSulf-2	2		
	Database: p	ome					HT	H				Link	to SLIF	query	HTH			
	Note: If you	ir hrowse	r manave	s only one	cookie		Hu	R				Link	to SLIF	query	: HuR			

	Safari	File	Edit	View	History	Bookmarks	Window	Help		Ð	S 🔳	$\bigcirc \bullet$	<b>(91%</b> )	Sun 11:42 PM	*	Q,
. 19	$\Theta \odot \odot$		_	_		htt	o://slif.cbi.o	cmu.edu/	PMC/scaleba	ar_sea	rch.do	_			_	
•			F	Image	Finder											
() () () () () () () () () () () () ()	Home			Intage	• muer • muer	Search by pix	el resolution	n, <i>within</i> 1 betweer	n <b>0.2</b> and <b>1.0</b>	micro	ns/pixel.					
	Search	<u>by wo</u>	ords in o	caption	0	Search by pix	el resolution	n			1					
20	Search Search	<u>by pix</u>	tein na	<u>intion</u>	 Th	e min resolutio	on is 12 Mic	rons/Pixe	el and the ma	x is 0.	.05 Miero	ons/Pixel				-
·⊡ ▼	Search	for FN	<u>/II figu</u>	res/pane		High: 0.05 ~ 0	0.311									
	GO ter	<u>ms</u>	ation ir	iferred t	rom 0	Low: 1.931 ~	12									
?	Search Images	<u>i by loc</u> §	ations i	inferred	from •	Other: 0.2	~ 1.0									
<u>ج</u> .	Brows locatio Terms	e by co ns infe and In	ombinat rred fro nages	<u>tion of</u> om GO	Ni	umber of reco	rds display	ed per p	age: 10 🛟							
• 🚳	Advan	iced Se	arch		e	earch										
• 🍪 ?	About Help	SLIF														
?																
<b>B</b> (0)	Murphy La	b SLIF S	ervice	n 2 I												
	Database:	pmc	ra veraior	n eu	coakie											4
	nor all the	ur prowse	r manage	a to the car	COOKIE											1

	Safa	ri File	Edit	View	History	Bookmarks	Window Help	Ð	S 🗖	. 🔿 🔹 🗠	⊡ (91%) Su	n 11:43 PM	* Q
-	0	0				http:/	//slif.cbi.cmu.edu/l	PMC/scalebar_sea	rch.do				
	Subce	ellular I	ocation	दे 1 Image	Finder								
	Hor	me	Sea	urched fo	r:								
S	Sea	rch by	Ima	ages with	n pixel resol	lution between	0.2 and 1.0 microns	/pixel.					
×	wor capt	<u>rds in</u> tion	Res	sults 1 - 1	10 of 5390								
	Sea pixe	<u>rch by</u> el		ext 10 Res	ults								
	reso Sea prot	olution rch by tein nam	e	Car	otion	Figure	Panel	Protein 🕇 ↓	Cell Type ↑↓	Location Inferred from Image ↑↓	Location Inferred from GO Terms †	Resolution (µm/pixel) ↑↓	FMI Confiden ↑↓
? 3	Sea FM figu	<u>rch for</u> II ires/pane irch by	els gr	Express HDGF member anular co cereber	sion of family ers in ells of the llum.		E: NeuN						
人 感	infe GO Sea	erred from terms terms		ntibodie DGF (A) ), HRP- NeuN (E	s against ), HRP-2 3 (G) and 3, E, H)			HDGF	N/A	ER	Cytoplasm	0.2375	0.9996
? · ? · ?	loca infe Ima	ations erred from ages	m w the sli	ere used e protein ces of ac Double s	to detect s in brain lult mice. tainings	<u>a new window</u>	<u>a new window.</u>						
20	Bro com	owse by abination	1 de	emonstra Express	te that in 💽	-							
	of lo infe	ocations erred from	<u>n</u>	HDGF membe	family ers in								
	and	Images	8	anular ce	lls of the llum.		E: NeuN						

Ś	Safari	File Edit	View	History	Bookmarks	Window	Help		00		(91%)	Sun 11:43 PM	*	Q
74	$\Theta \odot \odot$	_	_	_	H	nttp://slif.c	bi.cmu.edu	u/PMC/loc_m	natrix.do	_				_
	SI Subcellu	lar Locatio	⊴ n Image	Finder										
	Home			) In	Generate Loc ages with pixe	ation Comp l resolution	parison Ma n between	ttrix, <i>within</i> <b>0.2</b> and <b>1.0</b> i	microns/pi	ixel.				
5	Search	by words in	n caption	0	Generate Loc	ation Comp	parison Ma	ıtrix						
20	Search	by pixel res	solution	-	20									-
	Search	by protein	name	e										
ź	Search	for FMI fig	ures/pane	<u>els</u>										
X	Search GO ter	by location ms	inferred t	from										
2	Search Images	by location	s inferred	from										
**	Brows location	e by combir ns inferred f	nation of from GO											
	Terms	and Images												
<b>5</b>	Advan	ced Search												
2	<u>About</u>	<u>SLIF</u>												
?	<u>Help</u>													
?	Murphy La	b SLIF Service												
9	SLIF Web A Database: [	Application Vers	ion 2.1											
	Note: If you per all the y	ar browser mana vindows connect	nges only one ting to the sa	r cookie me server,										▼ //:

http://slif.cbi.cmu.edu/PMC/loc\_matrix.do

Ð

C

Help

4

9

00

F

2

2

(

#### Subcellular Location Image Finder

Home	GO Terms \ Images	Endosome	ER	Golgi	Lysosome	Mitochondria	Nucleolus	Tubulin	Unknown
Search by words in caption	Cytoplasm	78	23	1	1	20	2	1	6
Search by words in capiton	Cytoskeleton	4	2	0	1	1	0	0	1
Search by pixel resolution	Cytosol	6	0	0	0	0	0	0	0
Search by protein name	ER	7	0	0	0	0	0	0	0
Search for FMI figures/panels	Extracellular	0	0	0	0	1	2	0	0
Source in the inguies, purchs	Golgi apparatus	1	0	0	0	0	0	0	0
Search by location inferred from GO terms	Membrane	18	11	0	0	8	2	1	2
<u>oo tamis</u>	Mitochondrion	15	6	0	2	15	0	0	1
Search by locations inferred from	Nucleus	81	27	1	1	27	0	1	7
magoo	Other Component	5	5	0	0	1	0	0	0
Browse by combination of locations inferred from GO	Undecidable	0	1	0	0	1	0	0	0
Terms and Images									

Advanced Search

About SLIF

Help

Murphy Lab SLIF Service

SLIF Web Application Version 2.1 Database: pmc

Note: If your browser manages only one cookie per all the windows connecting to the same server,

Ś	Safa	ari Filo	e Eo	dit Vie	w History	/ Bo	ookmarks	Window	Help		Ð	C 🔳 <	$\bigtriangledown$	(93%)	Sun 11:55 P	M 🛞	Q
. 74	0	) 🔘	_				ht	ttp://slif.c	bi.cmu.eo	du/PMC/loc_r	natrix.	do	_				_
	Subo	cellular I	Locat	tion Ima	ge Finder												
	Но	ome		Searche	l for:												
1 ×		arch by ords in ption	1	Locatio Locatio Images	ns inferred fi n(s) inferred with <b>pixel r</b>	rom th I from r <b>esolu</b>	ne GO term in the images tion betwee	Mitochon Mitochon en 0.2 and	drion, wi dria, wit 1.0 micro	<i>ithin</i> hin ons/pixel.							u
· []	<u>Sea</u> pix res	arch by (el colution	]	Results	Results									[		1	
×	Se pro	arch by otein nan arch for	<u>ne</u>		Caption		Figure	Pa	anel	Protein ↑ ↓	Cell Type ↑↓	Locatio Inferro from Im	on ed age	Location Inferred from GO Terms †	Resolution (µm/pixel) ↑↓	FN Confic 1	fI lence ↓
) 🔂 🚱 🔄 🆓 🎢 🦿 📽 🧲 -	FM fig Sec loc inf GC Sec loc inf Im Br cor	<u>41</u> <u>ures/pan</u> <u>arch by</u> <u>cation</u> <u>erred fro</u> <u>arch by</u> <u>cations</u> <u>cerred fro</u> <u>ages</u> <u>owse by</u> <u>mbinatio</u>	m m	Muscle Gasti Mu Trk Stain Tibia P0(A) and S immun cross gasti mus The musc	Spindles in ocnemius scle and A/TrkC ing in the I Nerve at NF-M (red) 46 (green) ostaining in section of ocnemius cle at P0. re are no e spindles		lick to view new windov	in v. <u>Click ta</u> <u>a new</u>	view in window.	<u>E17</u>	N/A	Mitochor	ndria	Mitochondrion	0.546667	0.98	68
	of inf <u>GC</u> and	locations erred fro D Terms d Images	<u>m</u>	Muscle Gast Mu Trk	Spindles in ocnemius scle and A/TrkC ing in the		Fernandonuman Anto Patri Anton 2010										

8	<b>Safari</b> File E	dit View History	Bookmarks W	indow Help		£		> 🔹 🖅 (93%	Sun 11:5	5 PM 🖇 C	2
	$\bigcirc \bigcirc \bigcirc$		http:	//slif.cbi.cmu.ed	u/PMC/lo	oc_matr	ix.do				
	SLD	F									
<b>.</b>	Subcellular Loca	ition Image Finder									
	Home	Searched for:									
	<u>Search by</u> words in caption	Locations inferred fro Location(s) inferred f Images with pixel re	m the GO term M from the images M solution between (	itochondrion, with (itochondria, with 0.2 and 1.0 micro	<i>ithin</i> hin ns/pixel.						
2	Search by pixel resolution	Results 11 - 20 of 32 Next 10 Results									
	Search by protein name Search for	Caption	Figure	Panel	Protein ↑↓	Cell Type ↑↓	Location Inferred from Image ↑↓	Location Inferred from GO Terms ↑	Resolution (µm/pixel) ↑↓	FMI Confidence ↑↓	
	<u>FMI</u> figures/panels <u>Search by</u> location inferred from <u>GO terms</u> <u>Search by</u> locations inferred from <u>Images</u> <u>Browse by</u> combination	Immunohistochemic localization of OCTs in murine bronchi. OCT1- immunolabelling is localized to the apical membrane of ciliated epithelial cells in wild-type FVB mice (arrows in A). The specificity of this labelling is	ST S	OCT1 A sm Click to view in a new window.	OCT1	N/A	Mitochondria	Mitochondrion	0.75	0.9115	
	of locations inferred from GO Terms and Images	Immunohistochemic localization of OCTs in murine bronchi. OCT1- immunolabelling is	SCT and SCP.SCTUP	preab							

¢

Ś	Safari	File	Edit	View	History	Bookmarks	Window	Help	4	90		$\bigcirc \bullet$	(100%)	Mon 7:55 AM	*	Q
7	00	_	_	_		http	o://slif.cbi.c	mu.edu/P	MC/advanced	_search	.do				_	
	SI Subcellu	lar Loo	cation	Image 1	Finder											
	<u>Home</u>				A	dvanced Search	n									_
	Search	by wo	rds in c	aption	E	Search by wo	ords in capti	on								
21	Search	by pixe	el resol	ution												
?	Search	for FM	II figur	es/pane	<u>ls</u>	Search by pro	otein name									-
X	<u>Search</u> GO ter	<u>by loca</u> ms	ation in	ferred f	rom C	Protein names	extracted b	y CRFs								
<b>2</b>	Search Images	by loca	ations i	nferred	from	Protein names	extracted b	v Sprot m	atcher							
2%	Brows location	e by con ns infer	mbinat red fro	<u>ion of</u> m GO				, -1								
人	Advan	ced Sea	ages arch		e	Protein names	extracted b	y Uniref1	00 matcher							
	About	<u>SLIF</u>														
??	<u>Help</u>				E	Search for FM	/II figures/pa	anels								
?	Murphy La	b SLIF Se	rvice		0	) figure or 🔘 p	anel									_
	SLIF Web / Database: p	Applicatio omc	n Version	2.1	E	Search by pix	el resolutio	n								
1	Note: If you	ur browser	manage	s only one	cookie	he min resolutio	on is 12 Mic	crons/Pixe	l and the max	is 0.05	Micro	ons/Pixel				<b>v</b>

# Linking to SLIF from another website

- Queries against the database can be made from outside the website using http:// slif.cbi.cmu.edu/PMC/search.jsp?arguments
- The arguments are:
  - protein=<protein name>
  - level=figure OR level=panel
  - type=FMI (NOTE that BOTH level and type must be present if either is present)
  - pixel\_size\_lo=<lower bound>
  - pixel\_size\_hi=<upper bound> (NOTE that both upper and lower bounds must be specified)
- Carnegie Mellon location=<subcellular location>

# SOAP interface - query DTD

<!ELEMENT slif query (protein name, fmi level, pixel res lower bound, pixel res upper\_bound, subcellular location)> <!ELEMENT protein name (#PCDATA)> <!ELEMENT fmi level EMPTY> <!ATTLIST fmi level figure\_or\_panel (figure|panel) #REQUIRED> <!ELEMENT pixel res lower bound (#PCDATA)> <!ELEMENT pixel res upper bound (#PCDATA)> <!ELEMENT subcellular\_location (#PCDATA)>

### SLIF web client

0 0	SLIF Search	
	Search Criteria	
Protein Name	GO Location	Words in Caption
FMI Level	Max. Image Res.	Min. Image Res.
panel	•	
Article Id		
		Search SLIF
	Results	
		) •   )
		Save Results

### **SLIF** web client options

- -gui launches GUI search utility
- -p protein name
- -il image level (figure or panel)
- -go GO location
- -hr highest image resolution
- -Ir lowest image resolution
- -id article id
- -w words in caption

### java -jar SLIFPMCSearch.jar -lr 0.2 – hr 2.0 -il panel

<?xml version="1.0"?>

<!DOCTYPE slif\_query\_results SYSTEM "http://slif.cbi.cmu.edu/ slif\_result.dtd">

<slif\_query\_response>

<<u>query></u>

<imi\_level>panel</fmi\_level>

<s<mark>cale>0.2,2.0,</scale></mark>

</query>

<result\_summary>

<num\_records>8202</num\_records>

</result\_summary>

<query\_results>

<slif\_result>
Carnegie Melonite

### java -jar SLIFPMCSearch.jar -lr 0.2 – hr 2.0 -il panel

<slif\_result>

<caption>notch and delta gene expression in zebrafish ...

- <figure\_url>p005-paper/paper4439-figure/figure0001image\_box.jpg</figure\_url>
- <panel\_url>p005-paper/paper4439-figure/figure0001-panel/ panel\_6\_green-image.jpg</panel\_url>

<protein\_name>n1b</protein\_name>

<cell\_name>N/A</cell\_name>

<subcellular\_location>Endosome</subcellular\_location>

<pixel\_resolution>1.6</pixel\_resolution>

</slif\_result>

### java -jar SLIFPMCSearch.jar –p cadherin –il panel

<?xml version="1.0"?> <!DOCTYPE slif\_query\_results SYSTEM "http://slif.cbi.cmu.edu/ slif result.dtd"> <slif\_query\_response> <query> <protein>cadherin,</protein> <fmi\_level>panel</fmi\_level> </query> <result summary> <num records>8</num records> </result\_summary> <query results>

Carnegie Mellon <slif\_result>

# Summary

- Complex pipeline for parsing figure and caption pairs
- Web pages to provide interactive access to results of analysis
- Web service for programmatic access
- Work continues...



# Acknowlegments



Commonwealth of Pennsylvania Tobacco Settlement Funds NIH R01 GM078622



William Cohen

**Eric Xing** 

### Past:

Meel Velliste, Greg Porreca, Zhenzhen Kou, Juchang Hua Current:

Amr Ahmed, Andrew Arnold, Luis Coelho, Saboor Sheikh