

SCIENCE POLICY RESEARCH UNIT

The use of bibliometrics as a lens explore corporate innovation in cancer diagnostics

Daniele Rotolo, Frédérique Bone, and Michael Hopkins

SPRU – Science Policy Research Unit

University of Sussex

Brighton, United Kingdom

US
UNIVERSITY
OF SUSSEX

SPRU
SCIENCE POLICY
RESEARCH UNIT

Introduction to 'Cancerscreen' project

funded by an ERC grant held by Stuart Hogarth,
University of Cambridge

This project focuses on the nature and extent of changes underway in the field of cancer screening as a result of several postulated trends:

- The corporatisation of R&D for diagnostic innovation
- Regulatory expansion for diagnostic innovation
- Molecularisation of tests, supported by intellectual property rights, as a driver

The project involves study of industrial dynamics of the molecular diagnostics industry particularly at the cancer field level as well as micro-level case studies of leading firms as well as regulatory mapping of the US and EU markets.

Why do we expect corporate entities to publish in diagnostic innovation? (1)

There has been considerable scientometric research on examining firms' publication activity in pharma/biotech/life science (Camerani *et al.*, in preparation).

- Evidence that firms publish a significant amount of their R&D (e.g. Hicks, 1995; Tijssen *et al.*, 1996; Li *et al.*, 2015)
 - Growth (e.g. Arora *et al.*, 2018)
 - Decline (e.g. Godin, 1996)
- Why do firms publish?
 - "*Ticket of admission*" for firm to access the academic network (Rosenberg, 1990)
 - "*Scientists pay to be scientists*" (e.g. Stern, 2004; Sauermann & Roach, 2014)
 - Signalling quality of products (e.g. Nelson, 1990; Arora *et al.*, 2017)
 - Attracting investors (e.g. Belenzon & Patacconi, 2014, Simeth & Cincera, 2015)
 - Defensive publishing (e.g. Baker & Mezzetti, 2005)
 - Product approval (e.g. Polidor & Theeke, 2012)
 - Stimulating the adoption of a drug (e.g. Azoulay, 2002)

Our prior research has highlighted how in cervical cancer, Digene contributed to the clinical literature on HPV testing

Hogarth et al. (2016) Chapter 5 in Consoli et al. (eds) Medical Innovation.

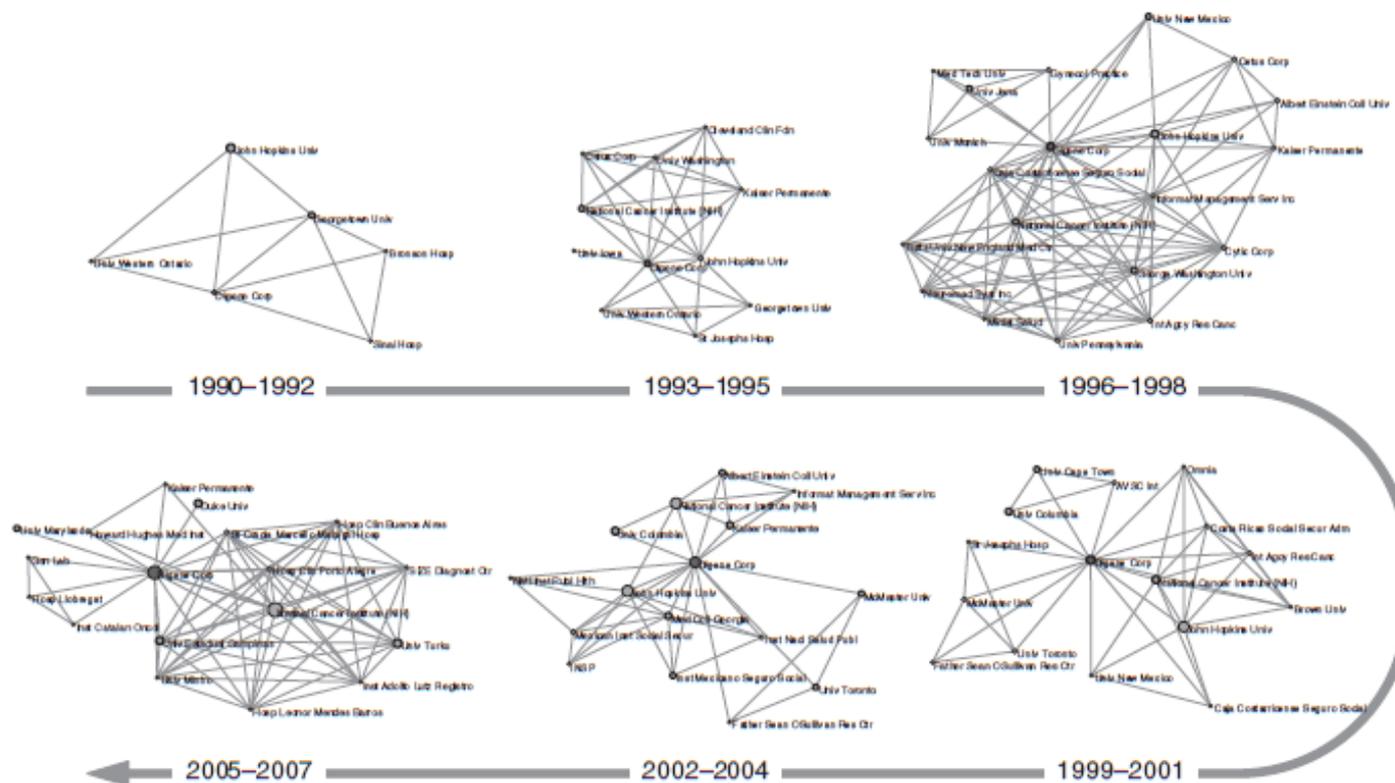


Figure 5.2 Digene's ego-network dynamics (co-authorship network). The size of nodes is proportional to the number of scientific articles related to HPV diagnostics that the given actor published in the relative time window.

Source: authors' elaboration on the basis of ISI Web of Science data and by using the Pajek software package.

Three levels of bibliometric analysis

Macro-level:

The cancer field (e.g. Lung, prostate, Colorectal)

Meso-level:

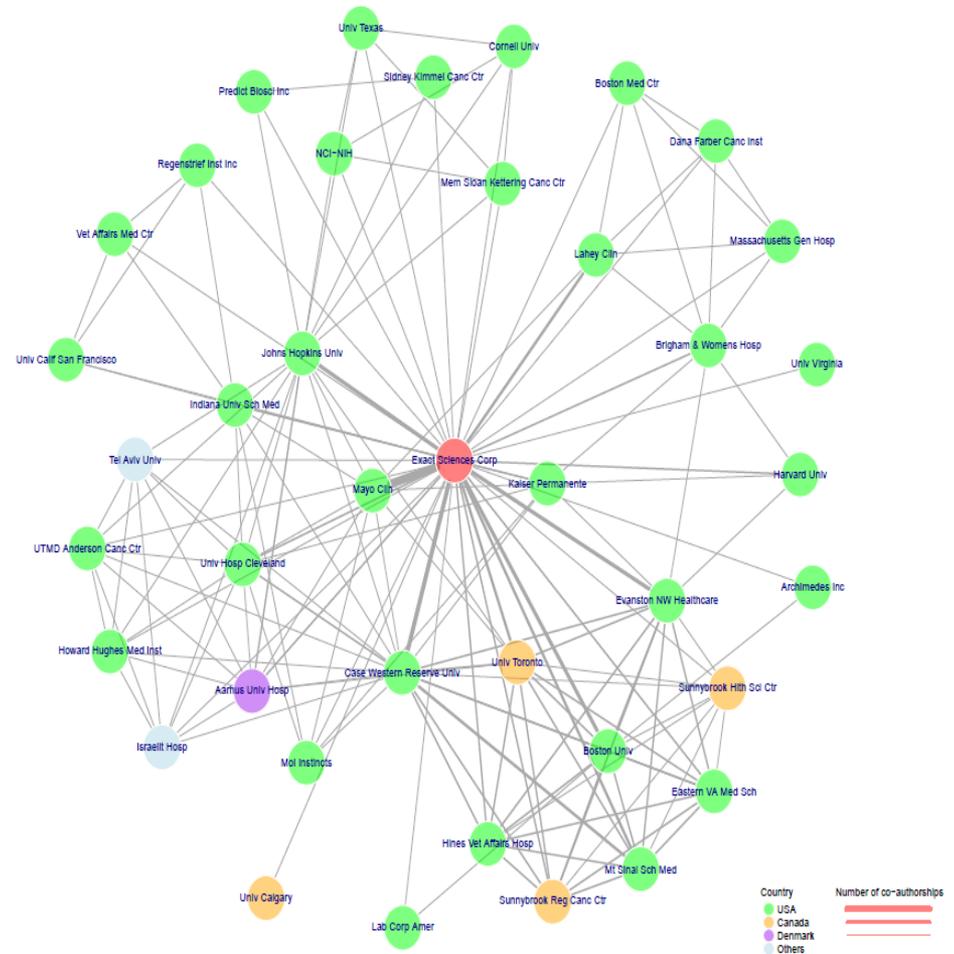
Diagnostic-specific activity within the cancer field

(e.g. early detection, screening)

Micro-level:

Activities relevant to a specific firm

Exact Sciences Corp: co-authorship network (2000–2016)



Identifying cancer related publications using MeSH terms and keywords

Case	Colorectal Cancer	Lung Cancer	Prostate Cancer
MeSH terms (Including children terms)	Colorectal Neoplasms	Lung Neoplasms	Prostatic Neoplasms
Specific terms (key words)	“familial adenomatous polyposis” HNPCC “lynch syndrome” “muir-torre syndrome”	NSCLC pneumoblastoma pneumoblastomas SCLC	-
Site terms to search with general terms (keywords)	bowel colon colonic colorectal rectal rectum	bronchial bronchogenic lung pulmonary	prostate prostatic
Excluded terms	“small bowel”	-	-

Macro level – activity in a cancer field over time

Lung 90-99

mg m2
regimen
response
rate

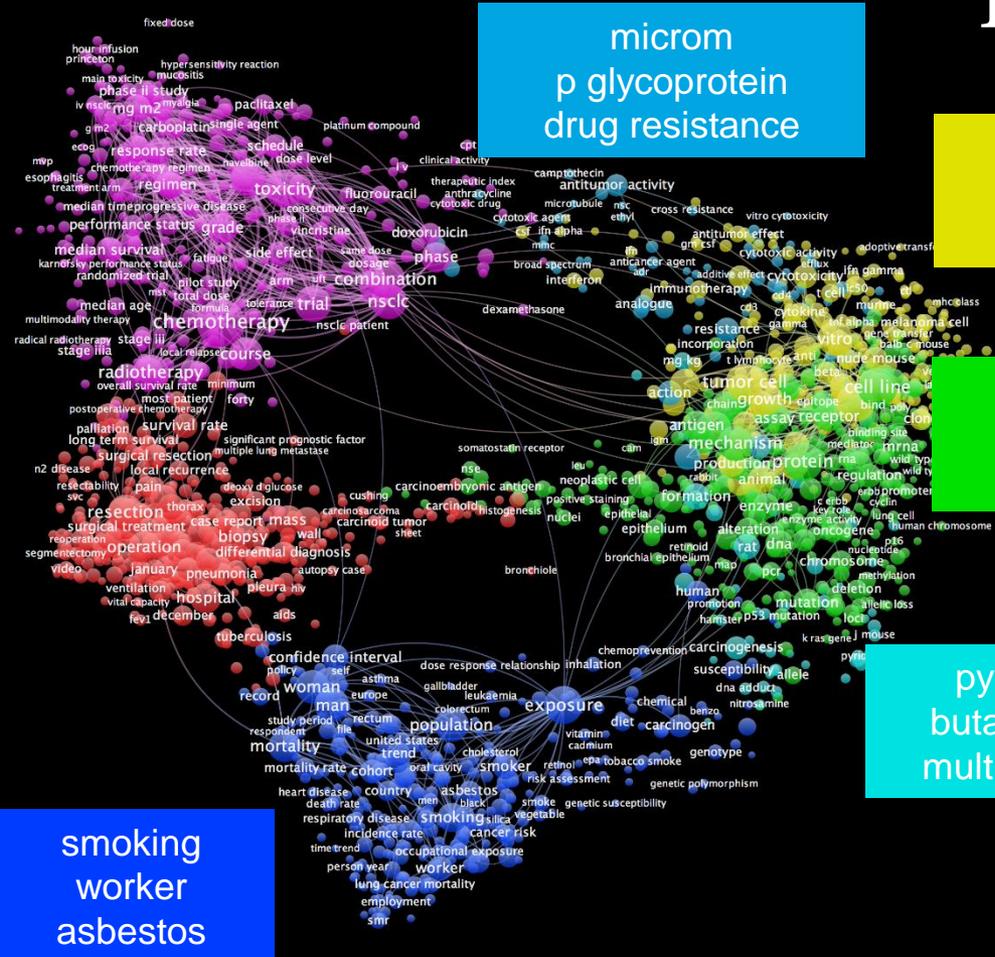
microm
p glycoprotein
drug resistance

clone
molecule
t cell

mutation
mrna
p53

resection
operation
thoracotomy

smoking
worker
asbestos



Macro level – activity in a cancer field over time

Lung 98-07

cycle
advanced non small
cell lung cancer
response rate

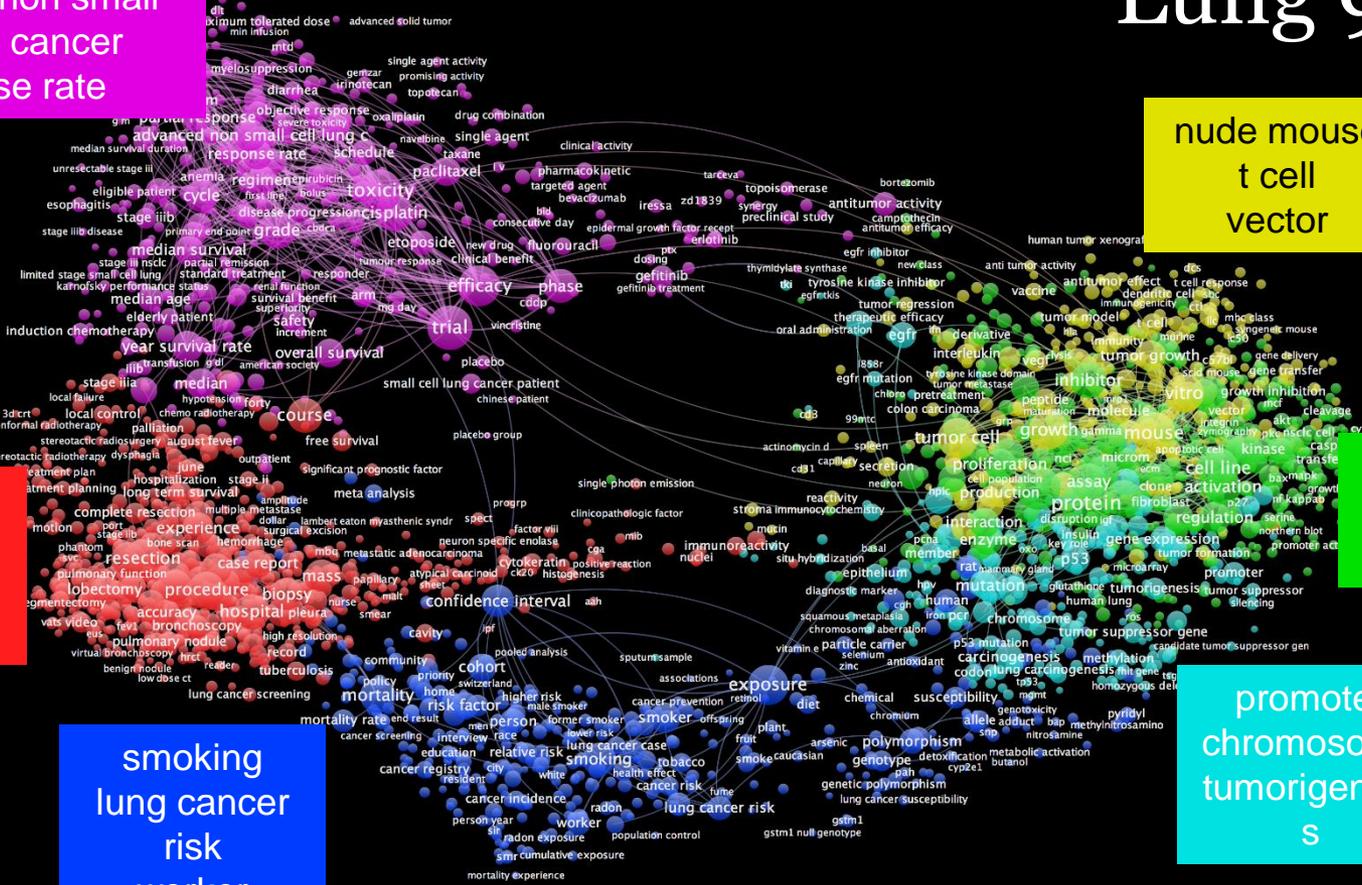
nude mouse
t cell
vector

resection
year old
man
staging

cell line
apoptosis
activation

smoking
lung cancer
risk
worker

promoter
chromosome
tumorigenesi
s



Trends in the “Corporatisation” of R&D?

We investigate corporatisation of research through addresses using the Web of Science:

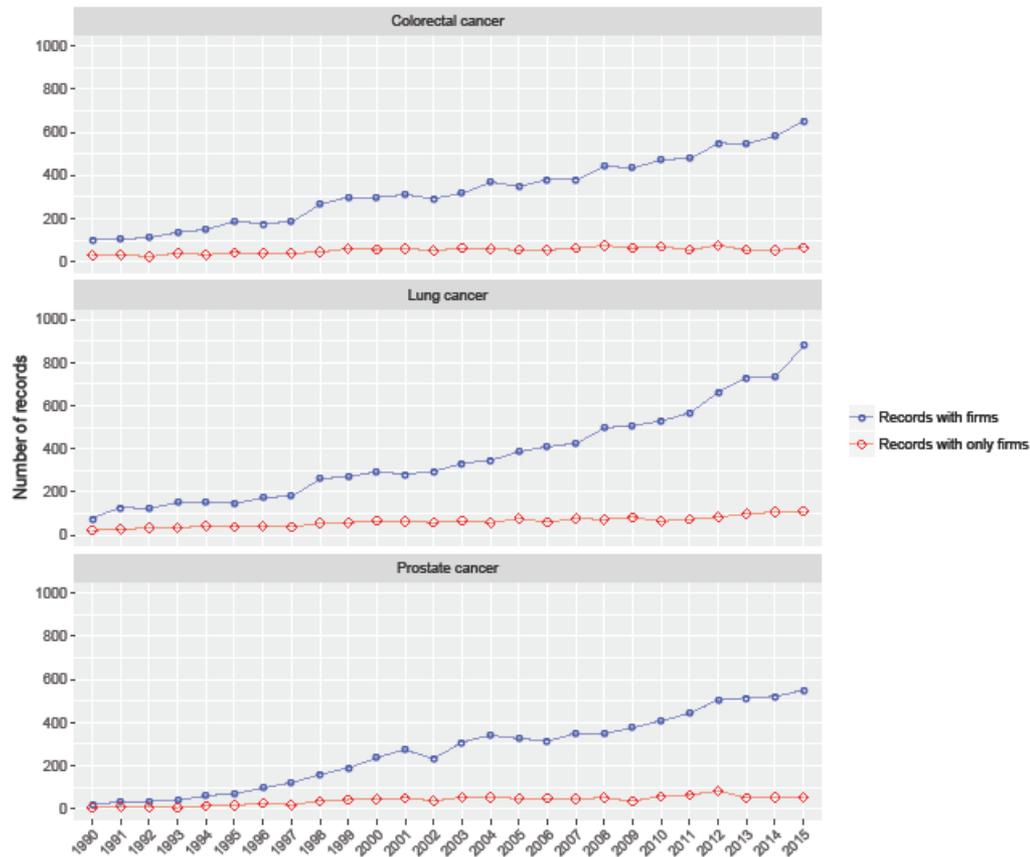
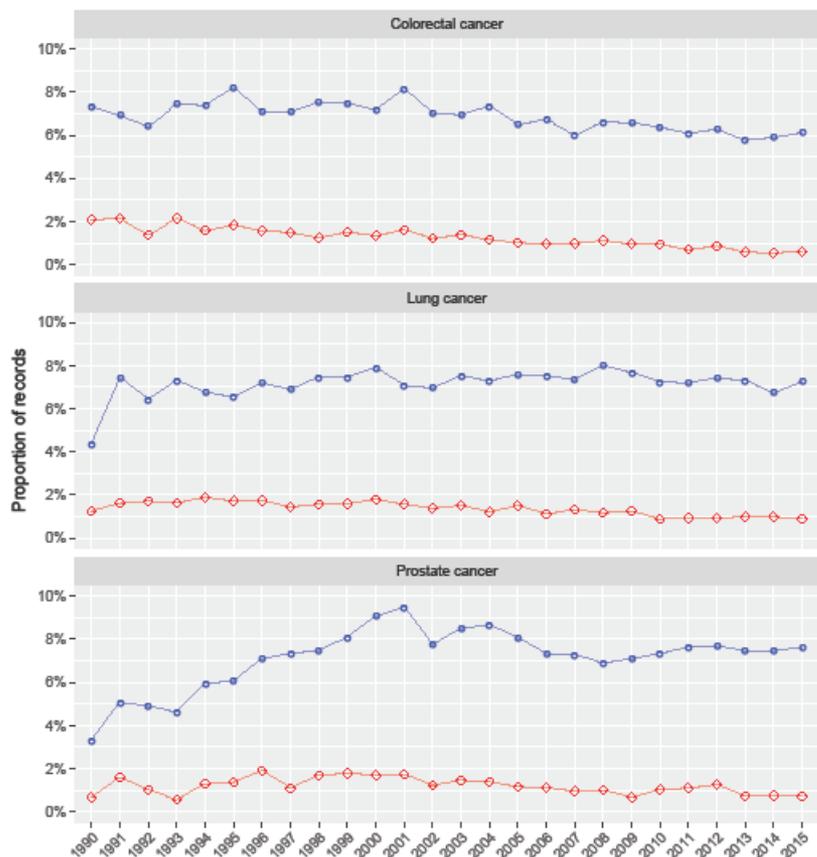
A semi-automated process was adopted to classify firm addresses:

- We use firm names (e.g. for known large pharmaceutical companies)
- We use Business entity abbreviations (e.g. Corp, Inc)
- Non-firm name abbreviations (e.g. Univ, Hosp, Canc Reg)

We were helped by research assistance to check the accuracy of classification and classify the remaining cases.

In the biomedical field, it may not be always clear what ‘corporate’ means: for example private health care providers, doctors ... are all not considered as corporate.

At the macro level: Proportional and absolute levels of corporate publishing in three cancer fields (not only diagnostics)



Proportion of publications with at least one corporate author in three cancer fields (macro and meso levels of analysis)



Micro level studies of corporate publishing: Illustrative comparison of Exact Sciences and Epigenomics Ag

Exact Sciences Corp.

'Our Vision:

Committed to helping win the war on cancer through early detection

Through our Mayo Clinic collaboration, we have found new biomarkers for 10 of the most prevalent cancers. Together, we developed Cologuard.

Now, we're collaborating on new blood-based, patient-friendly cancer screening and diagnostic tests.'

(Corporate website, 2018)

Founded: 1995 (IPO 2001)

HQ: Madison, Wisconsin

Product on the market for colorectal cancer

(DNA detection from stool)

Epigenomics Ag

'We revolutionize the way of cancer diagnostics using our unique, proprietary DNA methylation biomarker technology. Epigenomics develops and commercializes patient-friendly, blood-based diagnostic tests across multiple cancer indications with high medical need. Using blood as a liquid biopsy can improve patient access to cancer screening and thereby contribute to eradicate today's deadliest cancer types such as lung and colorectal cancer. By leveraging our product pipeline and strong intellectual property, we aim to become a global leader in blood-based cancer detection.'

(corporate Website 2018)

Founded: 1998 (IPO in 2004)

HQ: Berlin, Germany

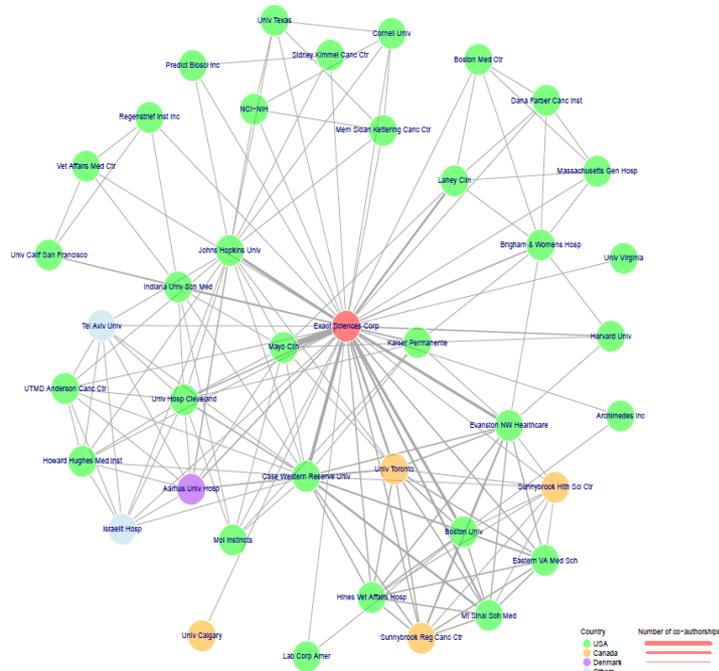
Products on market for lung and colorectal cancer

(DNA methylation detection from venous blood)

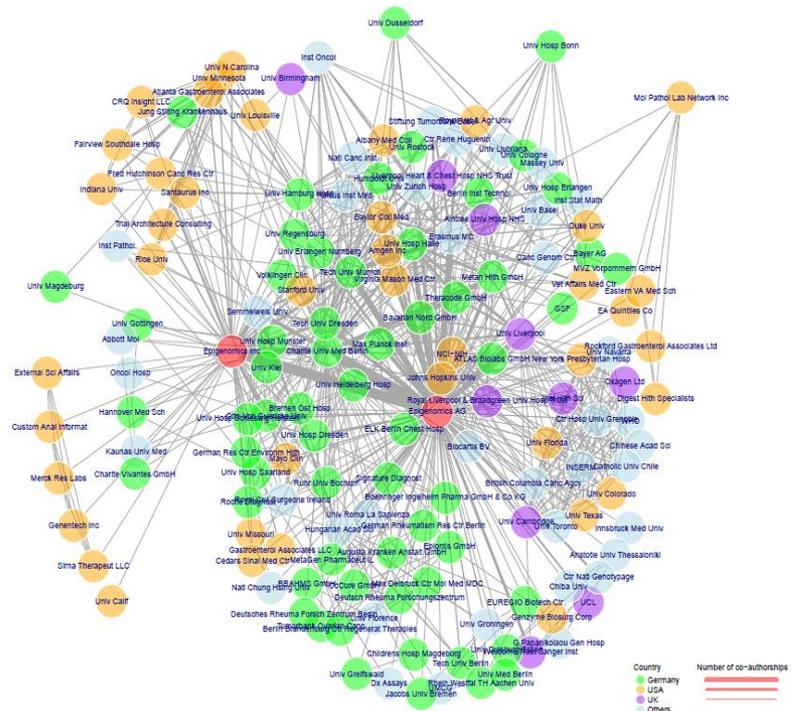


Ego networks of diagnostics firms: revealing key partners

Exact Sciences Corp: co-authorship network (2000–2016)



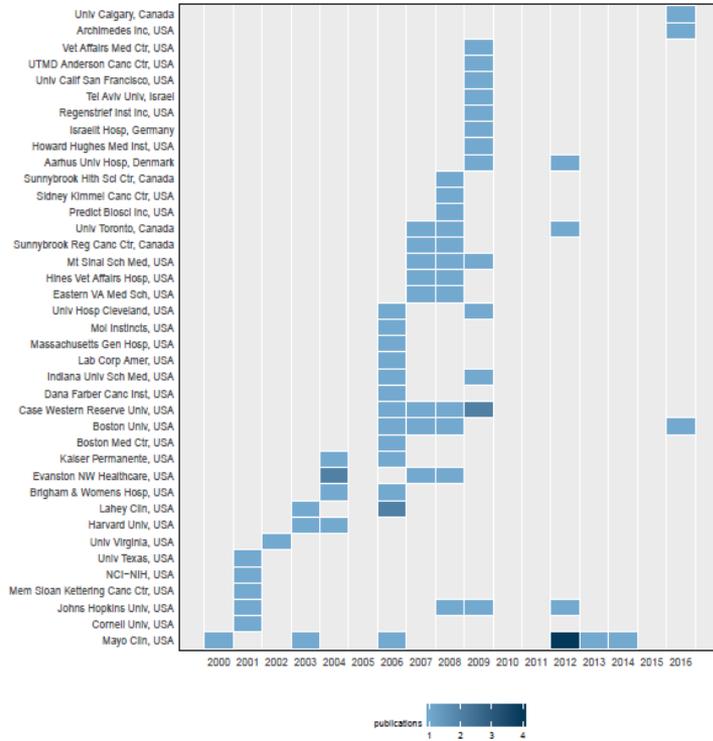
Epigenomics AG/Epigenomics Inc: co-authorship network (2001–2016)



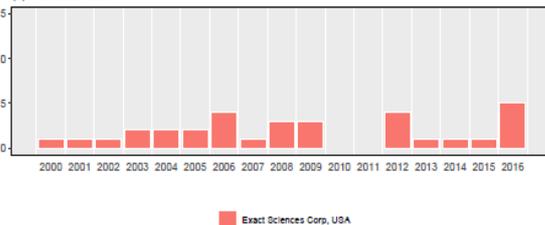
Matrices show persistent partnerships between diagnostics firms and collaborating author organisations over time

Exact Sciences Corp

(a) Co-author organisations (organisations sorted by first co-authorship)



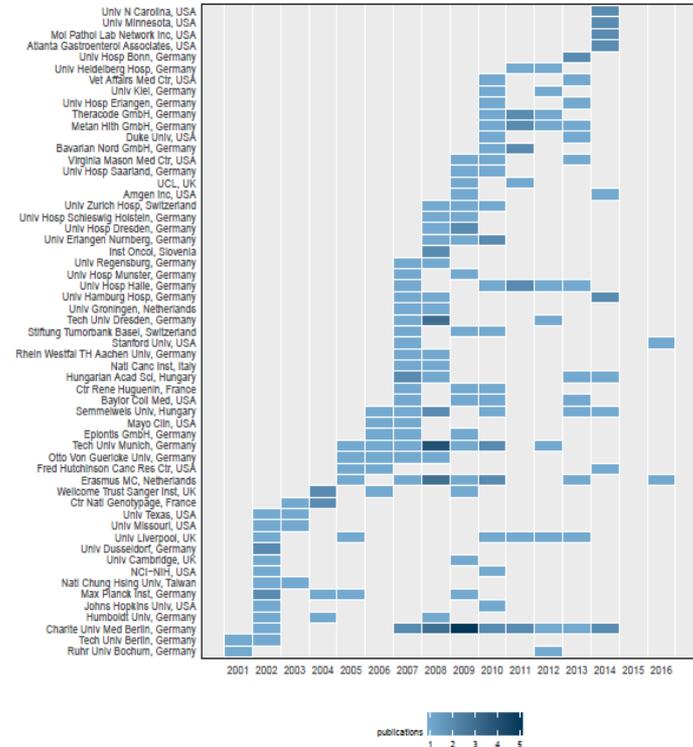
(b) Publication count



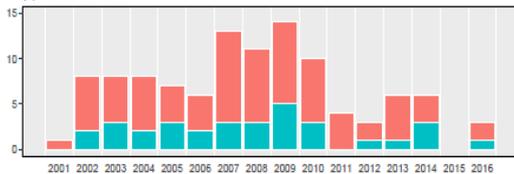
Exact Sciences Corp, USA

Epigenomics AG/Epigenomics Inc

(a) Co-author organisations (68/154 contributing to 90% of publications; organisations sorted by first co-authorship)



(b) Publication count



Epigenomics AG, Germany Epigenomics Inc, USA

Identifying cancer related publications using MeSH terms and keywords

Case	Colorectal Cancer	Lung Cancer	Prostate Cancer
MeSH terms (Including children terms)	Colorectal Neoplasms	Lung Neoplasms	Prostatic Neoplasms
Specific terms (key words)	“familial adenomatous polyposis” HNPCC “lynch syndrome” “muir-torre syndrome”	NSCLC pneumoblastoma pneumoblastomas SCLC	-
Site terms to search with general terms (keywords)	bowel colon colonic colorectal rectal rectum	bronchial bronchogenic lung pulmonary	prostate prostatic
Excluded terms	“small bowel”	-	-

Data retrieved

MEDLINE/PubMed data (1990-2015)

Case	Number of publications		
	Colorectal cancer	Lung cancer	Prostate cancer
Macro-level	172029	182137	117108
Meso 1-level	84319 (49.0%)	85263 (46.8%)	54520 (46.6%)
Meso 2-level	8051 (4.7%)	2650 (1.5%)	4356 (3.7%)

Source: Authors' elaboration.

Identifying relevant cancer publications using MeSH/Keywords*

MeSH terms: MEDLINE/PubMed records tagged with cancer specific MeSH terms and/ or keywords:

1. Data collection using keywords – On the basis of the US NLM and NCI's cancer search
 - Site specific cancer terms (e.g. lynch syndrome)
 - General cancer terms (e.g. tumour) to search with site terms (e.g. colon)

We are interested in the 'disease', so we excluded terms of specific diagnostic, therapeutic, or surgery techniques (e.g. PSA test, chemotherapy) to minimize selection bias

2. Data filtering – MEDLINE/PubMed records that satisfied one of the below conditions
 - Presence of at least **one site specific cancer term** in the title/abstract/keywords
 - Presence of at least **one general cancer term** and of a **site term** in the title/abstract/keywords and these terms were not distant more than 3 words (we tested for 4 and 5, but considerable noise is introduced)

Data matching of MEDLINE/PubMed with Web of Science (WoS) and Scopus on the basis of PMIDs, to retrieve authors' affiliation addresses (Rotolo and Leydesdorff, 2015)

We were helped by research assistants to check the accuracy of classification

General cancer terms used in search*

General cancer terms

adenoacanthoma	cancerous	lymphoma	papillomata
adenoacanthomas	cancers	lymphomas	papillomatosis
adenocanthoma	carcinogen	mesothelioma	sarcoma
adenocanthomas	carcinogenesis	mesotheliomas	sarcomas
adenocarcinoma	carcinogenic	metastasectomy	sarcomatosis
adenocarcinomas	carcinogens	metastases	"signet ring"
anaplasia	carcinoma	metastasis	tumor
angioreticuloma	carcinomas	metastatic	tumorigenesis
angiosarcoma	carcinomatosis	micrometastases	tumorigenic
angiosarcomas	carcinosarcoma	micrometastasis	tumorigenesis
"anti-neoplastic"	carcinosarcomas	"myelodysplastic-myeloproliferative"	tumorigenic
anticancer	"clonal evolution"	neoplasm	tumorigenesis
anticarcinogenesis	cocarcinogenesis	neoplasms	tumorigenic
anticarcinogenic	cystadenocarcinoma	"neoplastic stem cells"	tumors
antineoplastic	cystadenocarcinomas	oncocytoma	tumour
antioncogene	cystadenoma	oncocytomas	tumourgenensis
antioncogenes	cystadenomas	oncogen	tumourgenic
antitumor	cystoma	oncogene	tumourigenesis
antitumors	cystomas	oncogenes	tumourigenic
antitumour	desmoplastic	oncogens	tumourogenesis
antitumours	lymphangiosarcoma	oncoprotein	tumourogenic
blastoma	lymphangiosarcomas	oncoproteins	tumours
blastomas	lymphoepithelioma	papilloma	
cancer	lymphoepitheliomas	papillomas	

Excluding

TNF, "tumour necrosis factor", "tumor necrosis factor", *suppressor*

suppressor

FUS, *COH*

Source: Authors' elaboration based on the US NLM and NCI's cancer search

Data retrieved*

MEDLINE/PubMed data (1990-2015)

Case	Number of publications		
	Colorectal cancer	Lung cancer	Prostate cancer
MeSH terms (including children terms)	131250	132422	89059
Keywords			
Before filtering	63557	81108	29079
After filtering	41741	50658	29079
Total	172991	183080	118138

Note: MEDLINE/PubMed was queried on 27 September 2017.

Data retrieved*

Data coverage (1990-2015)

Case	Number of publications		
	Colorectal cancer	Lung cancer	Prostate cancer
All document types			
MEDLINE/PubMed	172991	183080	118138
WoS coverage	137490 (79.5%)	139633 (76.3%)	95796 (81.1%)
Scopus coverage	161184 (93.1%)	163791 (89.5%)	109098 (92.4%)
Selected document types			
MEDLINE/PubMed	172029	182137	117108
WoS coverage	137019 (79.6%)	139161 (76.4%)	95241 (81.3%)
Scopus coverage	160308 (93.2%)	162989 (89.5%)	108182 (92.4%)
Articles and reviews			
MEDLINE/PubMed	162549	173105	110139
WoS coverage	128978 (79.3%)	131694 (76.1%)	89369 (81.1%)
Scopus coverage	151214 (93.0%)	154692 (89.4%)	101455 (92.1%)

Note: See next slide for the list of selected/excluded document types.

Source: Authors' elaboration.

Sample coverage

