Mining for Medical Relations in Research Articles

Identification of Proteins

By Anna Palmqvist Sjövall and Eric Holmström

Table of Contents

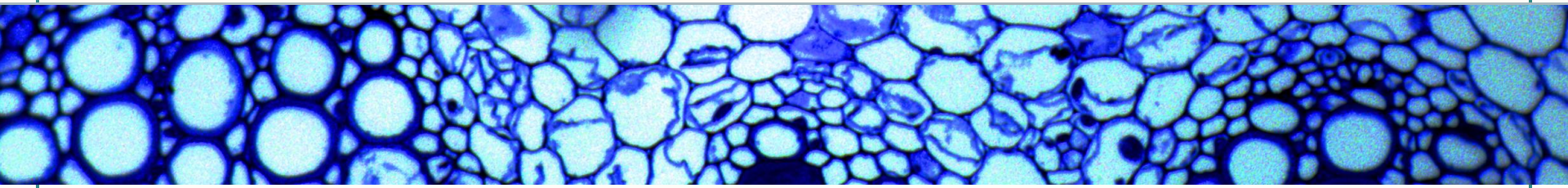
2

01

Introduction
Objective

02

Method



03

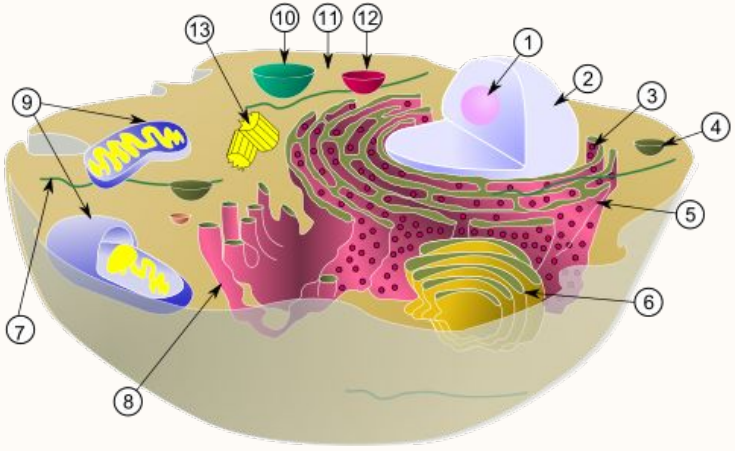
Results

04

Issues
Improvements

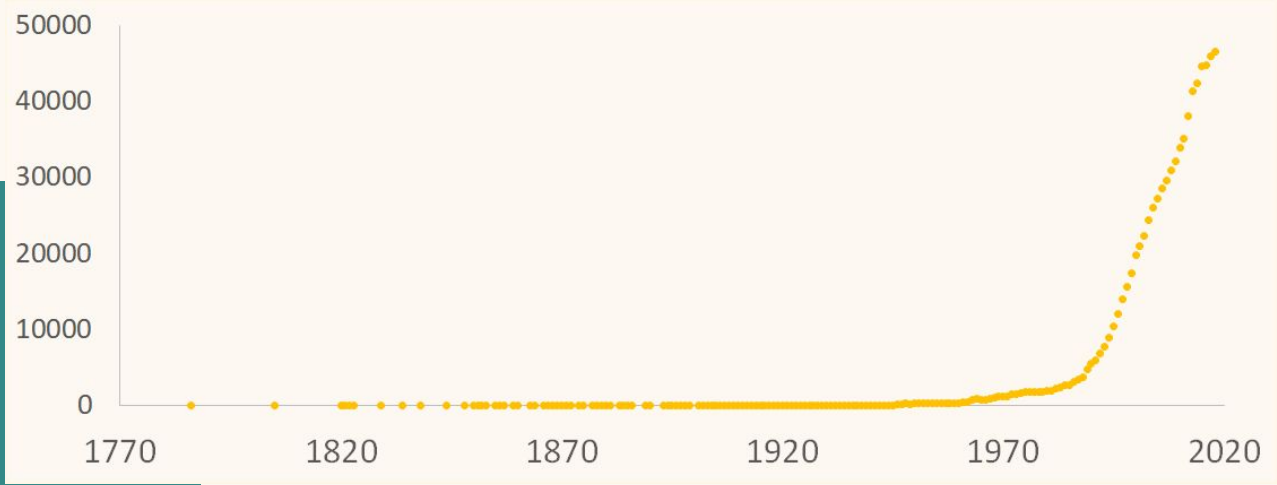
Lysosomal Cell Death

Eukaryotic cells
Apoptosis, Necrosis…
Alzheimer's and Parkinson's disease



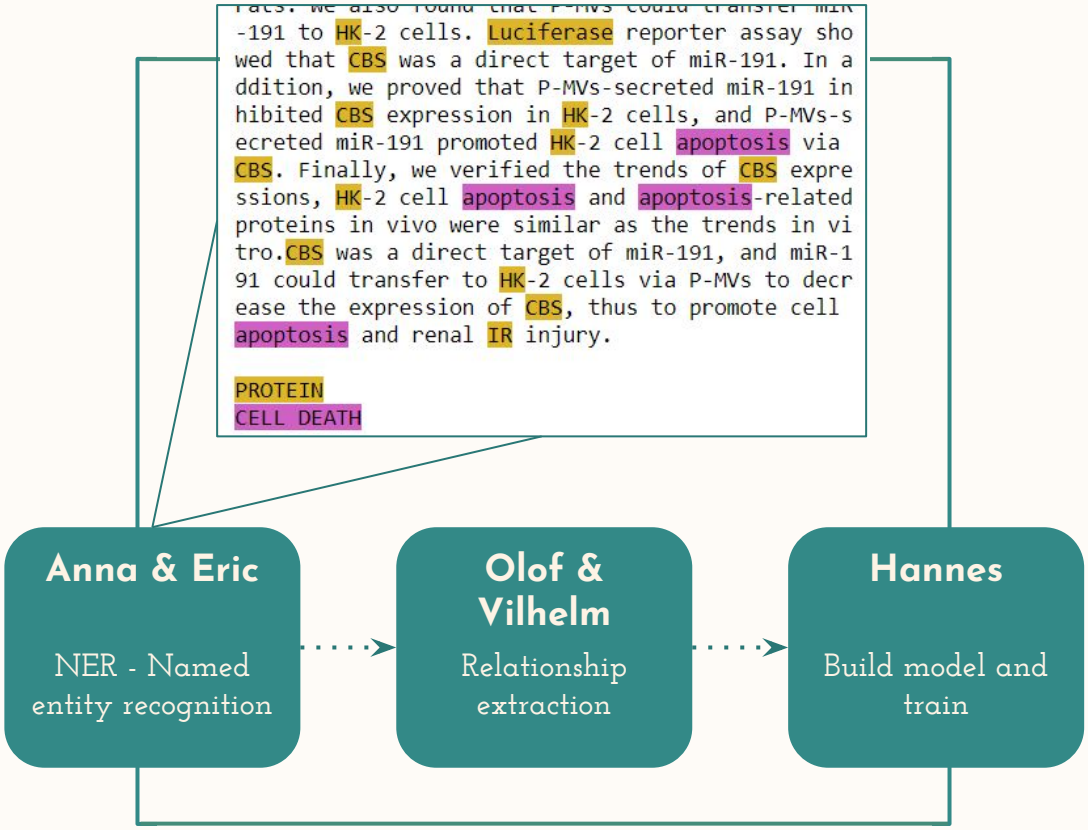
Genes and Proteins

Genes: blueprints for proteins.
Not all genes code for proteins.



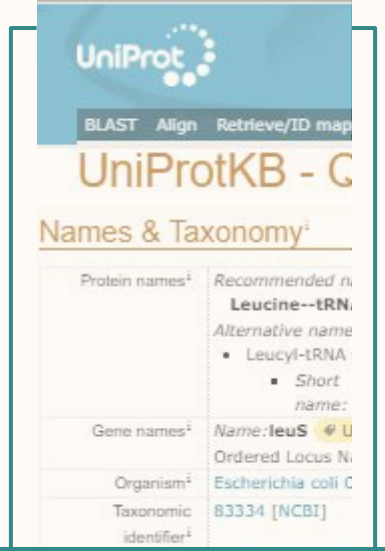
Cell death research articles/year

Over 800 000 articles in total.



Introduction

Investigate the role of protein/genes in relation to lysosomal cell death in order to induce/inhibit diseases..



PubMed

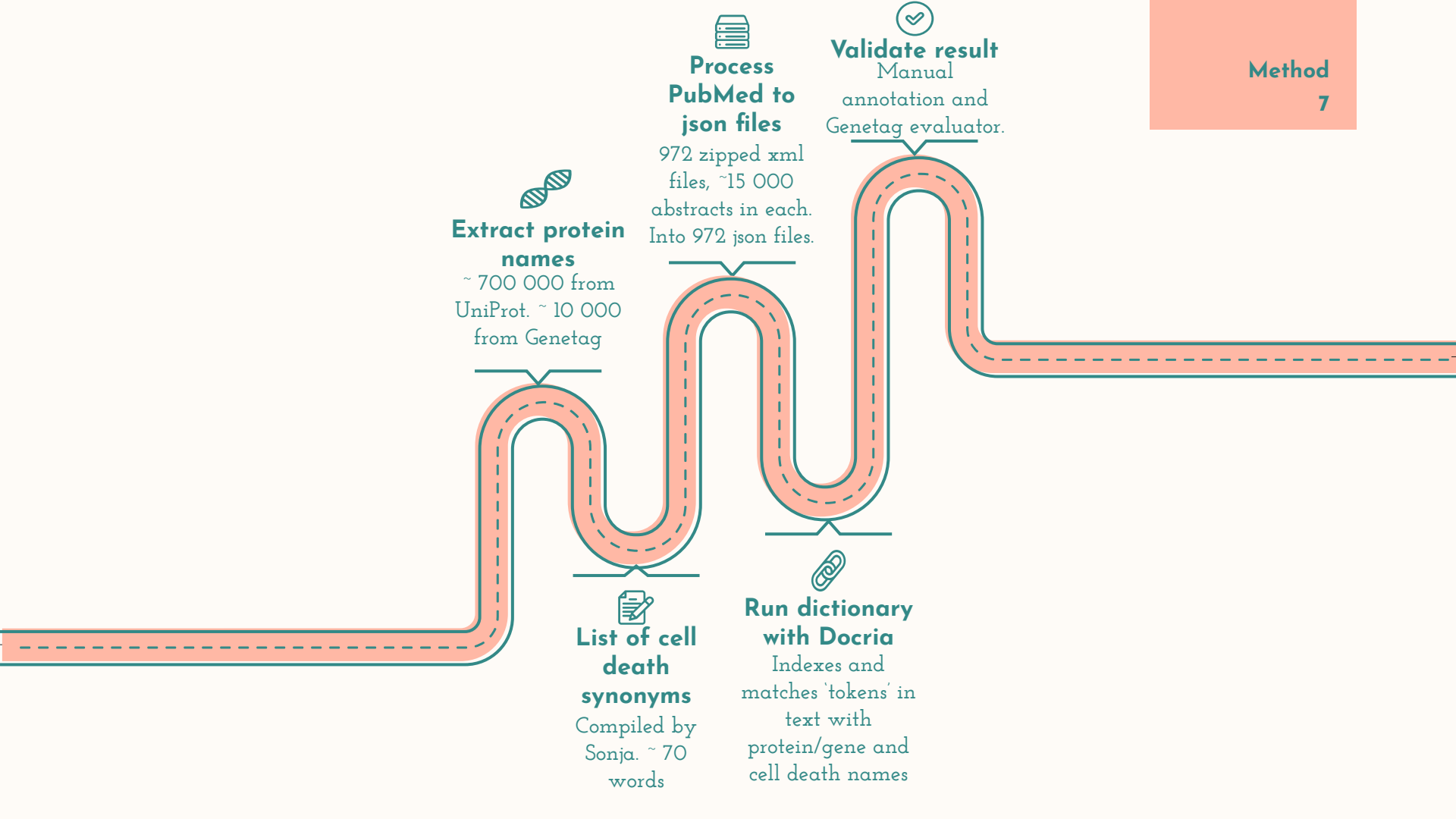
Database for life sciences and biomedical literature. ~20 000 000 abstracts.

UniProt

A manually annotated protein database. ~560 000 names of genes/proteins. One protein = multiple names.

Genetag

Contains 20 000 sentences tagged with gene/protein names. Used as validation set for text mining.


Extract protein names
~ 700 000 from UniProt. ~ 10 000 from Genetag
List of cell death synonyms
Compiled by Sonja. ~ 70 words
Process PubMed to json files
972 zipped xml files, ~15 000 abstracts in each. Into 972 json files.
Run dictionary with Docria
Indexes and matches 'tokens' in text with protein/gene and cell death names
Validate result
Manual annotation and Genetag evaluator.

Podocyte apoptosis is considered as the important element that promotes the development and progress of membranous nephropathy (MN). Unfortunately, the underlying mechanism of podocytes apoptosis in MN remains elusive. We compared the renal expressions of miR-130a-5p and M-type phospholipase A2 receptor (PLA2R) between MN patients (n = 30) and 30 controls by qRT-PCR and western blot, respectively. The podocyte damage model in vitro was established by angiotensin II (Ang II, 100 nmol/L) exposure for 24 h. Interaction between miR-130a-5p and PLA2R was determined using dual-luciferase reporter gene assay. MN mice were induced by intravenous injection of cBSA. In this study, miR-130a-5p expression was significantly decreased both in the renal biopsy specimens from MN patients and podocyte cell line AB8/13 following stimulation of Ang II. Overexpressed miR-130a-5p in AB8/13 cells significantly attenuated the Ang II induced-apoptosis in vitro. In contrast, down-regulated miR-130a-5p induced podocyte apoptosis. PLA2R was identified as the target of miR-130a-5p in AB8/13 cells. And up-regulated or down-regulated PLA2R could obviously attenuate the effect of miR-130a-5p overexpression or knockdown on the apoptosis of AB8/13 cells. Furthermore, it was also observed that overexpressed miR-130a-5p by miR-130a-5p agomir could obviously alleviate renal injury in MN mice. In conclusion, decreased miR-130a-5p was contributed to the pathological mechanism of MN through increasing PLA2R expression, which induced podocyte apoptosis.

id: 30394845

Layers

lysomatches N=6
protmatches N=13
terms N=624

Method: Dictionary
9

Layer: terms

#	term	text	type
#0	'Podocyte'	span(main[0:8]) = 'Podocyte'	'WORD_TITLE_CASE'
#1	'apoptosis'	span(main[9:18]) = 'apoptosis'	'WORD'
#2	'is'	span(main[19:21]) = 'is'	'WORD'
#3	'considered'	span(main[22:32]) = 'considered'	'WORD'
#4	'as'	span(main[33:35]) = 'as'	'WORD'
		. . .	
#621	'podocyte'	span(main[1520:1528]) = 'podocyte'	'WORD'
#622	'apoptosis'	span(main[1529:1538]) = 'apoptosis'	'WORD'
#623	'.'	span(main[1538:1539]) = '.'	'PERIOD'

Method:
Dictionary
10

Layer: protmatches

#	id	terms	text
#0	703295	[Node<terms#311>]	span(main[279:292]) = 'phospholipase'
#1	15984	[Node<terms#311>]	span(main[293:295]) = 'A2'
#2	462220	[Node<terms#311>]	span(main[306:311]) = 'PLA2R'
#3	469720	[Node<terms#311>]	span(main[365:368]) = 'PCR'
#4	713206	[2 nodes from layer: terms]	span(main[455:469]) = 'angiotensin II'
#5	33667	[2 nodes from layer: terms]	span(main[471:477]) = 'Ang II'
#6	462220	[Node<terms#311>]	span(main[546:551]) = 'PLA2R'
#7	705614	[3 nodes from layer: terms]	span(main[578:602]) = 'luciferase reporter gene'
#8	33667	[2 nodes from layer: terms]	span(main[838:844]) = 'Ang II'
#9	33667	[2 nodes from layer: terms]	span(main[917:923]) = 'Ang II'
#10	462220	[Node<terms#311>]	span(main[1020:1025]) = 'PLA2R'
#11	462220	[Node<terms#311>]	span(main[1122:1127]) = 'PLA2R'
#13	462220	[Node<terms#311>]	span(main[1488:1493]) = 'PLA2R'

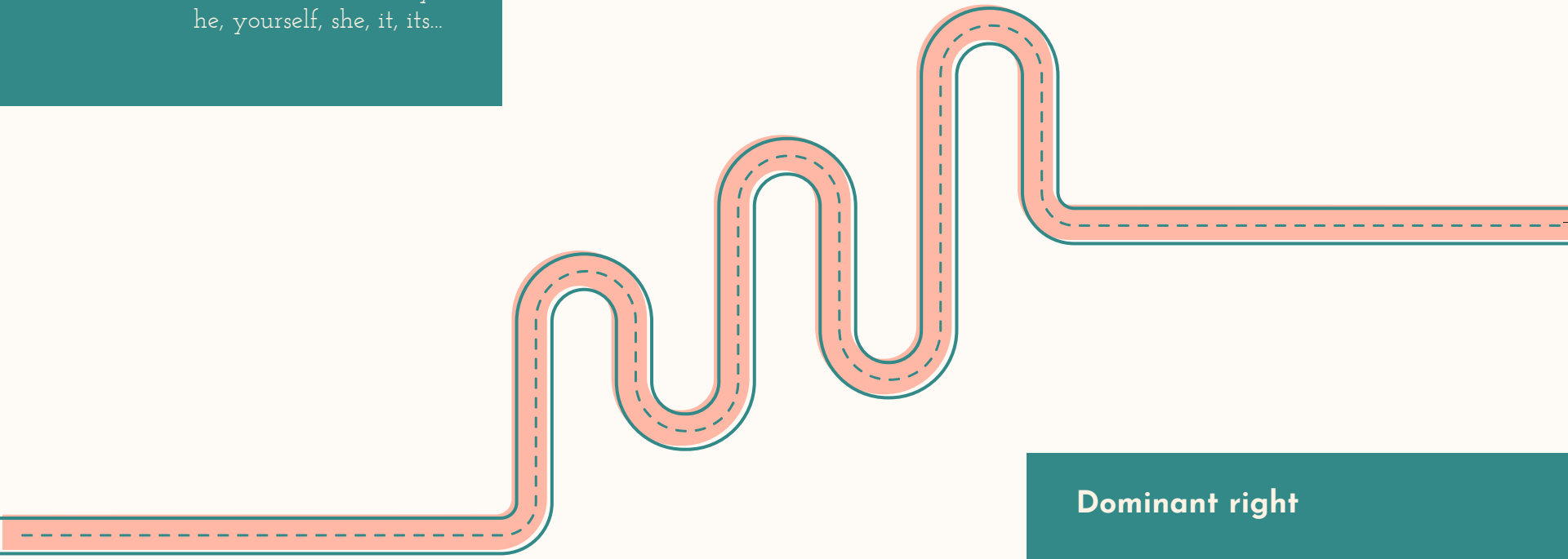
Layer: lysomatches

#	id	terms	text
#0	0	[Node<terms#623>]	span(main[9:18]) = 'apoptosis'
#1	0	[Node<terms#623>]	span(main[186:195]) = 'apoptosis'
#2	0	[Node<terms#623>]	span(main[932:941]) = 'apoptosis'
#3	0	[Node<terms#623>]	span(main[1009:1018]) = 'apoptosis'
#4	0	[Node<terms#623>]	span(main[1215:1224]) = 'apoptosis'
#5	0	[Node<terms#623>]	span(main[1529:1538]) = 'apoptosis'

Stopwords

the, and, in, i, me, my,
he, yourself, she, it, its...

Method:
Dictionary
11



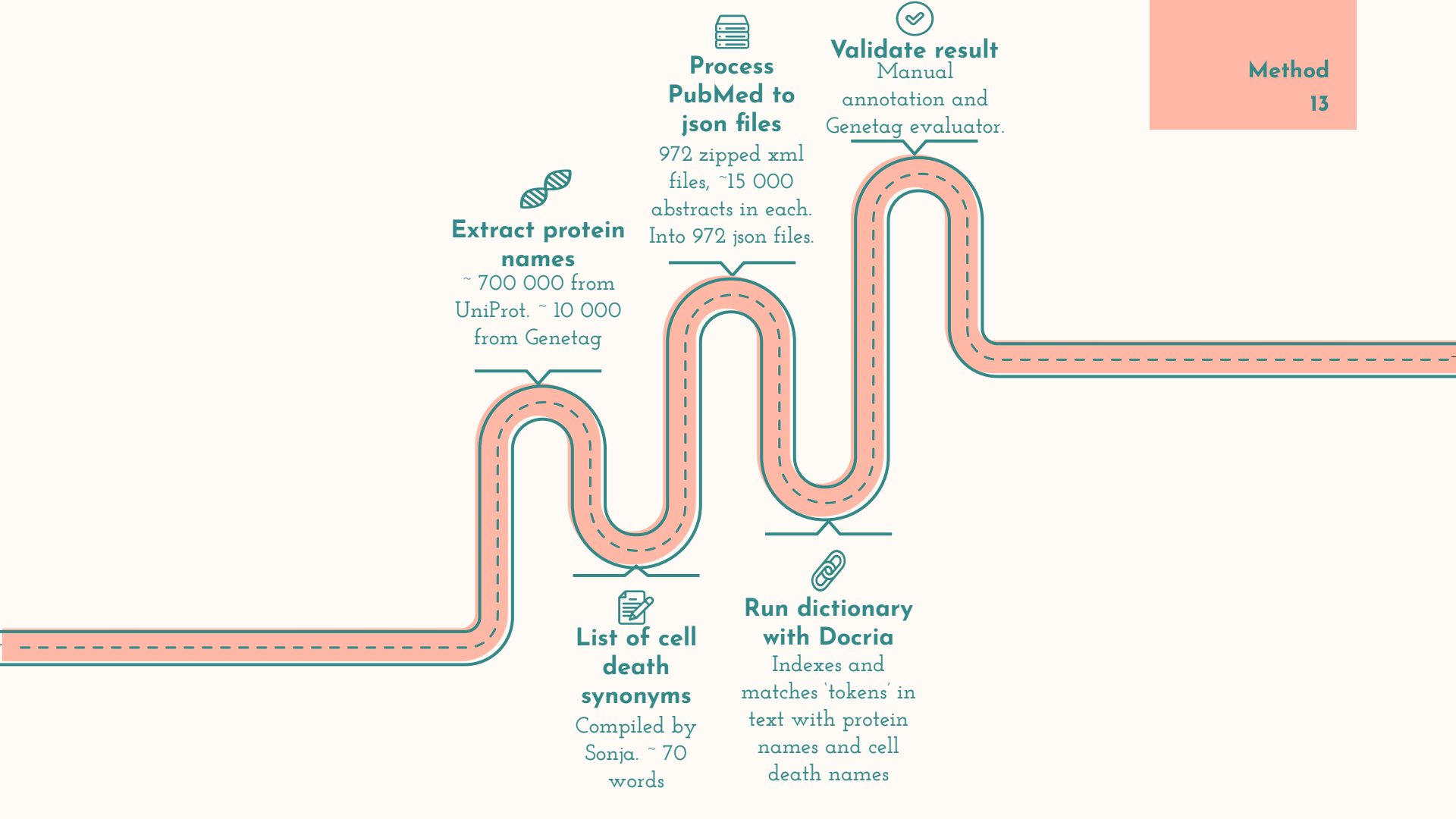
Dominant right

11S globulin seed storage protein G3

11S globulin seed storage protein G3 acidic chain

Podocyte apoptosis is considered as the important element that promotes the development and progress of membranous nephropathy (MN). Unfortunately, the underlying mechanism of podocytes apoptosis in MN remains elusive. We compared the renal expressions of miR-130a-5p and M-type phospholipase A2 receptor (PLA2R) between MN patients (n = 30) and 30 controls by qRT-PCR and western blot, respectively. The podocyte damage model in vitro was established by angiotensin II (Ang II, 100 nmol/L) exposure for 24 h. Interaction between miR-130a-5p and PLA2R was determined using dual-luciferase reporter gene assay. MN mice were induced by intravenous injection of cBSA. In this study, miR-130a-5p expression was significantly decreased both in the renal biopsy specimens from MN patients and podocyte cell line AB8/13 following stimulation of Ang II. Overexpressed miR-130a-5p in AB8/13 cells significantly attenuated the Ang II induced-apoptosis in vitro. In contrast, down-regulated miR-130a-5p induced podocyte apoptosis. PLA2R was identified as the target of miR-130a-5p in AB8/13 cells. And up-regulated or down-regulated PLA2R could obviously attenuate the effect of miR-130a-5p overexpression or knockdown on the apoptosis of AB8/13 cells. Furthermore, it was also observed that overexpressed miR-130a-5p by miR-130a-5p agomir could obviously alleviate renal injury in MN mice. In conclusion, decreased miR-130a-5p was contributed to the pathological mechanism of MN through increasing PLA2R expression, which induced podocyte apoptosis.

PROTEIN
CELL DEATH


Extract protein names
~ 700 000 from UniProt. ~ 10 000 from Genetag
List of cell death synonyms
Compiled by Sonja. ~ 70 words
Process PubMed to json files
972 zipped xml files, ~15 000 abstracts in each. Into 972 json files.
Run dictionary with Docria
Indexes and matches 'tokens' in text with protein names and cell death names
Validate result
Manual annotation and Genetag evaluator.

Statistics

Evaluation with Genetag

		Actual: Positive	Actual: Negative
Predicted	Positive	TP ✓ 2680	FP ✗ 5025
	Negative	FN ✗ 15553	TN ✓ -

Recall

15%
TPR = TP/(TP+FN)

Precision

35%
PPV = TP/(TP + FP)

F1-score

F1 = 2*PPV*TPR/(PPV+TPR)

21%

Podocyte apoptosis is considered as the important element that promotes the development and progress of membranous nephropathy (MN). Unfortunately, the underlying mechanism of podocytes apoptosis in MN remains elusive. We compared the renal expressions of miR-130a-5p and M-type phospholipase A2 receptor (PLA2R) between MN patients (n = 30) and 30 controls by qRT-PCR and western blot, respectively. The podocyte damage model in vitro was established by angiotensin II (Ang II, 100 nmol/L) exposure for 24 h. Interaction between miR-130a-5p and PLA2R was determined using dual-luciferase reporter gene assay. MN mice were induced by intravenous injection of cBSA. In this study, miR-130a-5p expression was significantly decreased both in the renal biopsy specimens from MN patients and podocyte cell line AB8/13 following stimulation of Ang II. Overexpressed miR-130a-5p in AB8/13 cells significantly attenuated the Ang II induced-apoptosis in vitro. In contrast, down-regulated miR-130a-5p induced podocyte apoptosis. PLA2R was identified as the target of miR-130a-5p in AB8/13 cells. And up-regulated or down-regulated PLA2R could obviously attenuate the effect of miR-130a-5p overexpression or knockdown on the apoptosis of AB8/13 cells. Furthermore, it was also observed that overexpressed miR-130a-5p by miR-130a-5p agomir could obviously alleviate renal injury in MN mice. In conclusion, decreased miR-130a-5p was contributed to the pathological mechanism of MN through increasing PLA2R expression, which induced podocyte apoptosis.

Podocyte apoptosis is considered as the important element that promotes the development and progress of membranous nephropathy (MN). Unfortunately, the underlying mechanism of podocytes apoptosis in MN remains elusive. We compared the renal expressions of miR-130a-5p and M-type phospholipase A2 receptor (PLA2R) between MN patients (n = 30) and 30 controls by qRT-PCR and western blot, respectively. The podocyte damage model in vitro was established by angiotensin II (Ang II, 100 nmol/L) exposure for 24 h. Interaction between miR-130a-5p and PLA2R was determined using dual-luciferase reporter gene assay. MN mice were induced by intravenous injection of cBSA. In this study, miR-130a-5p expression was significantly decreased both in the renal biopsy specimens from MN patients and podocyte cell line AB8/13 following stimulation of Ang II. Overexpressed miR-130a-5p in AB8/13 cells significantly attenuated the Ang II induced-apoptosis in vitro. In contrast, down-regulated miR-130a-5p induced podocyte apoptosis. PLA2R was identified as the target of miR-130a-5p in AB8/13 cells. And up-regulated or down-regulated PLA2R could obviously attenuate the effect of miR-130a-5p overexpression or knockdown on the apoptosis of AB8/13 cells. Furthermore, it was also observed that overexpressed miR-130a-5p by miR-130a-5p agomir could obviously alleviate renal injury in MN mice. In conclusion, decreased miR-130a-5p was contributed to the pathological mechanism of MN through increasing PLA2R expression, which induced podocyte apoptosis.

Statistics

Comparison for 10 abstracts

		Actual: Positive	Actual: Negative
Predicted	Positive	TP ✓ 119	FP ✗ 23
	Negative	FN ✗ 13	TN ✓ -

Recall

89.5%
TPR = TP/(TP+FN)

Precision

83.8%
PPV = TP/(TP + FP)

F1-score

F1 = 2*PPV*TPR/(PPV+TPR)

87%

Observations

- TNF-α vs. TNF-alpha
- IL1-13 = IL1, IL2, IL3, … IL13
- dual-luciferase reporter gene assay

Issues

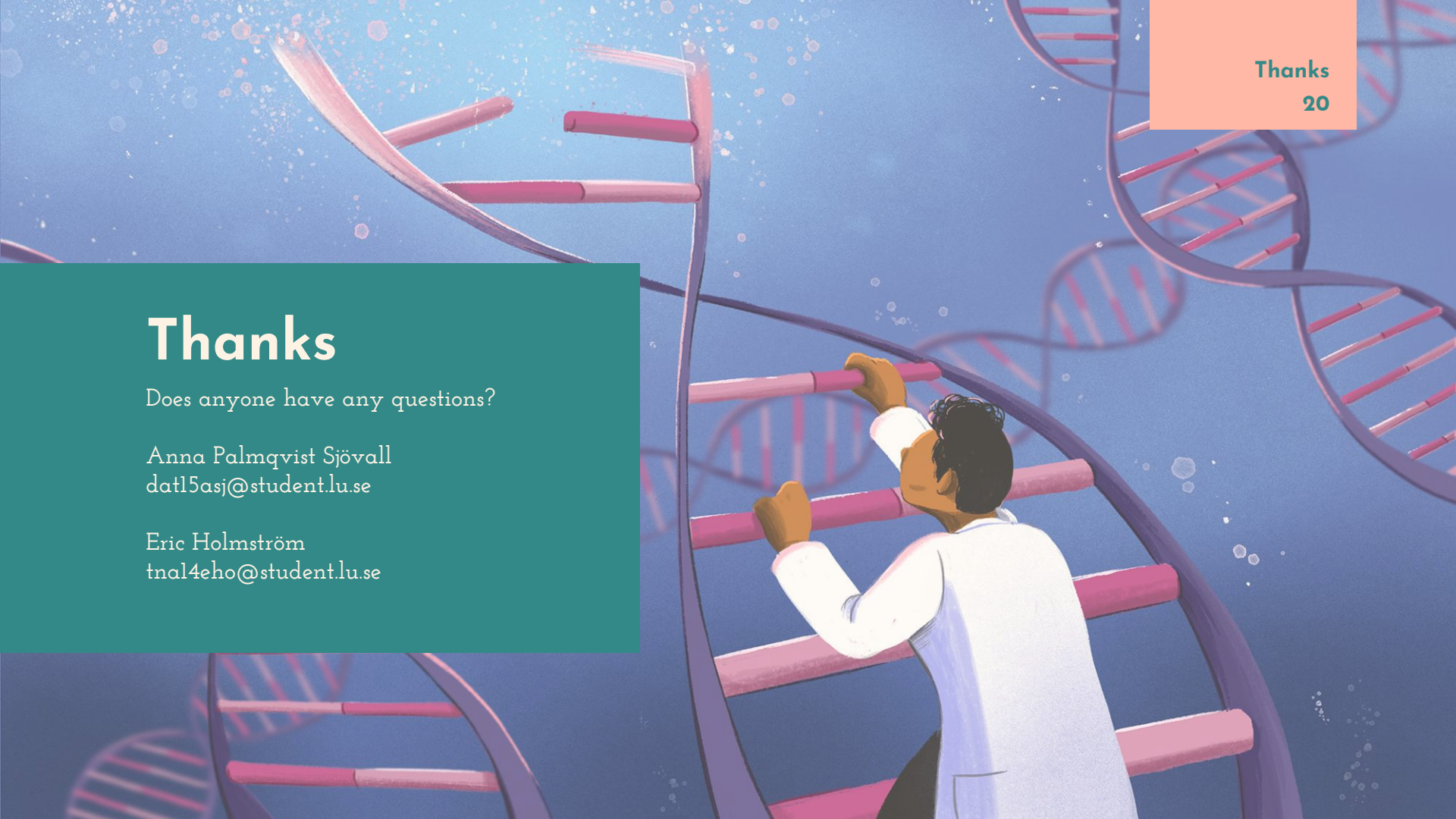
- Unicode characters
- Abbreviation ambiguities
- Ignoring context
- Family names

Future Improvements

- Machine Learning-model
- Species recognition (e.g. HUMAN & RAT)
- GUI
- More extensive evaluation

Benefits

- Includes context → higher accuracy
- Visualize results, easier to use



Thanks

Does anyone have any questions?

Anna Palmqvist Sjövall
dat15asj@student.lu.se

Eric Holmström
tna14eho@student.lu.se