



# #FigTag

Find Papers with the Data You're Looking For!



# The Team

- Meng Cheng
- Ryan Connor
- Marie Gallagher
- Alex Kotliarov
- David Shao
- Ricardo V.

# The Problem

- ***As a researcher:***
  - I look for articles in pubmed
    - That have data that speak to a hypothesis or problem I have

**BUUUUUUT ...**

- ***What I get***
  - Lots of articles that contain my keywords but with few figures about data “with” all my keywords

# The Solution:

## *Index Papers based on Figure Attributes*

### **Image Processing**

- Split Multipanel Figures
- Extract Text from Figures
- Cluster and Classify Figures

### **Text Mining**

- Map text to MeSH Hierarchy
- Sources
  - Figure Legend
  - Related Material & Methods
  - Relevant Results

## Search results

Items: 1 to 20 of 857

&lt;&lt; First &lt; Prev Page 1 of 43 Next &gt; Last &gt;&gt;

- ☐ [PSGL1-deficient mice develop spontaneous pulmonary hypertension associated to systemic sclerosis.](#)

González-Tajuelo R, de la Fuente-Fernández M, Morales-Cano D, Muñoz-Callejas A, González-Sánchez E, Silván J, Serrador JM, Cadenas S, Barreira B, Espartero-Santos M, Gamallo C, Vicente-Rabaneda EF, Castañeda S, Pérez-Vizcaíno F, Cogolludo Á, Jiménez-Borreguero LJ, Urzainqui A. Arthritis Rheumatol. 2019 Sep 11. doi: 10.1002/art.41100. [Epub ahead of print]

PMID: 31509349 **Figs 1,3 of 5**

[Similar articles](#)

- ☐ [Acute Myeloid and Lymphoblastic Leukemia Cell Interactions with Endothelial Selectins: Critical Role of \*\*PSGL-1\*\*, CD44 and CD43.](#)

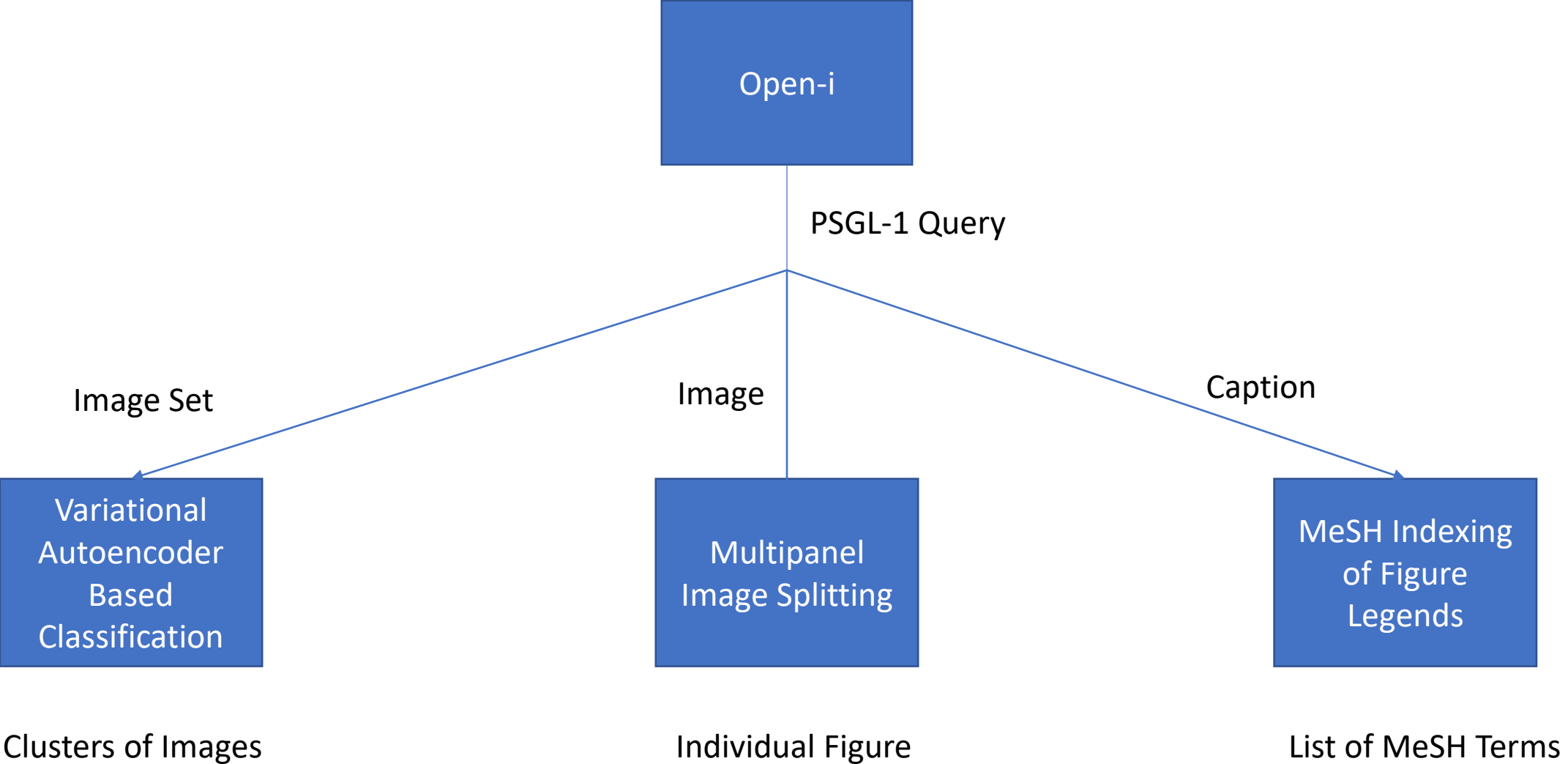
Spertini C, Baïsse B, Bellone M, Gikic M, Smirnova T, Spertini O. Cancers (Basel). 2019 Aug 27;11(9). pii: E1253. doi: 10.3390/cancers11091253.

PMID: 31461905 **Free Article**

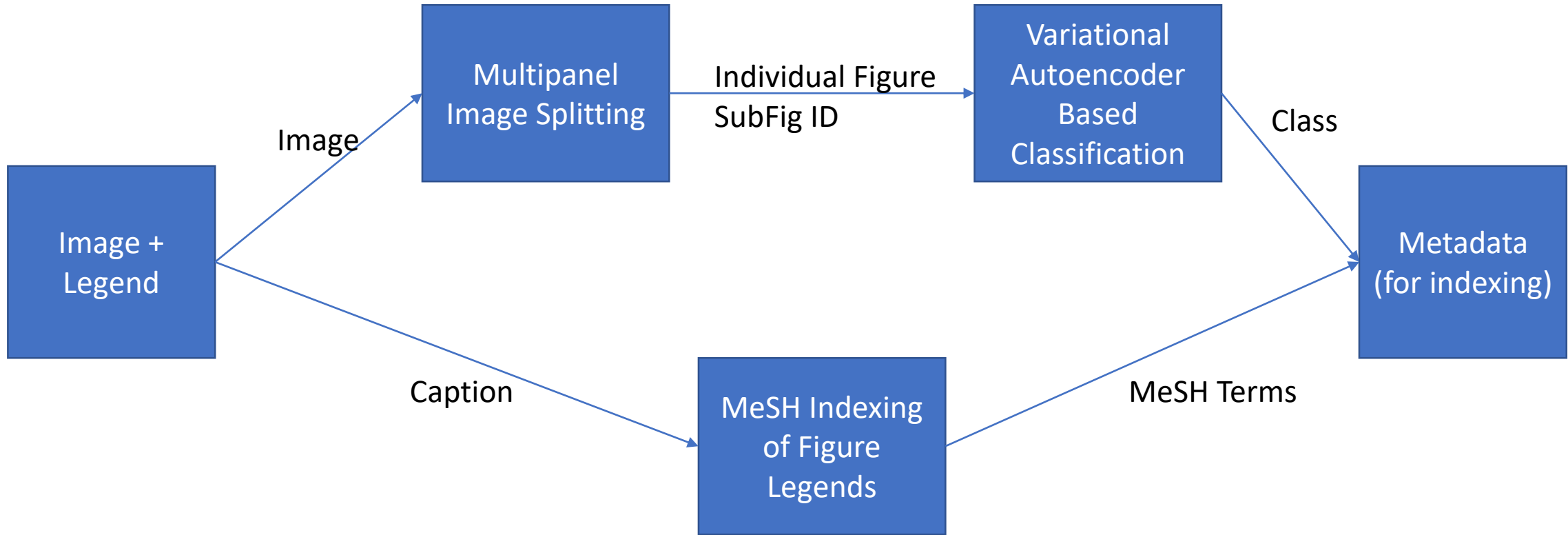
[Similar articles](#)

**Figs 2 of 4**

# Development Pipeline



# Product Pipeline



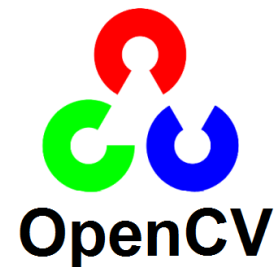
# Split multi-panel figure

Why?

- Better input for image clustering
- More granular mesh indexing for individual sub image

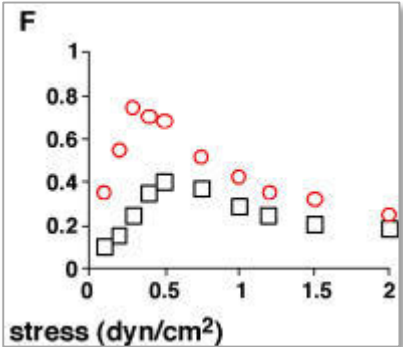
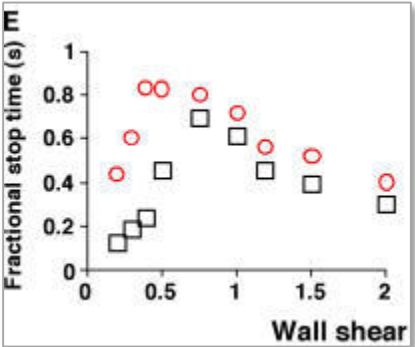
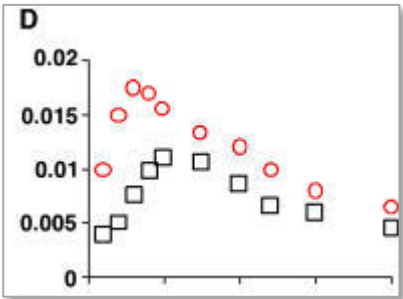
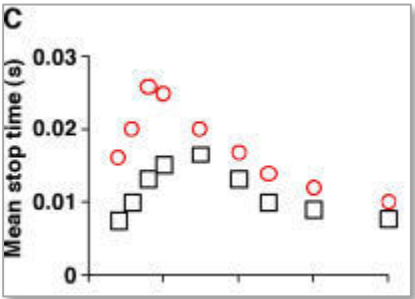
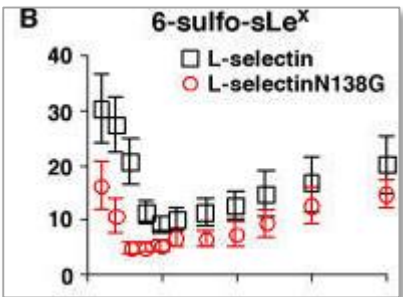
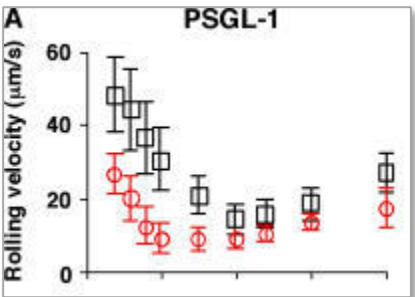
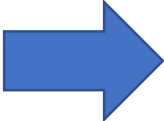
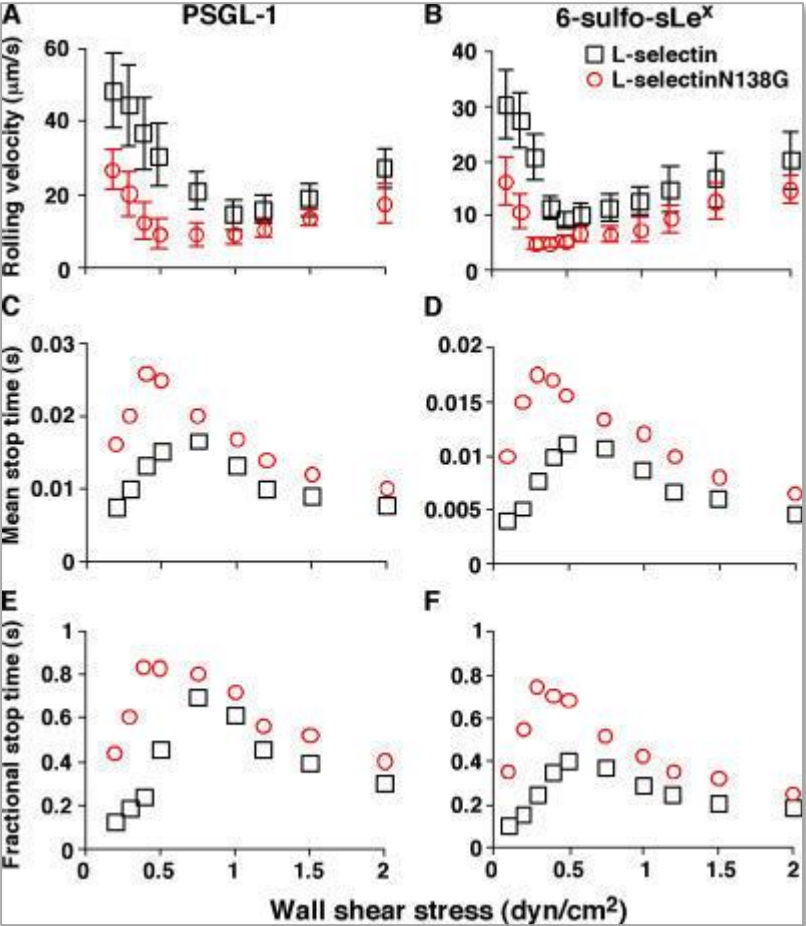
How?

- Use OpenCV library
- Horizontal/Vertical projection
- Assume sub panels are arranged in grid

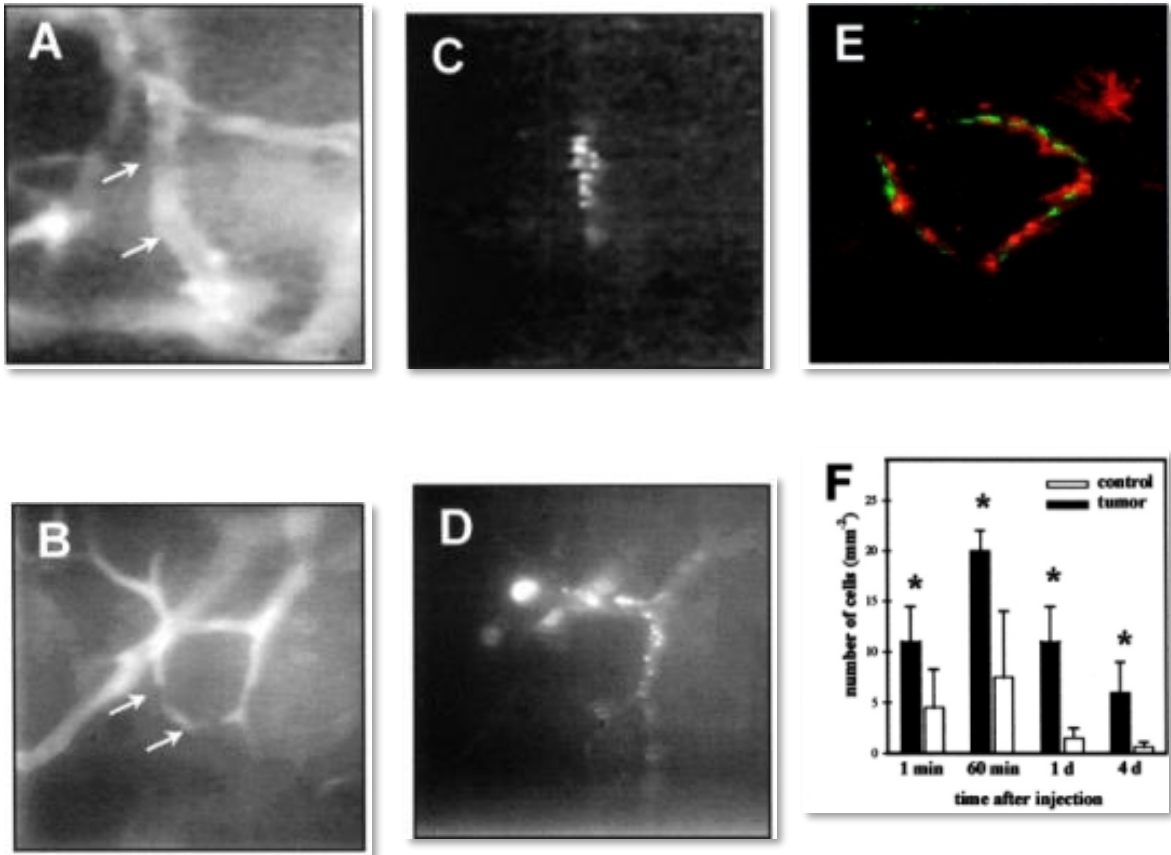
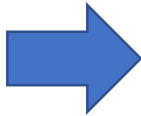
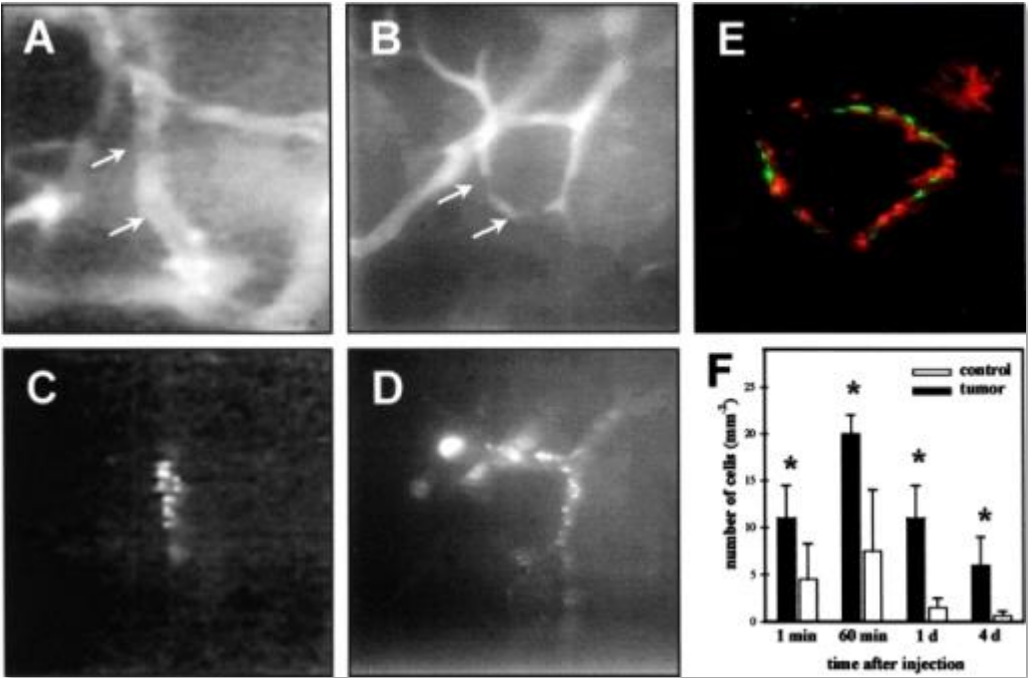




Work when sub panels are arranged in grid



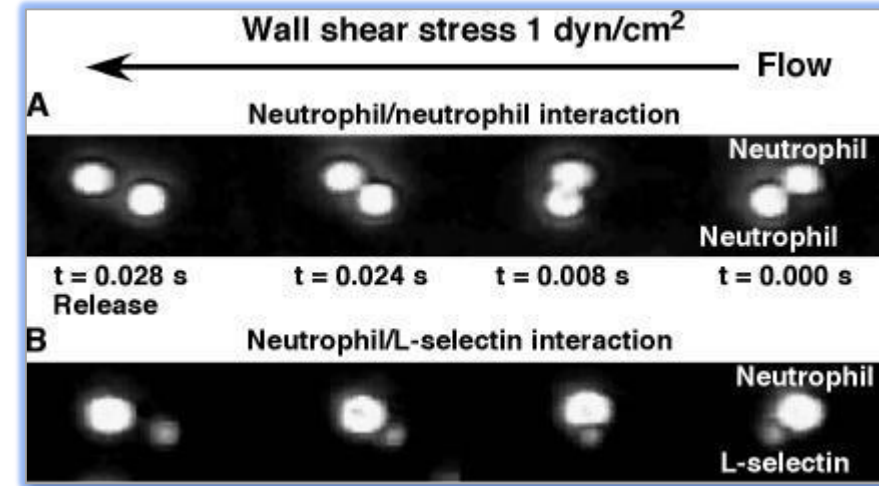
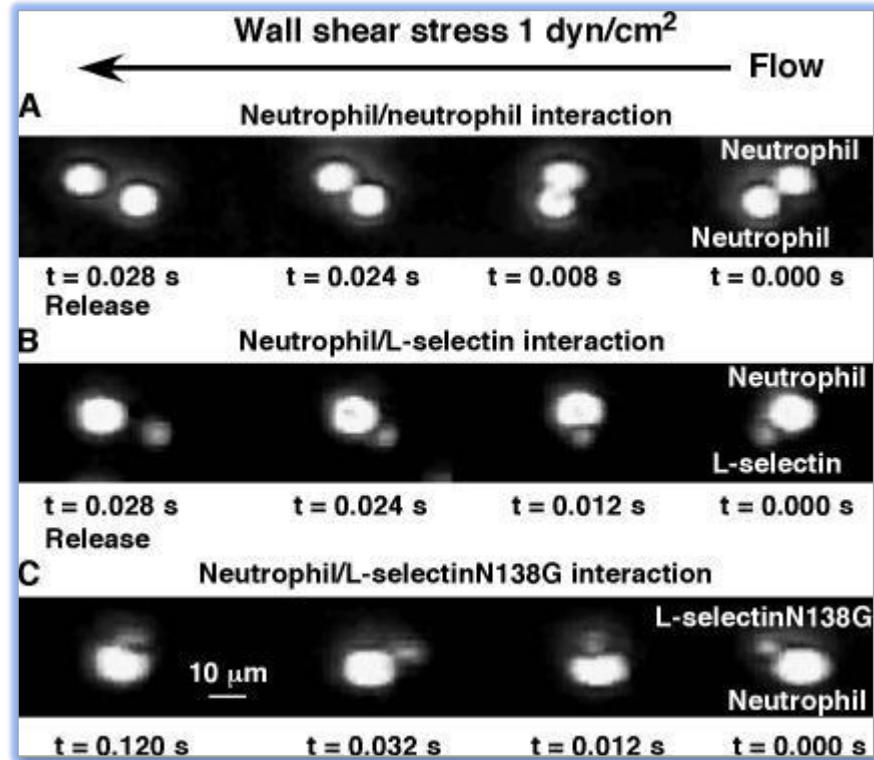
Heterogeneous sub panels



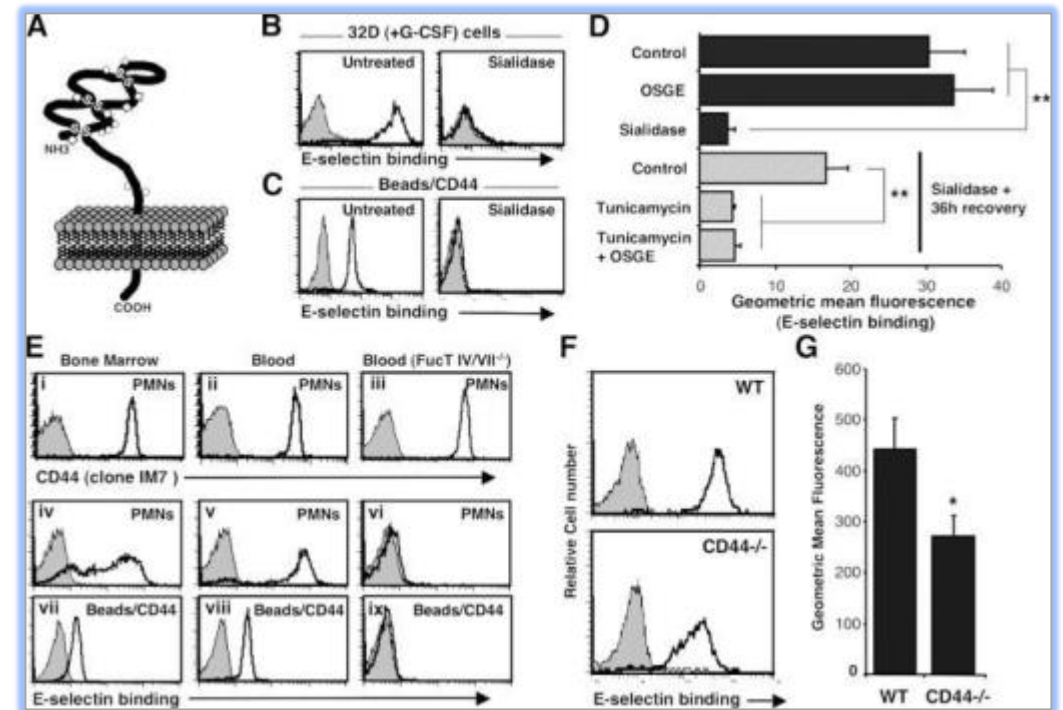
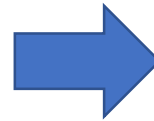
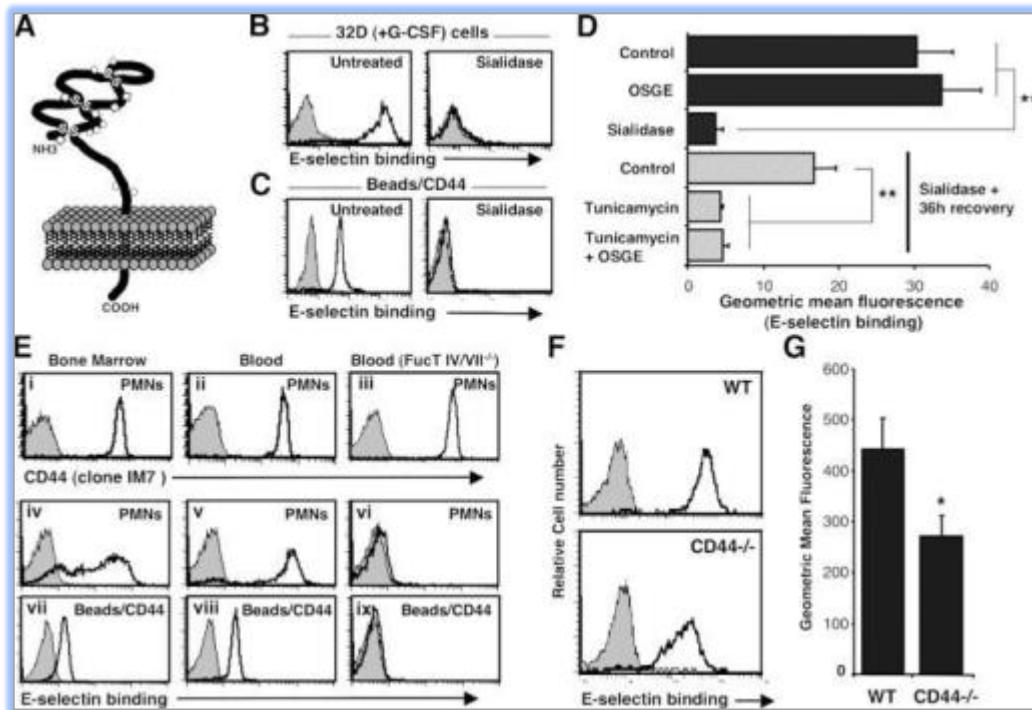
Notice now the bar chart F is a separate image

## Cut at the wrong places!

- Sub panels and labels are not well separated

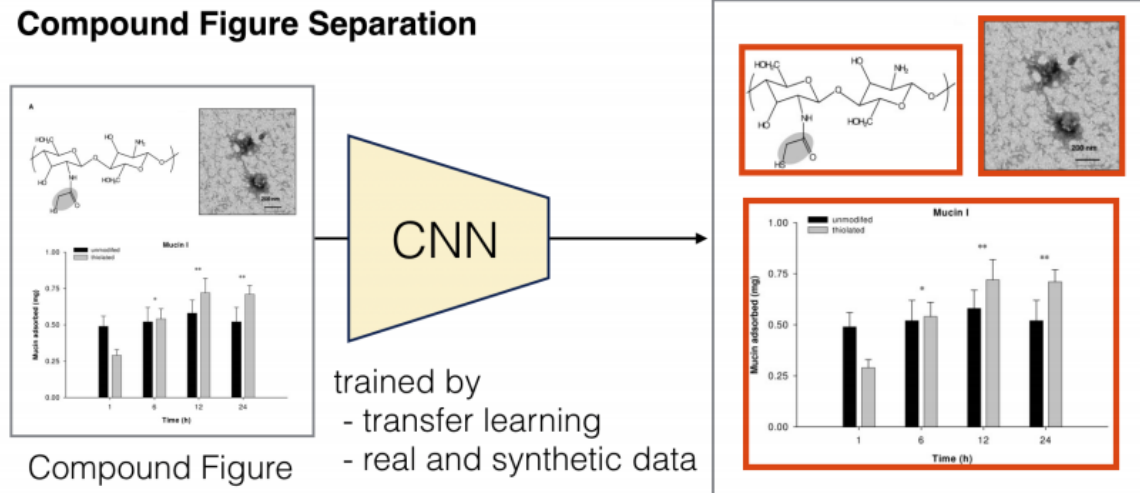


Does not detect sub panels if they are not aligned



# Potential Improvement

- Use AI/ML methods to detect and split multi-panel figures.
  - E.g. *A Data Driven Approach for Compound Figure Separation Using Convolutional Neural Networks* (Tsutsui, et al, 2017)



# Mesh term indexer



Affinity isolation of PSGL-1 glycoprotein from platelets and neutrophils. (A) Purified preparations of neutrophils or platelets were biotinylated and lysed. Cell lysates were incubated with P-selectinIgG ... ..

Male  
Mice  
Animals  
Humans  
SELP protein, human  
P-Selectin  
Blood Platelets  
DNA Primers  
Ligands  
Membrane Glycoproteins  
... ..

### MeSH terms

[Animals](#)  
[Antibodies, Monoclonal](#)  
[Base Sequence](#)  
[Blood Platelets/metabolism\\*](#)  
[Blood Platelets/physiology](#)  
[Blood Platelets/ultrastructure](#)  
[DNA Primers/genetics](#)  
[Endothelium, Vascular/physiology](#)  
[Gene Expression](#)  
[Humans](#)  
[Leukocytes/metabolism](#)  
[Ligands](#)  
[Male](#)  
[Membrane Glycoproteins/blood\\*](#)  
[Membrane Glycoproteins/genetics](#)  
[Membrane Glycoproteins/immunology](#)  
[Mice](#)  
[Mice, Inbred C57BL](#)  
[Microscopy, Immunoelectron](#)  
[P-Selectin/blood\\*](#)  
[Platelet Activation](#)  
[RNA, Messenger/blood](#)  
[RNA, Messenger/genetics](#)

MeSH terms associated with article

```
"MeSH": {  
  "minor": [  
    "Animals",  
    "Antibodies, Monoclonal",  
    "Base Sequence",  
    "DNA Primers/genetics",  
    "Endothelium, Vascular/physiology",  
    "Gene Expression",  
    "Humans",  
    "Leukocytes/metabolism",  
    "Ligands",  
    "Male",  
    "Mice",  
    "Mice, Inbred C57BL",  
    "Microscopy, Immunoelectron",  
    "Platelet Activation",  
    "RNA, Messenger/blood/genetics"  
  ],  
  "major": [  
    "Blood Platelets/metabolism*/physiology/ultrastructure",  
    "Membrane Glycoproteins/blood*/genetics/immunology",  
    "P-Selectin/blood*"   
  ]  
},
```




Open Access Biomedical Image  
Search Engine


Search by text or dropping an image.



Article's MeSH terms snippet from the Open-i API




**Indexing Initiative**  
 U.S. National Library of Medicine



[Home](#) > [NLM Medical Text Indexer \(MTI\)](#)

### Information & Resources

- [About Indexing Initiative](#)
- [Datasets & Test Collections](#)
- [MEDLINE Baseline Statistical Reports \(BSD link\)](#)
- [MEDLINE Co-Occurrence File \(MRCOC\)](#)
- [MEDLINE Baseline Repository \(MBR\)](#)
- [MetaMapped MEDLINE Baselines](#)
- [MUID to PMID Conversion Files](#)
- [Publications](#)

### Tools

- [Terms of Service](#) NEW
- Access to our Tools
  - [Batch Access](#)
  - [Interactive Access](#)
  - [Web API Access](#)
- [Medical Text Indexer \(MTI\)](#)
- [Phrase2MeSH](#)
- [MeSH on Demand \(MeSH link\)](#)
- [MetaMap](#)
- [MetaMapLite](#) NEW
- [Custom Taxonomy Builder](#) NEW
- [MTI ML \(Machine Learning Package\)](#)
- [Specialist Lexicon Information and Tools](#)
- [Semantic Knowledge Representation Web Site \(Retired\)](#)
- [SemRep/SemMedDB Access Web Site](#)

### Areas of Interest

- [Full Text Processing](#)
- [Structured Abstracts](#)
- [Word Sense Disambiguation \(WSD\)](#)

### NLM Medical Text Indexer (MTI)

The NLM Medical Text Indexer (MTI) combines human [NLM Index Section](#) expertise and Natural Language Processing technology to curate the biomedical literature more efficiently and consistently.

[Batch MTI](#)
[Interactive MTI](#)

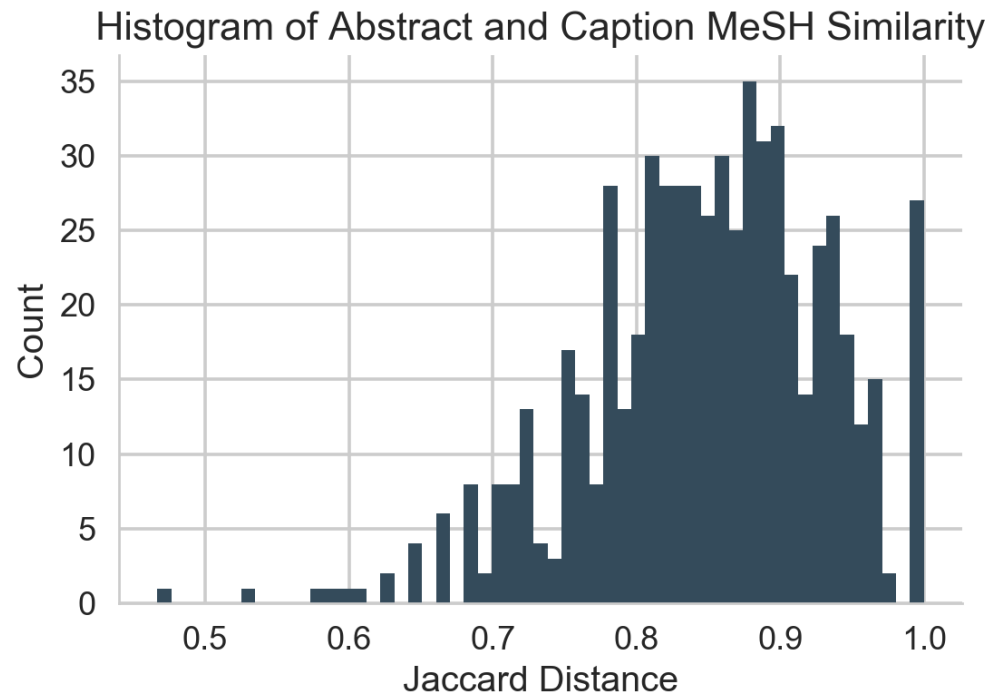
[+ Expand All](#)

- ▼ **What is MTI?**
- ▼ **What is MTIFL and how does it differ from MTI?**
- ▼ **Is there a recommended publication I should start with?**
- ▼ **Is there a detailed description of how MTI works?**
- ▼ **How can I use MTI?**
- ▼ **What does the MTI Logo look like?**
- ▼ **What about the rest of the MTI related publications?**
- ▼ **What does the MTI Indexing Life Cycle look like?**
- ▼ **What is the history of MTI?**

Indexing Initiative's NLM Medical Text Indexer Tools generated MeSH headings for each Caption in our dataset  
 Used MTI Batch Access first. Used MTI Interactive Access later.



# MeSH Indexing of Figure Legends Adds Value!



the Jaccard coefficient by

$$J_{\mu}(A, B) = \frac{\mu(A \cap B)}{\mu(A \cup B)},$$

and the Jaccard distance by

$$d_{\mu}(A, B) = 1 - J_{\mu}(A, B)$$

- 1115 Figures from 371 Papers
- 615 Figure for which both source paper and caption have MeSH terms (~55%)
- 588 Figures with at least 1 MeSH term in common between article and caption (~96%)

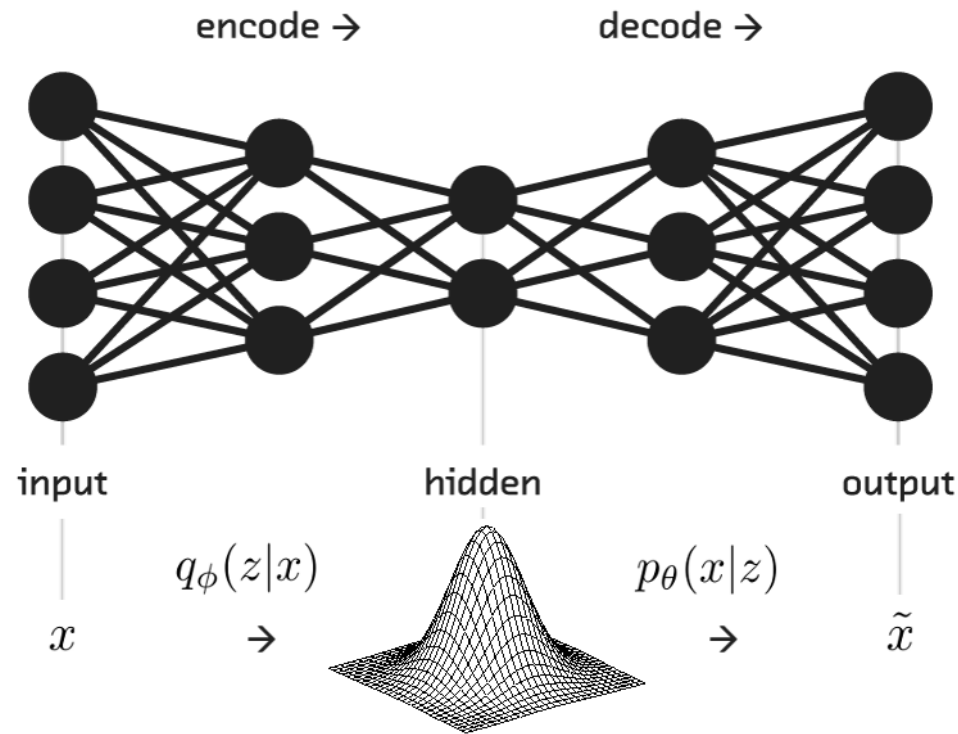
# Machine-Learning for Image Classification

Variational Autoencoder

==



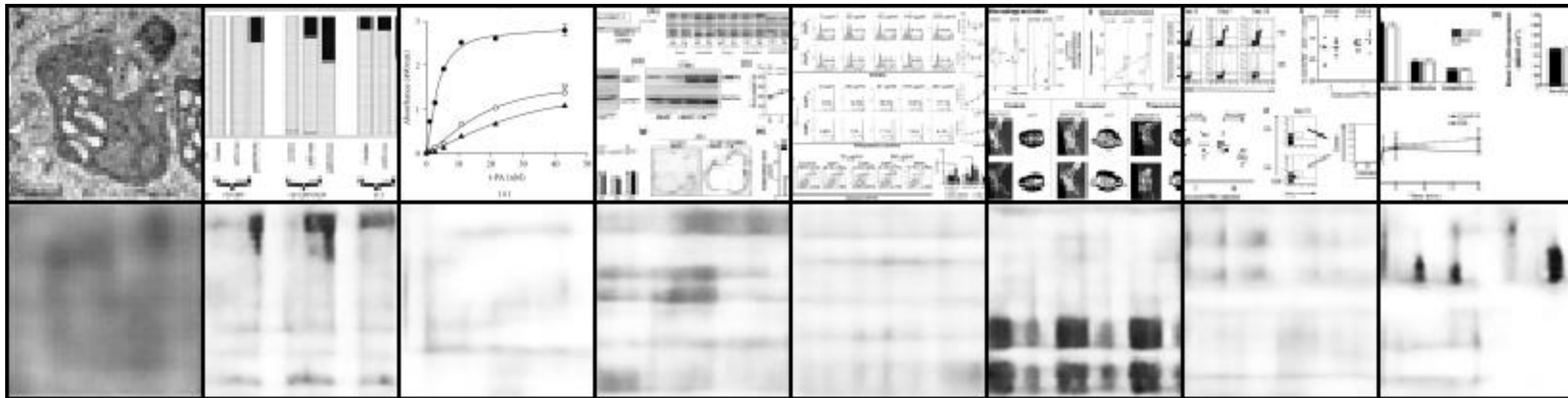
# What does Magic look like?



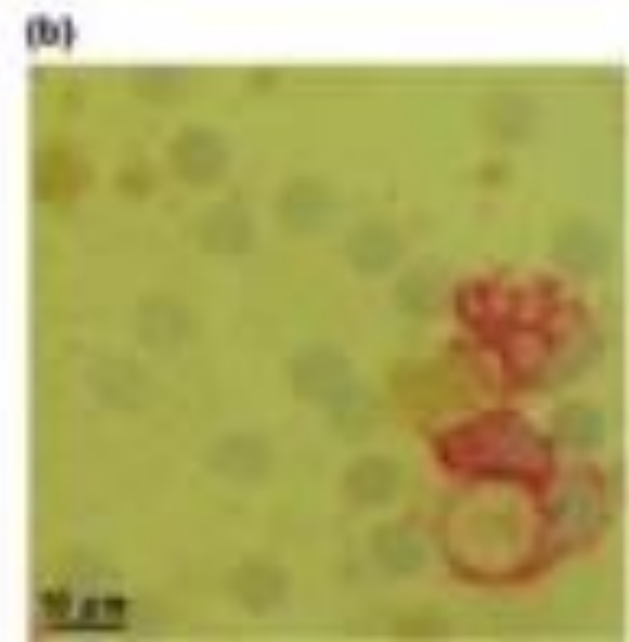
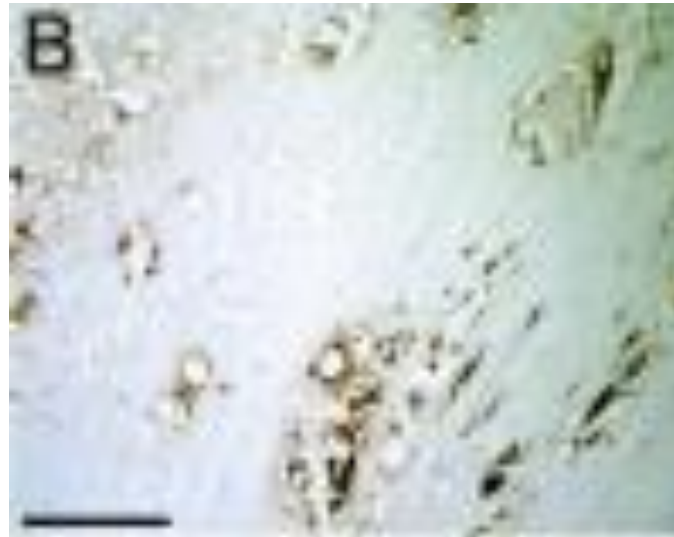
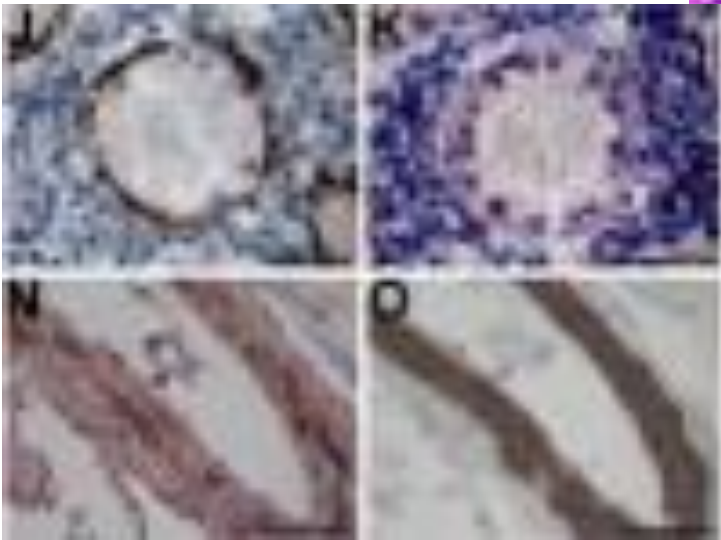
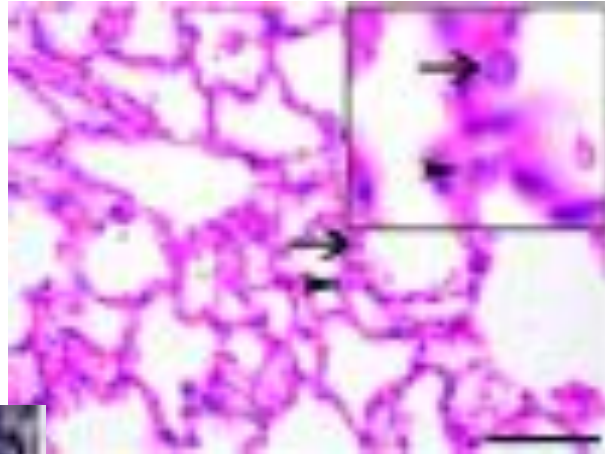
- In English?

- Variational Autoencoder model consists of encoder, decoder and a loss function.
- Encoder is a neural network that outputs a latent representation of an image - features of an image that represent a point in the D-dimensional feature space; The encoder serves as inference model.
- Decoder is a neural network that learns to reconstruct the data - input image - given its representation (latent variables).
- We will produce features vector for each image of a training set and will use these features to fit a KMeans clustering model and decide on number of clusters using "elbow" heuristic.

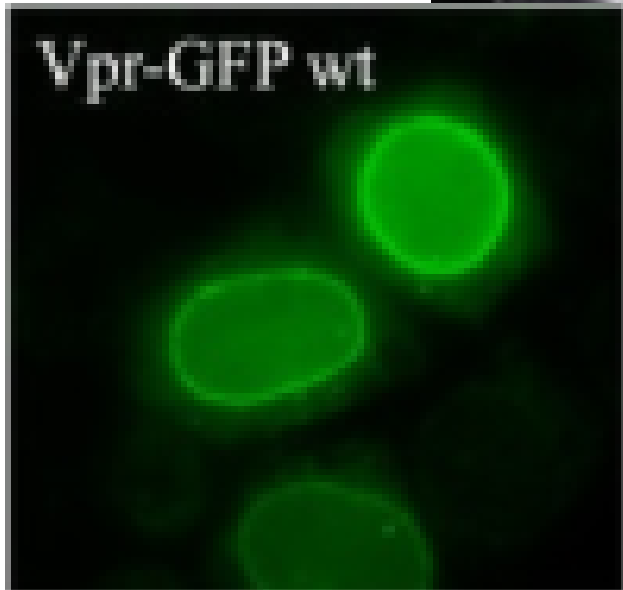
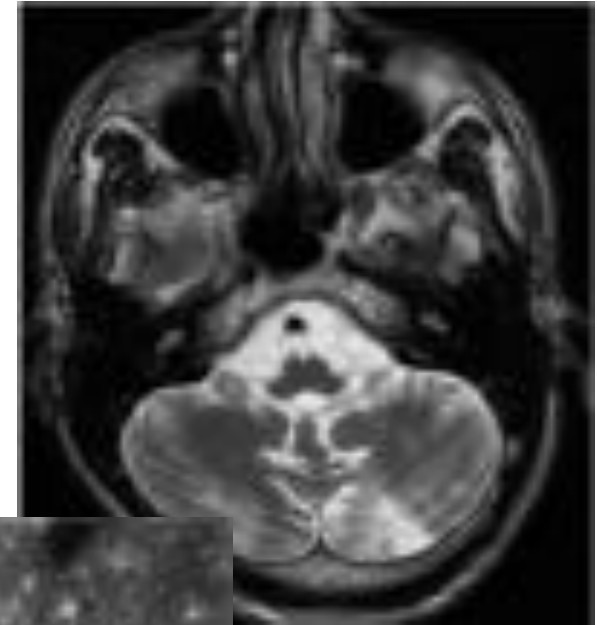
# Example Decoder Output



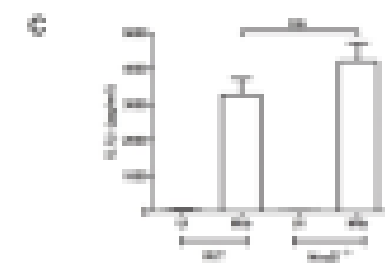
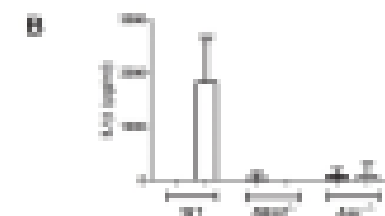
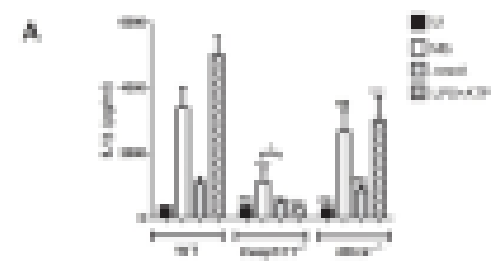
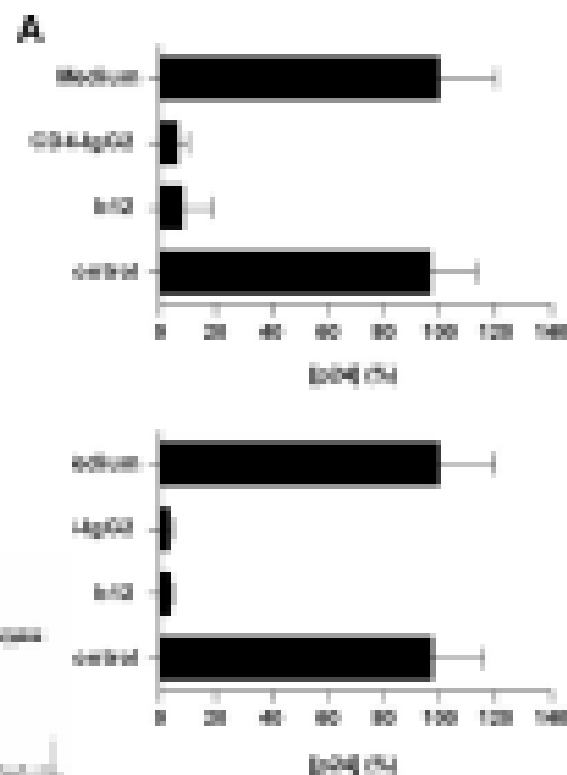
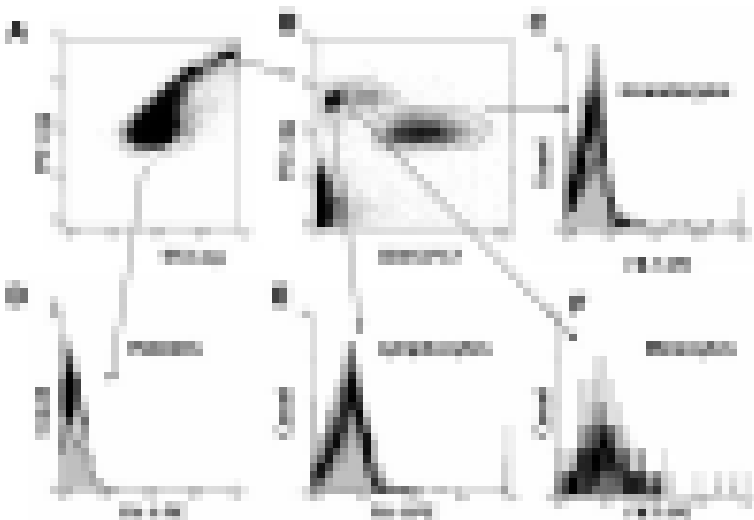
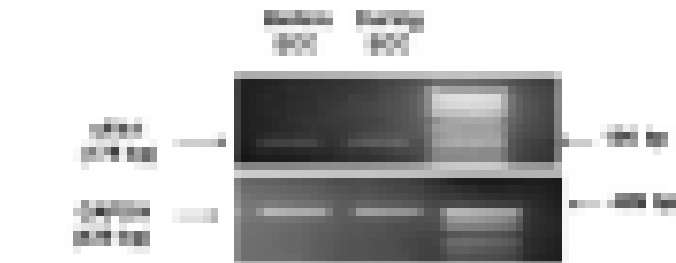
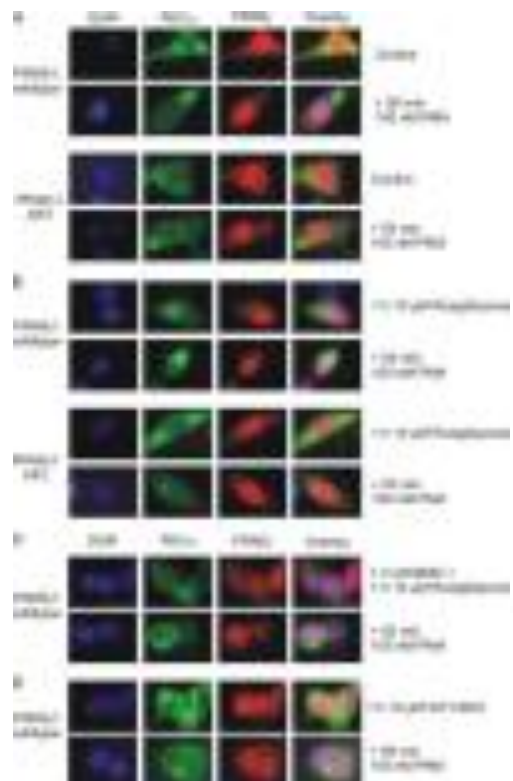
# Example Clusters



# Example Clusters



# Example Clusters



# Thank You

*Questions? Comments? Suggestions?*

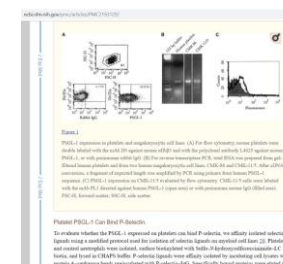


```

"uid": "PMC2193129",
"pmcid": "2193129",
"pmid": "10770806",
"docSource": "PMC",
"articleType": "ra",
"pmc_url": "http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2193129",
"pubMed_url": "http://www.ncbi.nlm.nih.gov/pubmed/10770806",
"image": {
  "id": "F1",
  "caption": "<b>PSGL-1</b> expression in platelets and megakaryocytic cell lines. (A) For flow cytometry, mouse platelets were double labeled with the mAb D9 against mouse  $\alpha$ IIb $\beta$ 3 and with the polyclonal antibody L4025 against mouse <b>PSGL-1</b>, or with preimmune rabbit IgG. (B) For reverse transcriptase PCR, total RNA was prepared from gel-filtered human platelets and from two human megakaryocytic cell lines, CMK-86 and CMK-11/5. After cDNA conversion, a fragment of expected length was amplified by PCR using primers from human <b>PSGL-1</b> sequence. (C) <b>PSGL-1</b> expression on CMK-11/5 evaluated by flow cytometry. CMK-11/5 cells were labeled with the mAb PL1 directed against human <b>PSGL-1</b> (open area) or with preimmune mouse IgG (filled area). FSC-H, forward scatter; SSC-H, side scatter."
},
"imgThumb": "/imgs/100/60/2193129/PMC2193129_JEM991708.f1.png",
"imgLarge": "/imgs/512/60/2193129/PMC2193129_JEM991708.f1.png",
"imgThumbLarge": "/imgs/137/60/2193129/PMC2193129_JEM991708.f1.png",
"imgGrid150": "/imgs/150/60/2193129/PMC2193129_JEM991708.f1.png",
"similarInCollection": "/search?simCollection=PMC2193129_JEM991708.f1&query=psgl-1%20OR%20sleplg&req=3",
"getArticleFigures": "/search?uid=PMC2193129&req=5",
"detailedQueryURL": "/search?img=PMC2193129_JEM991708.f1&query=psgl-1%20OR%20sleplg&it=x,u,ph,p,mc,m,g,c&coll=pmc&req=4",
"similarInResults": "/search?simResults=PMC2193129_JEM991708.f1&query=psgl-1%20OR%20sleplg&it=x,u,ph,p,mc,m,g,c&coll=pmc&req=2"

```

Caption snippet from the Open-i API



PMC2193129\_JEM991708.f1|PSGL-1 expression in platelets and megakaryocytic cell lines. (A) For flow cytometry, mouse platelets were double labeled with the mAb D9 against mouse  $\alpha$ Ib $\beta$ 3 and with the polyclonal antibody L4025 against mouse PSGL-1, or with preimmune rabbit IgG. (B) For reverse transcriptase PCR, total RNA was prepared from gel-filtered human platelets and from two human megakaryocytic cell lines, CMK-86 and CMK-11/5. After cDNA conversion, a fragment of expected length was amplified by PCR using primers from human PSGL-1 sequence. (C) PSGL-1 expression on CMK-11/5 evaluated by flow cytometry. CMK-11/5 cells were labeled with the mAb PL1 directed against human PSGL-1 (open area) or with preimmune mouse IgG (filled area). FSC-H, forward scatter; SSC-H, side scatter.

### Input to MTI (Identifier | Caption)

PMC2193129\_JEM991708.f1|Humans|C0086418|156604  
PMC2193129\_JEM991708.f1|Mice|C0026809|156604  
PMC2193129\_JEM991708.f1|Rabbits|C3887509|156604  
PMC2193129\_JEM991708.f1|Animals|C0003062|156604  
PMC2193129\_JEM991708.f1|DNA, Complementary|C0006556|81641  
PMC2193129\_JEM991708.f1|Blood Platelets|C0005821|45514  
PMC2193129\_JEM991708.f1|Reverse Transcriptase Polymerase Chain Reaction|C0599161|40552  
PMC2193129\_JEM991708.f1|Flow Cytometry|C0016263|37414  
PMC2193129\_JEM991708.f1|DNA Primers|C0206416|155604  
PMC2193129\_JEM991708.f1|Megakaryocytes|C0025166|87831  
PMC2193129\_JEM991708.f1|Polymerase Chain Reaction|C0032520|45810  
PMC2193129\_JEM991708.f1|Cell Line|C0007600|11652  
PMC2193129\_JEM991708.f1|Immunoglobulin G|C0020852|1000  
PMC2193129\_JEM991708.f1|RNA|C0035668|1000

### Output from MTI (includes Identifier and MeSH heading)

NCBI Resources How To Sign in to NCBI

PMCID: PMC2193129

Journal List > J Exp Med > v.191(8); 2000 Apr 17 > PMC2193129

**JEM** Journal of Experimental Medicine

This article at JEM.org Editors Contact Instructions for Authors

J Exp Med. 2000 Apr 17; 191(8): 1413–1422. PMID: [10770806](#)

### P-Selectin Glycoprotein Ligand 1 (Psgl-1) Is Expressed on Platelets and Can Mediate Platelet–Endothelial Interactions in Vivo

Paul S. Frenette,<sup>a,c</sup> Cécile V. Denis,<sup>a</sup> Linnea Weiss,<sup>c</sup> Kerstin Jurk,<sup>e</sup> Sangeetha Subbarao,<sup>a</sup> Beate Kehrel,<sup>e</sup> John H. Hartwig,<sup>b</sup> Dietmar Vestweber,<sup>d</sup> and Denisa D. Wagner<sup>a</sup>

► Author information ► Article notes ► Copyright and License information [Disclaimer](#)

This article has been [cited by](#) other articles in PMC.

#### Abstract

The platelet plays a pivotal role in maintaining vascular integrity. In a manner similar to leukocytes, platelets interact with selectins expressed on activated endothelium. P-selectin glycoprotein ligand 1 (PSGL-1) is the main P-selectin ligand expressed on leukocytes. Searching for platelet ligand(s), we used a P-selectin–immunoglobulin G (IgG) chimera to affinity purify surface-biotinylated proteins from platelet lysates. P-selectin–bound ligands were eluted with ethylenediaminetetraacetic acid. An ~210-kD biotinylated protein was isolated from both human neutrophil and platelet preparations. A band of the same size was also immunopurified from human platelets using a monoclonal anti-human PSGL-1 antibody and

**Formats:**  
Article | [PubMed](#) | [ePub \(beta\)](#) | [PDF \(424K\)](#) | [Citation](#)

**Share**  
[Facebook](#) [Twitter](#) [Google+](#)

**Save items**

**Similar articles in PubMed**

Primary and secondary capture of platelets onto inflamed femoral artery endothelium is dependent on P-selectin [Eur J Pharmacol. 2008]

PSGL-1 regulates platelet P-selectin-mediated endothelial activation and shedding of P-selectin from [Thromb Haemost. 2007]

P-selectin glycoprotein ligand-1 mediates rolling of human neutrophils on P-selectin. [J Cell Biol. 1995]

The biology of P-selectin glycoprotein ligand-1: its role as a selectin counterreceptor in leukocyte-enc [Thromb Haemost. 1999]

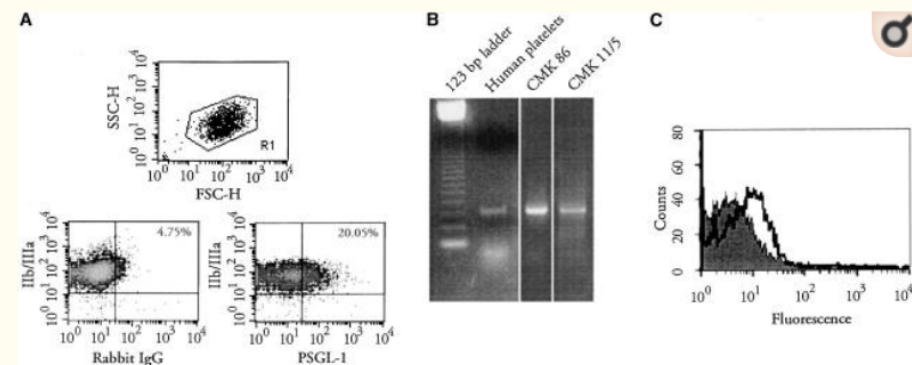
Structure and function of P-selectin glycoprotein ligand-1. [Leuk Lymphoma. 1998]

[See reviews...](#)  
[See all...](#)

**Cited by other articles in PMC**

Peptide Antagonists for P-selectin Discriminate between Sulfate-Dependent Platelet Aggregation a [Journal of Clinical Medicine. ...]

## Full text article in PubMed Central (PMC)



**Figure 1**

PSGL-1 expression in platelets and megakaryocytic cell lines. (A) For flow cytometry, mouse platelets were double labeled with the mAb D9 against mouse  $\alpha$ IIb $\beta$ 3 and with the polyclonal antibody L4025 against mouse PSGL-1, or with preimmune rabbit IgG. (B) For reverse transcriptase PCR, total RNA was prepared from gel-filtered human platelets and from two human megakaryocytic cell lines, CMK-86 and CMK-11/5. After cDNA conversion, a fragment of expected length was amplified by PCR using primers from human PSGL-1 sequence. (C) PSGL-1 expression on CMK-11/5 evaluated by flow cytometry. CMK-11/5 cells were labeled with the mAb PL1 directed against human PSGL-1 (open area) or with preimmune mouse IgG (filled area). FSC-H, forward scatter; SSC-H, side scatter.

### Platelet PSGL-1 Can Bind P-Selectin.

To evaluate whether the PSGL-1 expressed on platelets can bind P-selectin, we affinity isolated selectin ligands using a modified protocol used for isolation of selectin ligands on myeloid cell lines [20](#). Platelets and control neutrophils were isolated, surface biotinylated with Sulfo-*N*-hydroxysulfosuccinimide–LC biotin, and lysed in CHAPS buffer. P-selectin ligands were affinity isolated by incubating cell lysates with protein A–sepharose beads preincubated with P-selectin–IgG. Specifically bound proteins were eluted with

## Figure and caption in full text article