Analysing biomedical image and text data with machine learning and CellProfiler

Sonja Aits

Cell Death, Lysosomes and Artificial Intelligence Group, Lund University & Lund Institute of Advanced Neutron and X-Ray Science (LINXS)

sonja.aits@med.lu.se @Aitslab



Medical knowledge is abundant but scattered and unstructured

Patient journals



Scientific literature and other texts



Bioinformatics databases



Large unstructured research datasets





Al can help in processing of large and unstructured biomedical data



Natural language processing (NLP)

The number of medical research articles is enormous



Natural language processing tools can identify and connect pieces of information in medical texts



www.aitslab.org

Named entity recognition (NER) "This is a protein"

Disease Drug/treatment Gene/protein Process/location Relation

PLoS One. 2012;7(10):e45381. doi: 10.1371/journal.pone.0045381. Epub 2012 Oct 11.

Identification of cytoskeleton-associated proteins essential for lysosomal stability and survival of human cancer cells.

Groth-Pedersen L¹, Aits S, Corcelle-Termeau E, Petersen NH, Nylandsted J, Jäättelä M.

Author information

Abstract

Microtubule-disturbing drugs inhibit vsosomal trafficking and induce vsosomal membrane permeabilization followed by cathepsin-dependent cell death. To identify specific trafficking-related proteins that control cell survival and vsosomal stability, we screened a molecular motor siRNA library in human MCF7 breast cancer cells. SiRNAs targeting four kinesins (KIF11/Eg5, KIF20A, KIF21A, KIF25), myosin 1G (MYO1G), myosin heavy chain 1 (MYH1) and tropomyosin 2 (TPM2) were identified as effective inducers of non-apoptotic cell death. The cell death induced by KIF11, KIF21A, KIF25, MYH1 or TPM2 siRNAs was preceded by vsosomal membrane permeabilization, and all identified siRNAs induced several changes in the endo-lysosomal compartment, i.e. increased vsosomal volume (KIF11, KIF20A, KIF25, MYH1), increased cysteine cathepsin activity (KIF20A, KIF25), altered vsosomal localization (KIF25, MYH1, TPM2), increased dextran accumulation (KIF20A), or reduced autophagic flux (MYO1G, MYH1). Importantly, all seven siRNAs also killed human cervix cancer (HeLa) and osteosarcoma (U-2-OS) cells and sensitized cancer cells to other vsosome-destabilizing treatments, i.e. photo-oxidation, siramesine, etoposide or cisplatin.

www.aitslab.org

Relationship extraction "(SARS-CoV-2)-binds to-(ACE2)"

Relation: binds to SARS-CoV2 binding to ACE2 can lead to excessive angiotensin II signaling, which activates the STING pathway in mice.

CovidQ: extracting information from electronic health records with NLP

SYMPTOM ICD-10: R43.0 Anosmia

SYMPTOM ICD-10: R50 Fever

75y old male with reduced sense of smell and mild fever. No cough or other respiratory symptoms. Chronic hypertension.

DISEASE ICD-10: I10 Essential (primary) hypertension

Medical NLP has many applications

COVID19 knowledge sources

Scientific literature Electronic health records Social media posts News reports Bioinformatics databases



Knowledge extraction and connection

Symptoms Risk factors Co-morbidity Virus-host interaction Drug candidates Animal/cell models

Actionable COVID19 insights for

Therapy development Research tool identification Clinical decision-support Pandemic surveillance Public health measures etc, etc



Natural language processing uses deep neural networks trained on GPUs



Computer vision

Computer vision enables large-scale microscopy



Computer vision models find objects and draw outlines (= instance segmentation)

Identification of nuclear outlines

Annotated outlines U-net prediction







CellProfiler measures hundreds of features to generate phenotypic fingerprints



Computer vision models for image classification







Images from collaborator Darcy Wagner

Why and how we use LUNARC

- Temporary storage for large datasets
- Connection to Swestore storage
- Parallel processing on CPUs \rightarrow faster results
- GPU nodes for machine learning
- Helpful support

WinSCP lftp Python Pytorch Kersas/Tensorflow Anaconda virtual environments Jupyter notebooks MATLAB CellProfiler **MySQL**

Other needs

- Expansion of GPU nodes
- GPUs for sensitive patient data
- Large memory
- Permanent storage of large datasets that allows public access (FAIR)
- Support for web-based citizen science projects (e.g. cloud sql, python flask)





Cell Death, Lysosomes and AI Group



Karl Åström, Lund University Darcy Wagner, Lund University Alexandros Sopasakis, Lund University Johan Nilsson, Lund University Dennis Medved, Lund University Johan Rung, Uppsala University Ricky Johnstone, Peter MacCallum Cancer Centre Kaylene Simpson, Peter MacCallum Cancer Centre

Pierre Nugues, Lund University

Marcus Klang, Lund University

AIR Lund team, Lund University

sonja.aits@med.lu.se Johan Frid, Humanities Lab, Lund University Johanna Berg, Skåne University Hospital Martin Gerdin Wärnberg, Karolinska Institute

Salma Kazemi Rashed Rafsan Ahmed Petter Berntsson Alexander Skafte Mariam Miari Malou Arvidsson Carl Ollvik

@Aitslab



AI Lund

RESEARCH, EDUCATION AND INNOVATION REA OF ARTIFICIAL INTELLIGENCE







