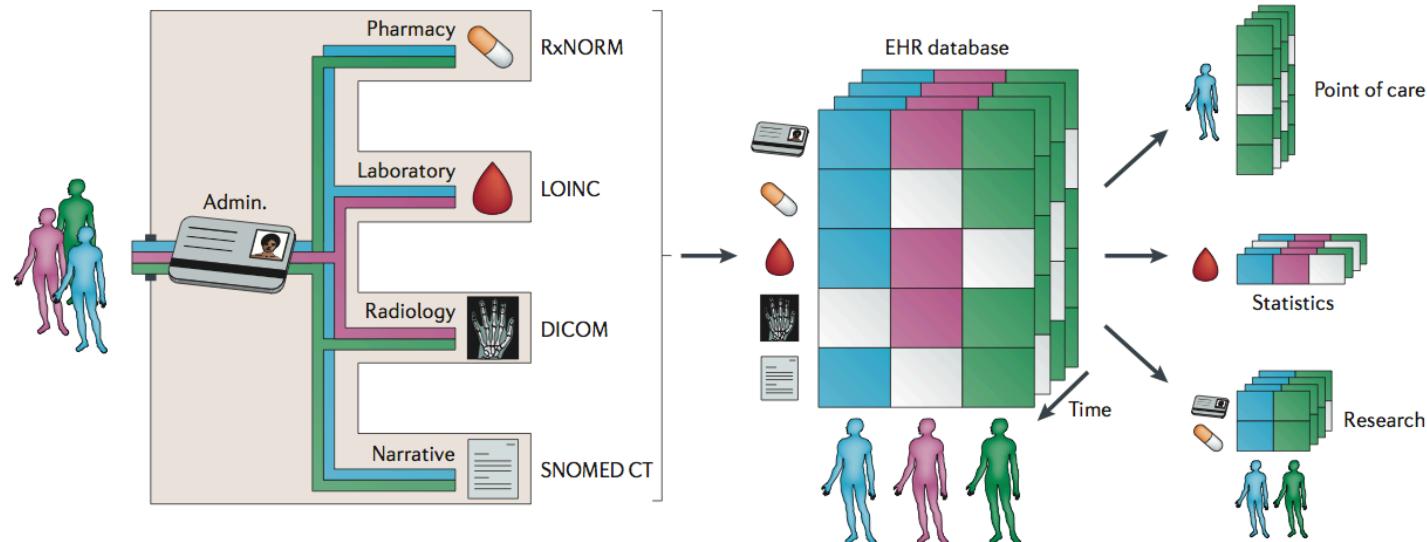


Medical data and text mining

Linking diseases, drugs, and adverse reactions



Lars Juhl Jensen
jensenlab.org



structured data

	Assigned diagnosis				Medications			Laboratory values			Demographics	
	C1	C2	C3	C4	M1	M2	M3	L1	L2	L3	D1	D2
Patient 1	■		■					■			■	
Patient 2												
Patient 3			■									
Patient 4	■						■		■			
Patient 5			■									
Patient 6		■										
Patient 7			■								■	■
Patient 8			■				■					
Patient 9	■					■	■	■	■		■	
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

unstructured data

symptomer på paranoid skizofreni. i denne beskriver hun at "hendes største problem nok er den manglende sociale evne, som er en følge af sygdommen (paranoid skizofreni) og henviser til contras beskrivelse" Pt. Nævner sin mor, som han mener har en nervøs lidelse, muligvis social fobi pt. har her til aften angivet tiltagende bivirkninger i form af trækninger i nakken, indre uro og stivhed af fingre. pt. har fået svar på sit ekg, som viser sinus rytme med enkelte ventrikulære ekstrasystoler uforandret fra tidl. med baggrund i oplysninger om tidligere maniske episoder præget af irritabilitet, hyperaktivitet og øget seksuel interesse revurderes diagnosen til bipolar affektiv sindslidelse. følges i distrikt vest med psykologsamtaler. har i dag tydeligvis brug for en faglig forklaring på hendes symptomer. det drejer sig om paranoia , uvirkelighedsfølelser , influenssymp. og koncentrationsbesvær. det største problem er dog samværet med andre. det er specielt om natten det påvirker hendes astma., klg. desuden over uro i benene. ,xxx nævner på et tidspunkt, hun er bange for, tidlige tiders spiseforstyrrelser er ved at dukke op igen. xxx har haft søvnbesvær og har af vagtlægen i aftes fået tabl. imovane 7,5 mg med god effekt. kl 19, pinex, tabletter, 500 mg, indtaget dosis: 1 gram, for hovedpine. pt. er henvist til at drøfte sagen nærmere med soc.rdg. xxx og lade sig rådgive her før hun evt. kontakter den kommunale sagsbeh. pt. bliver i den del af samtalen irritabel og devaluerende.

central registries

civil registration system

	Assigned diagnosis				Medications			Laboratory values			Demographics	
	C1	C2	C3	C4	M1	M2	M3	L1	L2	L3	D1	D2
Patient 1	■		■		...			■			■	...
Patient 2				
Patient 3			■	
Patient 4	■				...		■		
Patient 5			■		...	■				
Patient 6	■			
Patient 7			■		...	■				
Patient 8			■		...		■			
Patient 9	■				...			■	■		■	...

national discharge registry

	Assigned diagnosis				Medications			Laboratory values			Demographics	
	C1	C2	C3	C4	M1	M2	M3	L1	L2	L3	D1	D2
Patient 1	■		■					■			■	
Patient 2												
Patient 3			■									
Patient 4	■						■		■			
Patient 5			■									
Patient 6		■										
Patient 7			■								■	
Patient 8			■									
Patient 9	■							■	■			
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

19 years

6.6 million patients

163 million diagnoses

only hospitals

reimbursement

linking diseases

comorbidity

contingency table

	C4	$\neg C4$
C2	10	40
$\neg C2$	90	860
	100	900
		1,000

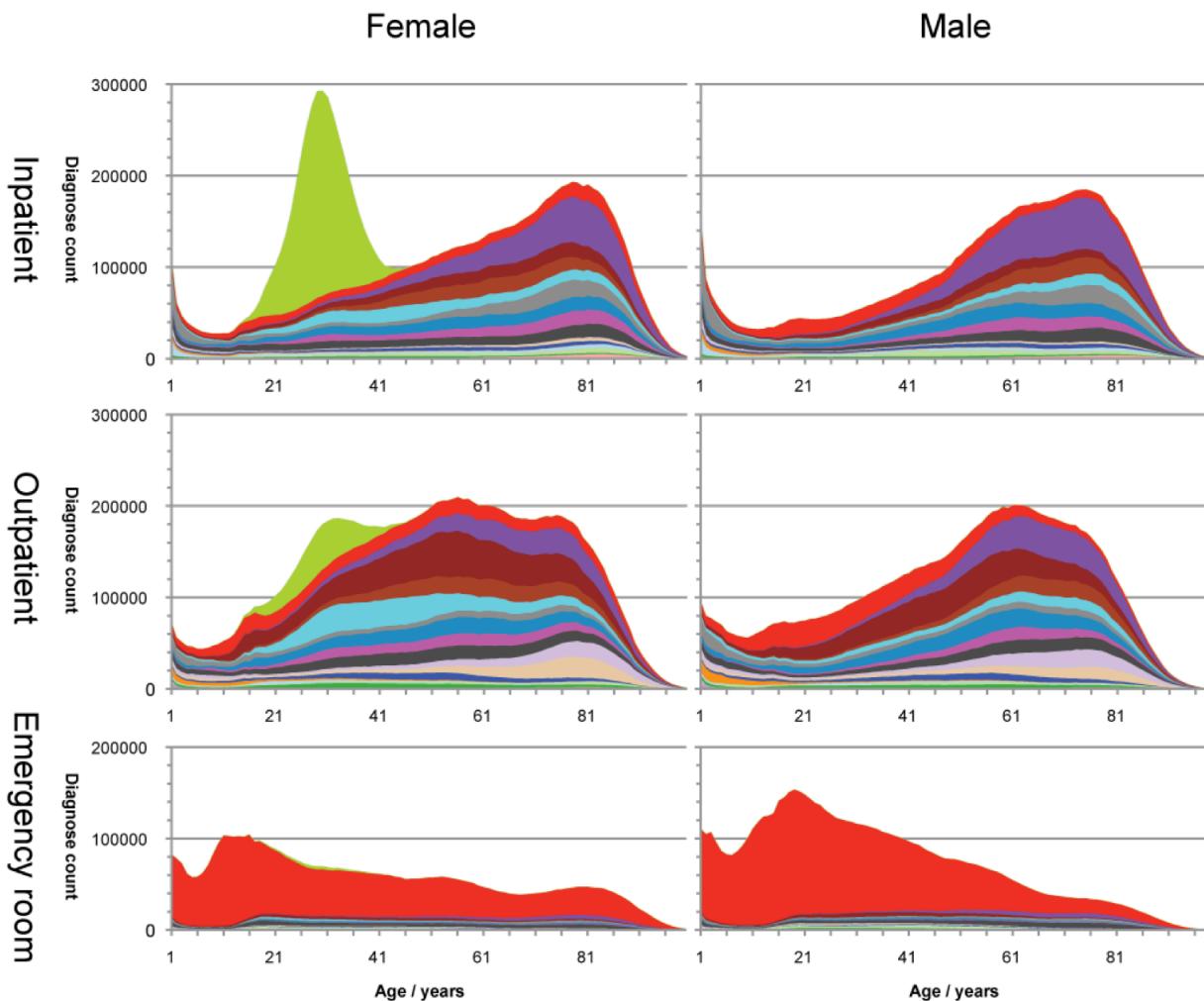
confounding factors

“known knowns”

sex

age

type of hospital encounter



“known unknowns”

smoking

diet

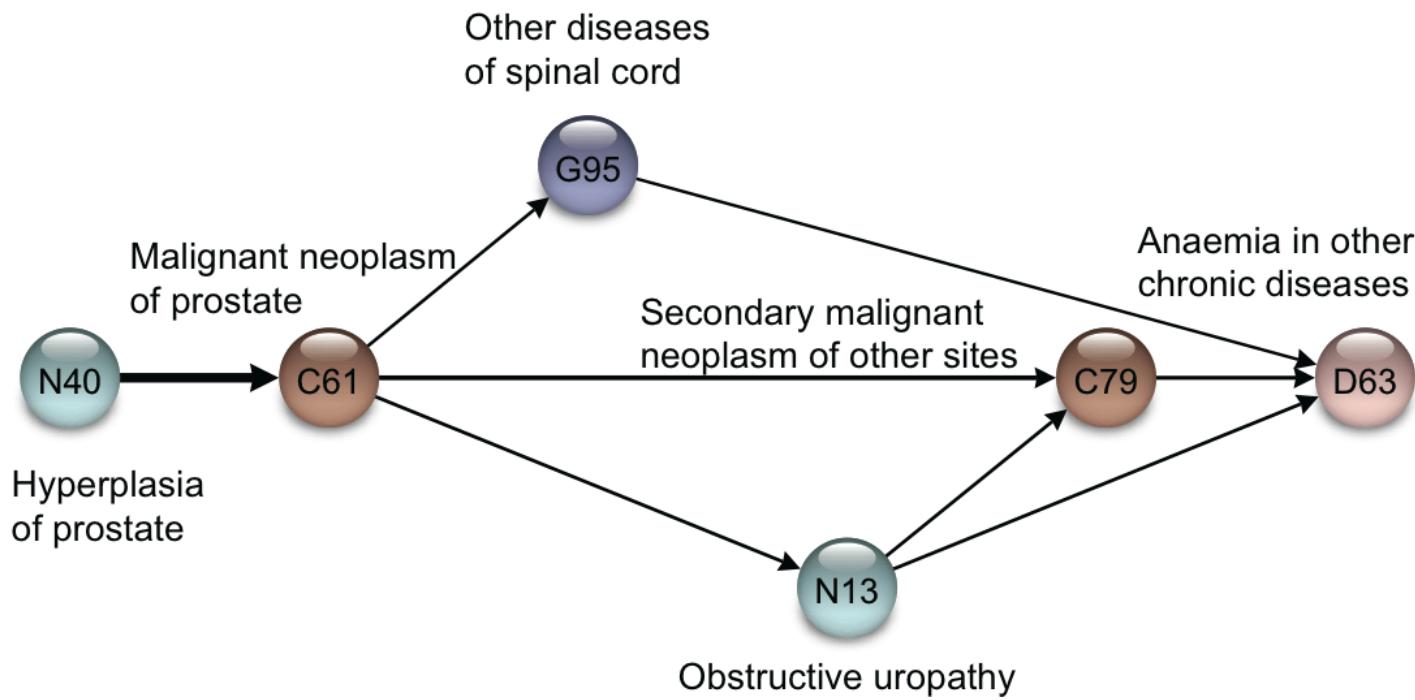
“unknown unknowns”

reporting biases

matched controls

temporal correlations

trajectories



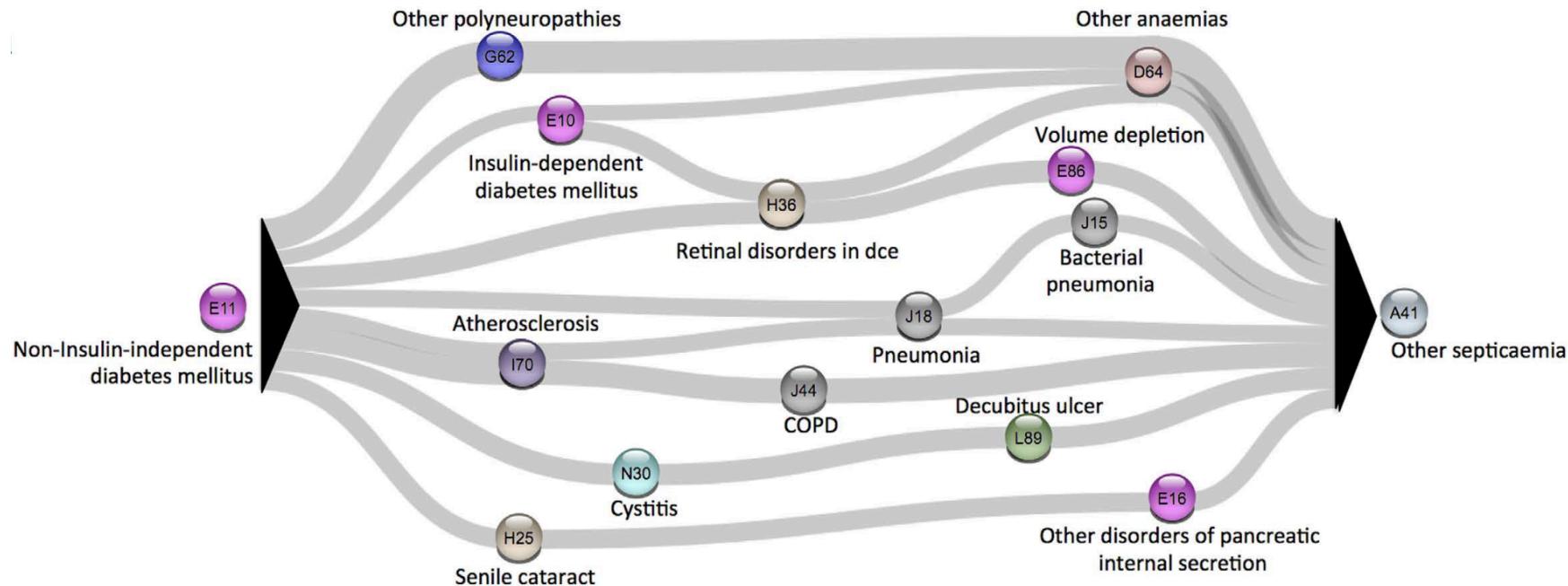
global map of diseases

targeted trajectories

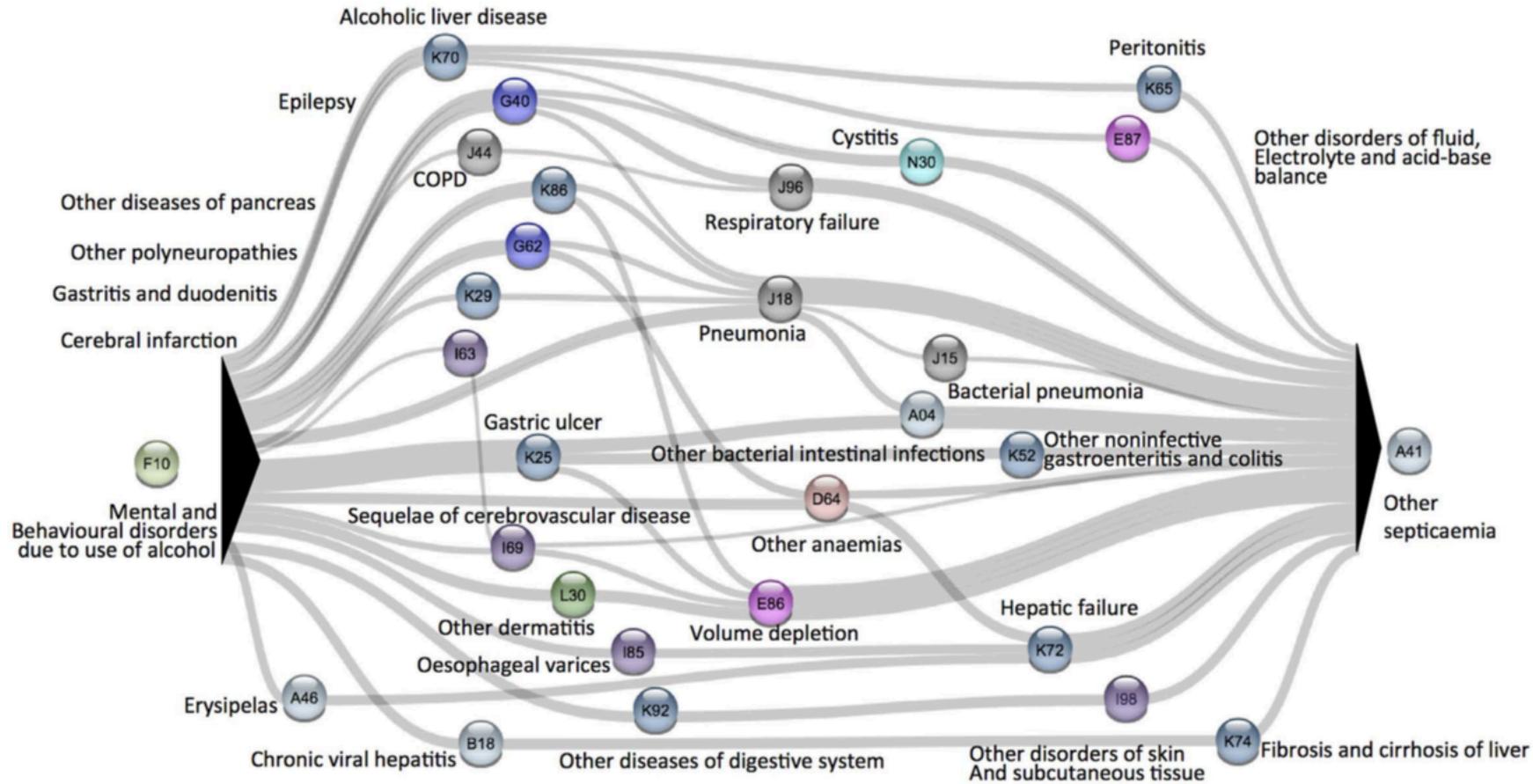
sepsis

stratify patients

diabetes-related sepsis



alcohol-related sepsis

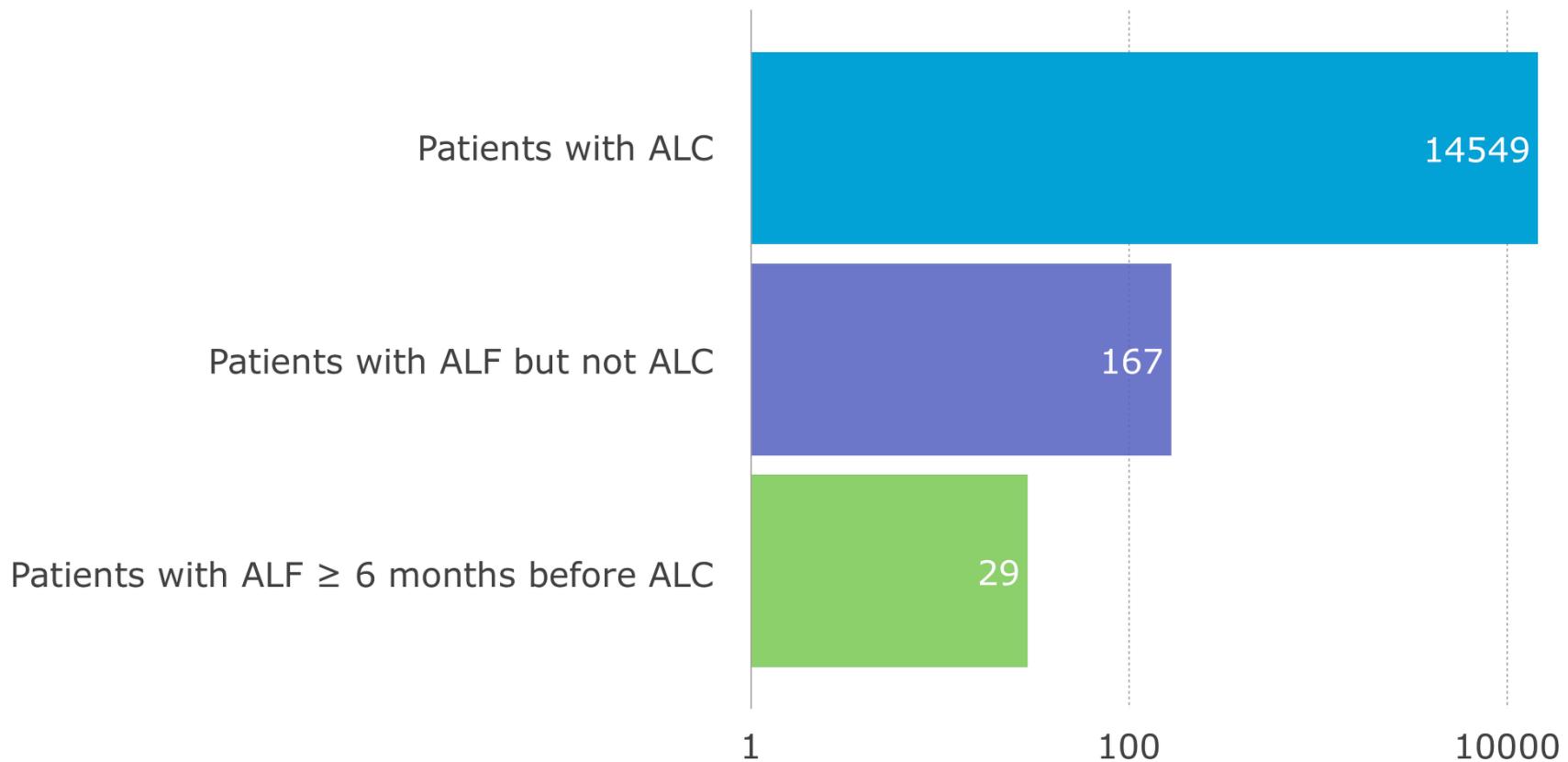


predict mortality

alcoholic liver disease

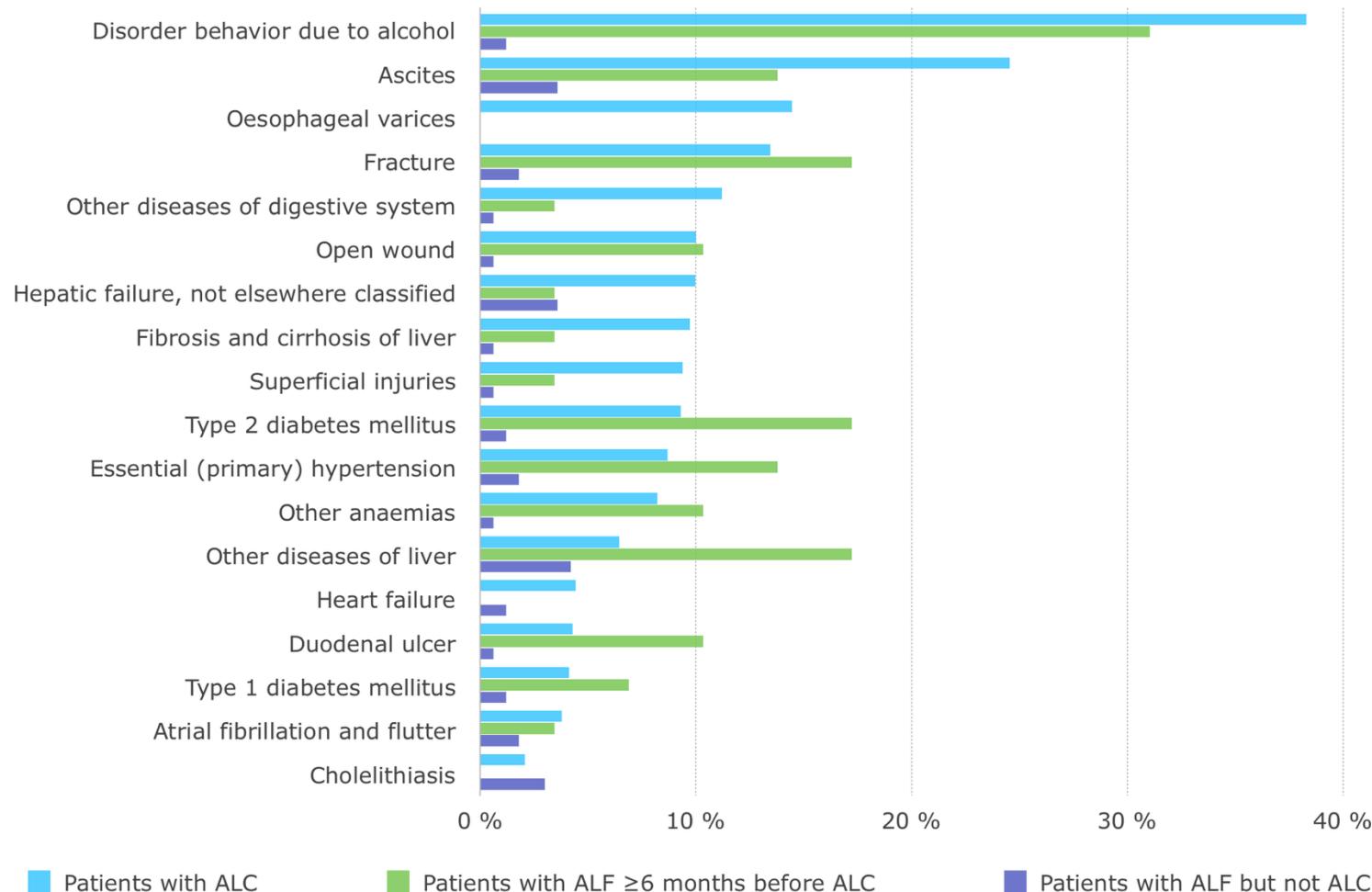
fibrosis → cirrhosis

Number of Patients after Filtering step



upstream diagnoses

Comparison of Most Representative ICD 10 codes



combine with omics

identify biomarkers

linking drugs and ADRs

clinical trials

spontaneous reports

**Report of Suspected Adverse Drug Reaction
including Birth Defects** **224289**

(Note: Identities of Reporter, Patient and Institution will remain Confidential)

Patient (Initials or Record # only)	Age	Sex	Weight	Height	
[REDACTED]	05 DEC 2005	55	M	80	168

Adverse Reaction Description: **DESC** Date of Onset of Reaction: **29/11/06**

Patient with a **NON ST ELEVATION MI**
HAD DIAGNOSTIC ANGIOGRAM SHOWING
SEVERE STENOSIS IN LAD.
THE SAME DAY HAD PCI TO LAD DURING
WHICH EXPERIENCED PROFOUND AND SUSTAINED
NEO HYPOTENSION NOT RELIEVED WITH AMARININE
6mg (several 0.5mg doses) AND IABP. IMPROVED
AFTER HYDROCOCLONE 200mg + PHENERGAN
GIVEN. ??ALERGIC REACTION TO CONTRAST
(ISOVUE 370)

All Drug Therapy Prior to Reaction Asterisk Suspected Drug(s) (please use trade names)	Daily Dosage and Route	Date Begun	Date Stopped	Reason for Use
ASPIRIN	300mg PO	29/11/06	—	NSTEMI
CLOPIROXEL	300mg PO	29/11/06	—	NSTEMI
TEMAZEPAM	10mg PO	29/11/06	—	Sedation
TIVASILAN	1V, 500US	29/11/06	29/11/06	NSTEMI
midazolam	2mg IV	29/11/06	29/11/06	SCYATIN
OMANIPRAQUE	90mg IC	29/11/06	29/11/06	? ANESTHESIA
ISOVUE 370	—	29/11/06	29/11/06	ANGIOGRAM

Treatment (of reaction): **AMARININE, HYDROCOCLONE, PHENERGAN**

Outcome: Recovered Not Yet Recovered Unknown Fatal Date of Death **11/11**

Sequelae: No Yes (describe) **MYOCARDIAL INFARCTION**

Comments (eg. relevant history, allergies, previous exposure to this drug):

NO KNOWN ALLERGIES BEFORE THE EPISODE.
HAD ANGIOGRAM IN ANOTHER HOSPITAL.
THEN PCI SAME DAY. REACTION DURING
PCI

Reporting Doctor, Pharmacist, etc:

Name: **[REDACTED]**

Address: **[REDACTED]**

POSS

Signature _____

30/11/06

underreporting

how hard can it be?

drugs

	Assigned diagnosis				Medications			Laboratory values			Demographics	
	C1	C2	C3	C4	M1	M2	M3	L1	L2	L3	D1	D2
Patient 1	■		■					■			■	
Patient 2												
Patient 3			■									
Patient 4	■						■		■			
Patient 5			■									
Patient 6	■											
Patient 7			■									
Patient 8		■					■					
Patient 9	■					■	■	■	■			
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

adverse events

symptomer på paranoid skizofreni. i denne beskriver hun at "hendes største problem nok er den manglende sociale evne, som er en følge af sygdommen (paranoid skizofreni) og henviser til contras beskrivelse" Pt. Nævner sin mor, som han mener har en nervøs lidelse, muligvis social fobi pt. har her til aften angivet tiltagende bivirkninger i form af trækninger i nakken, indre uro og stivhed af fingre. pt. har fået svar på sit ekg, som viser sinus rytme med enkelte ventrikulære ekstrasystoler uforandret fra tidl. med baggrund i oplysninger om tidligere maniske episoder præget af irritabilitet, hyperaktivitet og øget seksuel interesse revurderes diagnosen til bipolar affektiv sindslidelse. følges i distrikt vest med psykologsamtaler. har i dag tydeligvis brug for en faglig forklaring på hendes symptomer. det drejer sig om paranoia , uvirkelighedsfølelser , influenssymp. og koncentrationsbesvær. det største problem er dog samværet med andre. det er specielt om natten det påvirker hendes astma., klg. desuden over uro i benene. ,xxx nævner på et tidspunkt, hun er bange for, tidlige tiders spiseforstyrrelser er ved at dukke op igen. xxx har haft søvnbesvær og har af vagtlægen i aftes fået tabl. imovane 7,5 mg med god effekt. kl 19, pinex, tabletter, 500 mg, indtaget dosis: 1 gram, for hovedpine. pt. er henvist til at drøfte sagen nærmere med soc.rdg. xxx og lade sig rådgive her før hun evt. kontakter den kommunale sagsbeh. pt. bliver i den del af samtalen irritabel og devaluerende.

free text

Danish

busy doctors

typos

too much to read

computer

as smart as a dog

teach it specific tricks

What we say to dogs

Okay, Ginger! I've had it!
You stay out of the garbage!
Understand, Ginger? Stay out
of the garbage, or else!



what they hear



named entity recognition

comprehensive dictionary

drugs

adverse events

synonyms

orthographic variation

spelling mistakes

Clozapine

klozapine

klosapin

klozapin

Clozapine

clozapin

chlozapin

klossapin

chlosapin

chlozapine

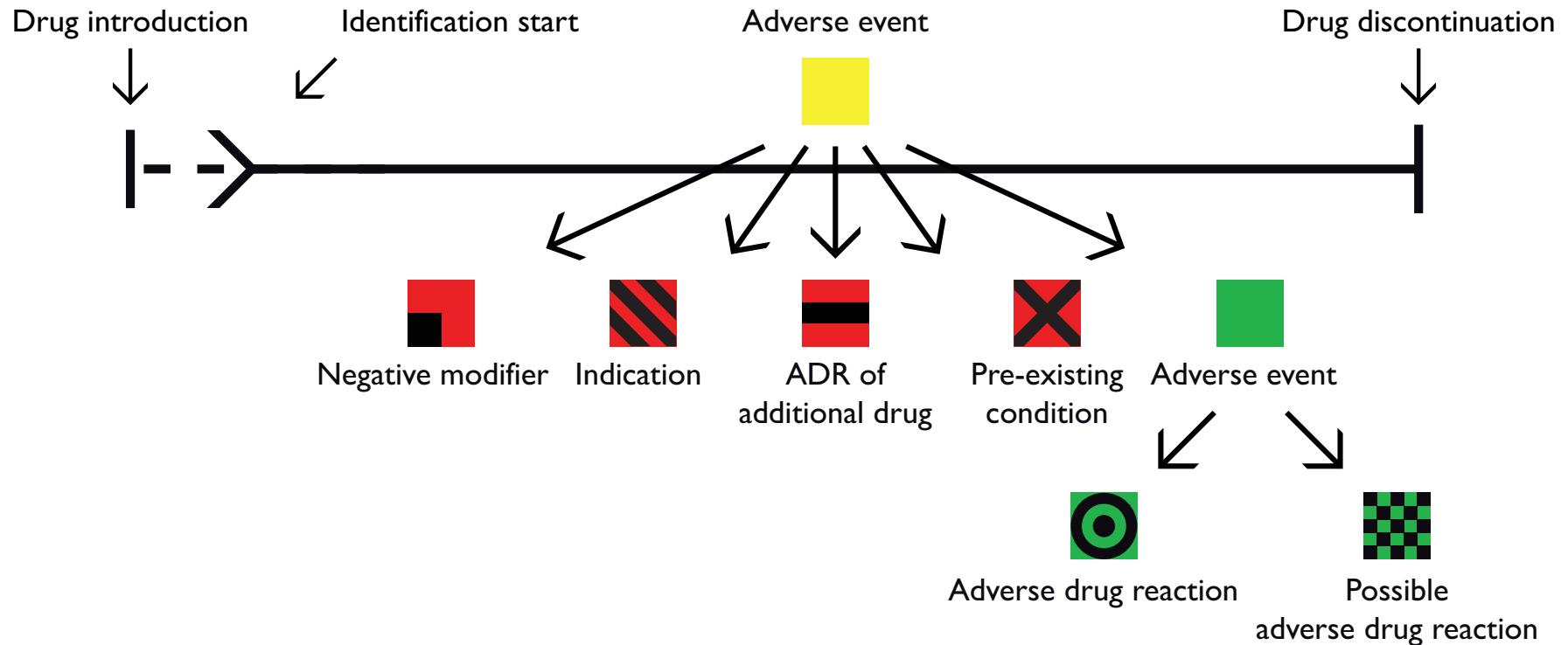
closapine

chlosapine

curated blacklist

temporal correlations

filtering rules

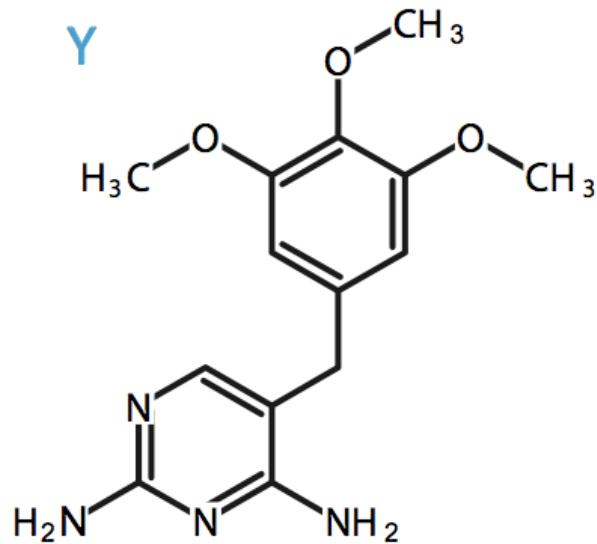
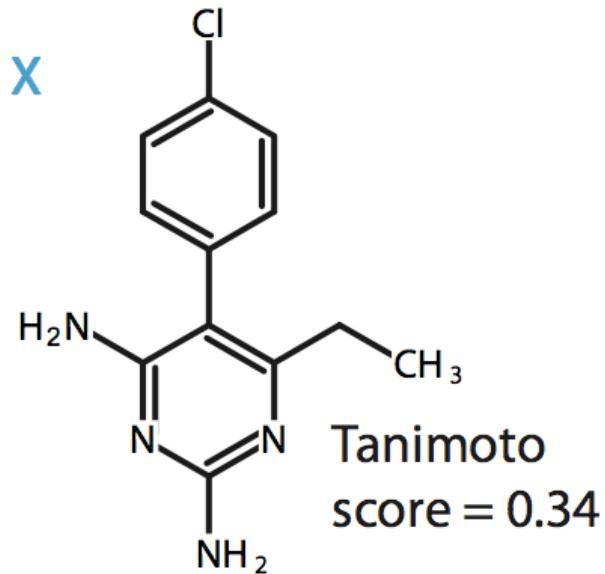


larger cohort

drug combinations

drug target prediction

chemical similarity



similar drugs share targets

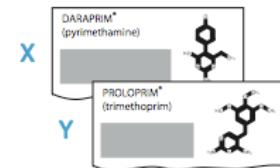
only trivial predictions

chemical perturbations

phenotypic readout

ADR similarity

E Calculation of raw score $S_{X,Y}$

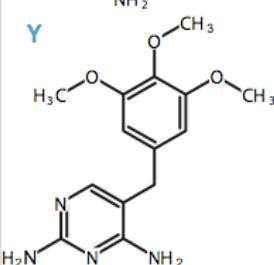
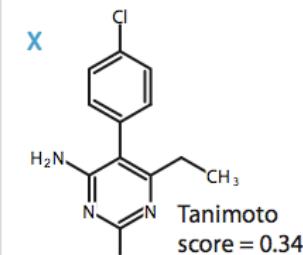


$$S_{X,Y} = \sum_{i \in X \cap Y} r_i \cdot g_i$$

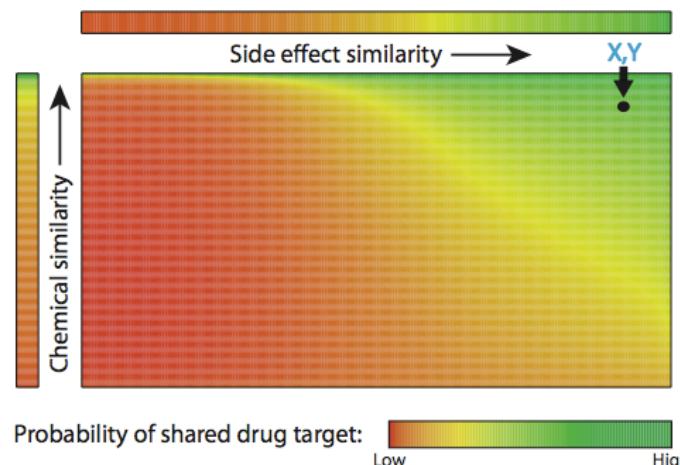
F Randomization yields p-values used as side effect similarity measure



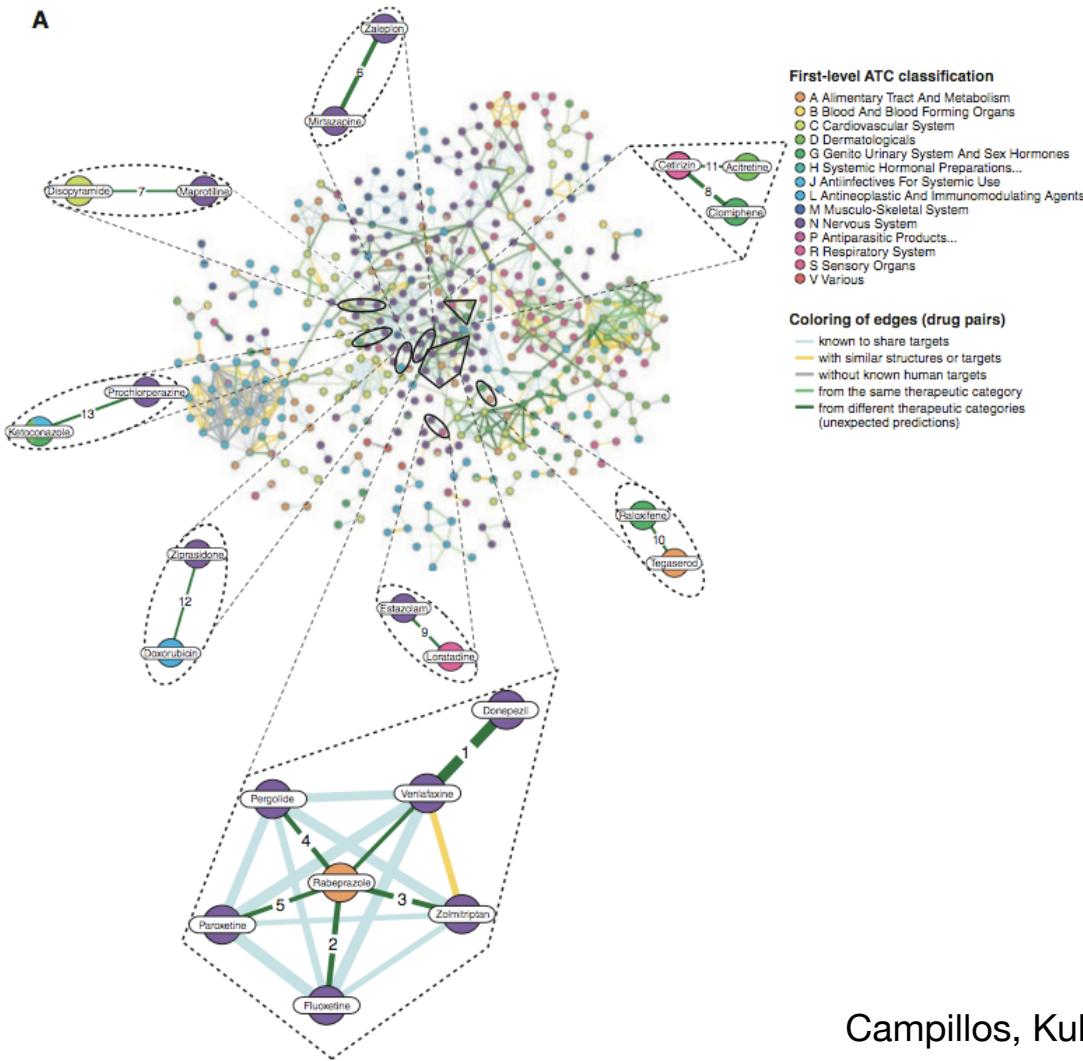
G Chemical similarity



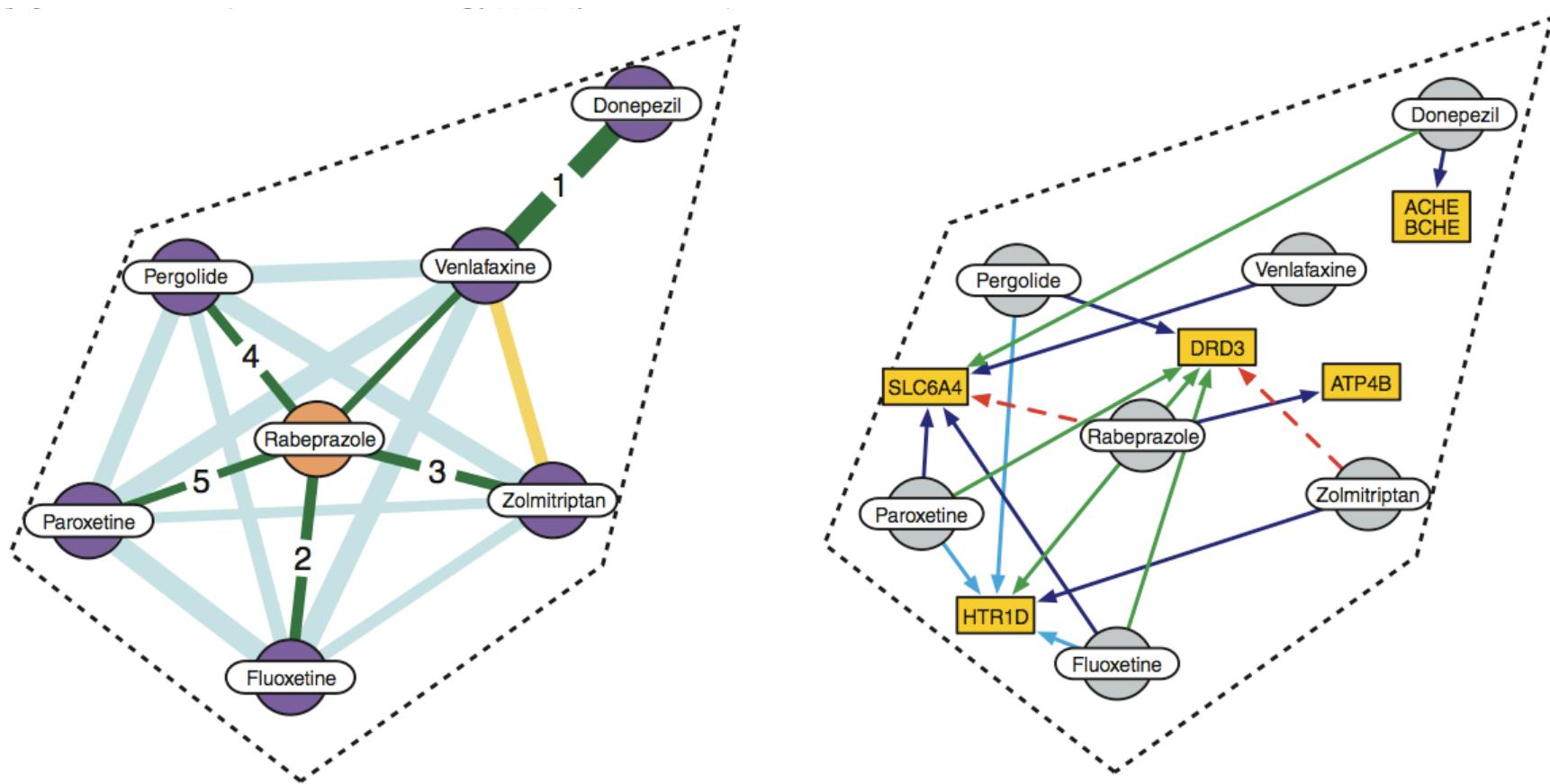
H Combination of side effect and chemical structure similarity for prediction of shared targets



drug-drug network

A

visualize known targets



predict new targets

experimental validation

summary

rewarding

global map of diseases

patient stratification

ADR monitoring

challenging

biased

correlated

incomplete

unstructured

Acknowledgments

Disease trajectories

Anders Bæk Jensen

Mette Kristina Beck

Annelaura Bach Nielsen

Dhouha Grissa

Anders Perner

Tudor Oprea

Pope Moseley

Søren Brunak

Adverse drug reactions

Robert Eriksson

Thomas Werge

Søren Brunak

Drug target prediction

Monica Campillos

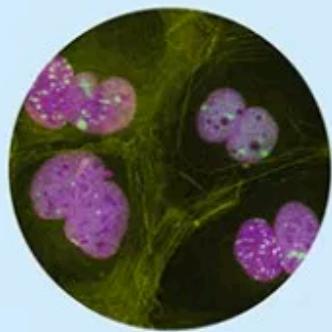
Michael Kuhn

Anne-Claude Gavin

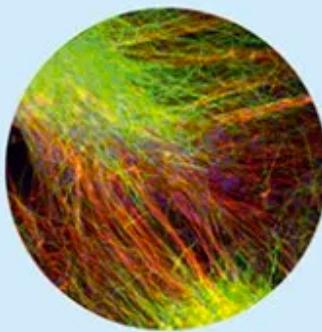
Peer Bork

Copenhagen Bioscience PhD

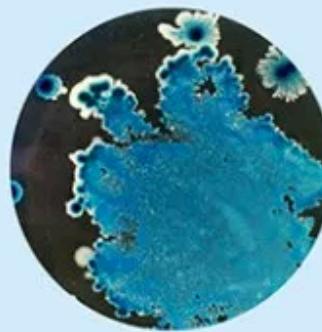
APPLICATION DEADLINE JANUARY 6, 2019:
www.novonordiskfonden.dk/da/cphbiosciencephd



Novo Nordisk Foundation
Center for Protein Research



Novo Nordisk Foundation
Danish Stem Cell Center



Novo Nordisk Foundation
Center for Biosustainability



Novo Nordisk Foundation
Center for Basic Metabolic Research