



KfK-Praktikum, KfK-Seminar

Studienrichtung: Masterstudium Medizinische Informatik

Thema: Novel Targets in Pain Research

Beschreibung: Für einen aktuellen Übersichtsartikel wurde cursorisch erhoben, was es an neuen Ansatzpunkten für schmerzlindernde Substanzen (Analgetika) gibt. Hierfür wurde über die Pubmed E-Utilities automatisiert gesucht (Web scrape), als Basis dient eine Liste aller biologischen Moleküle ('targets and family' der IuPhar/BPS). Als Suchbegriffe wurde je ein Molekül mit 'pain' kombiniert. Um in diesen Daten Trends zu identifizieren, wurden die letzten 3 Jahre mit den 3 Jahren zuvor verglichen. Als Ergebnisse wurden 1) Large Targets, 2) Rising Topics und 3) New Topics identifiziert. Diese Ergebnisse sind für das wissenschaftliche Feld von Interesse. Nebenstehend finden sich exemplarisch die Large Targets, zudem mit Klassifikation der Molekülklasse.

A	Total	Last 3y	3y earlier
TRPV1	2695	612	564
COX-2	2469	418	339
Mu opioid receptor	2022	391	371
Cysteine aminotransferase	1852	279	252
Mitogen-activated protein kinase	1418	240	345
CD4	1286	238	212
Plasminogen	1183	87	88
Peptide transporter 4	1112	149	230
Angiotensin-converting enzyme	1046	94	99
TRPA1	999	299	260
Hydroxymethylglutaryl-CoA reductase	934	123	225
Kappa opioid receptor	724	105	109
COX-1	678	77	75
Nav1.7	669	218	184
Delta opioid receptor	634	95	86
Myeloperoxidase	597	114	146
P2X3	488	88	92
Caspase 3	479	158	130
Tyrosine kinase non receptor 1	461	72	92
Nav1.8	452	100	107
Tyrosine kinase non receptor 2	415	70	84
TLR4	414	167	119
Fatty acid amide hydrolase	383	76	75
TRPM8	382	91	96
Renin	380	46	48
CYP2D6	366	78	63
CB1 receptor	347	75	67
catalase	334	101	77
Protein kinase A	320	51	57
Catechol-O-methyltransferase	307	60	68

Enzyme	VGIC	GPCR	TC	LC

Aufgaben:

- 1) Kontinuierliches Update dieser Daten mit Webpublikation
- 2) Die Liste der Ergebnisse soll als CSV herunterladbar sein
- 3) Eine Weiterentwicklung des Suchparadigmas in den NCBI E-Utilities*
- 4) Automatische Klassifikation der Targets mit visueller Aufbereitung
- 5) Mitarbeit an der ordentlichen Beschreibung für eine Publikation
- 6) Publikation des Codes zur Nutzbarkeit durch andere wissenschaftliche Felder

*Die einmalig durchgeführte Analyse der jeweils 30 Topergebnisse (siehe Tabelle der Large Targets) hat verschiedene Herausforderungen aufgezeigt. Es sind verschiedene Identifier für die biologischen Moleküle möglich, z.B. human gene nomenclature name, rat genome database name oder die zugehörigen Kürzel. Alle diese Ansätze führen für sich genommen zu falsch-positiven oder falsch-negativen Ergebnissen, dies gilt es zu optimieren.