

# Identifying and Validating Networks of Oncology Biomarkers Mined From the Scientific Literature

Kim Wager<sup>1</sup> , Dheepa Chari<sup>2</sup>, Steffan Ho<sup>2</sup>,  
Tomas Rees<sup>1</sup> , Orion Penner<sup>3</sup> and Bob JA Schijvenaars<sup>3</sup>

<sup>1</sup>Oxford PharmaGenesis, Oxford, UK. <sup>2</sup>Pfizer, Inc., New York, NY, USA. <sup>3</sup>Digital Science, London, UK.

Cancer Informatics  
Volume 21: 1–15  
© The Author(s) 2022  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/11769351221086441



**ABSTRACT:** Biomarkers, as measurements of defined biological characteristics, can play a pivotal role in estimations of disease risk, early detection, differential diagnosis, assessment of disease progression and outcomes prediction. Studies of cancer biomarkers are published daily; some are well characterized, while others are of growing interest. Managing this flow of information is challenging for scientists and clinicians. We sought to develop a novel text-mining method employing biomarker co-occurrence processing applied to a deeply indexed full-text database to generate time-interval-delimited biomarker co-occurrence networks. Biomarkers across 6 cancer sites and a cancer-agnostic network were successfully characterized in terms of their emergence in the published literature and the context in which they are described. Our approach, which enables us to find publications based on biomarker relationships, identified biomarker relationships not known to existing interaction networks. This search method finds relevant literature that could be missed with keyword searches, even if full text is available. It enables users to extract relevant biological information and may provide new biological insights that could not be achieved by individual review of papers.

**KEYWORDS:** Biomarkers, cancer, natural language processing, network analysis, text-mining

**RECEIVED:** November 4, 2021. **ACCEPTED:** February 18, 2022.

**TYPE:** Original Research

**FUNDING:** The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was conducted as a collaboration between Digital Science, Oxford PharmaGenesis and Pfizer. Pfizer is the study sponsor. Editorial support was provided by Alessia Gonzato at Oxford PharmaGenesis, Oxford, UK and was funded by Pfizer.

**DECLARATION OF CONFLICTING INTERESTS:** The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Kim Wager and Tomas Rees are employees of Oxford PharmaGenesis, which was a paid consultant to Pfizer in connection with the development of this manuscript. Orion Penner and Bob JA Schijvenaars are employees of Digital Science, which was a paid consultant to Pfizer in connection with the development of this manuscript. Dheepa Chari and Steffan Ho are employees of Pfizer.

**CORRESPONDING AUTHOR:** Dheepa Chari, Pfizer, Inc., 235 East 42nd Street, New York, NY 10017, USA. Email: dheepa.chari@pfizer.com

## Introduction

Biomarkers, as measurements of defined biologic characteristics, can play a pivotal role in estimations of disease risk, early detection, differential diagnosis, assessment of disease progression and outcomes prediction.<sup>1</sup> Cancer biomarkers used in clinical practice often measure specific molecular markers, including somatic gene alterations or protein expression, but may also measure characteristics defined by gene expression signatures or tumour imaging.<sup>2</sup>

Applications of biomarkers in drug development enable patient stratification to optimize therapeutic efficacy and safety, and measurement of drug-dependent biologic responses (pharmacodynamics) to evaluate mechanisms of drug action.<sup>3</sup> Owing to the wide variety of biomarker categories, the advent of 'omic' data and the interest in patient stratification for personalized medicine, studies of cancer biomarkers are published daily. As a result of in-depth research over time, some are well characterized, whereas others are emerging biomarkers of growing interest.

Although the biomedical literature represents a valuable source of cancer biomarker-related information, managing this flow of information is challenging for scientists and clinicians. There are limited biomarker data in publicly available databases, but it is likely that further insights remain hidden in the academic literature owing to the limitations of standard keyword-based searching and the sheer volume of available literature. To help to address this challenge, semi-automated text-mining and

innovative approaches to the synthesis and visualization of the output are required.

Recently, text-mined gene-disease term co-occurrence in abstracts has been used to suggest additional genes to be included in cancer gene panels, by identifying the characteristics of an existing gene panel and suggesting genes with related features.<sup>4</sup> Other text-mining pipelines seek to identify disease-related mutations using content derived from titles and abstracts,<sup>5,6</sup> while others utilize full-text publications but are restricted by inclusion of open-access articles only.<sup>7,8</sup> Exploration of gene-gene co-occurrence networks from abstract-derived text-mined data indicate that clusters may contain genes that are directly or indirectly related based on physical interactions, co-expression or through signalling pathways.<sup>8</sup> However these approaches, which are limited in scope, have so far demonstrated limited utility in enhancing or facilitating the process through which researchers access the literature to identify and to prioritize biomarkers of interest.

Here we report the development of a novel text-mining method that employs not only biomarker co-occurrence processing applied to a deeply indexed full-text database (Dimensions),<sup>9</sup> but also utilizes time-interval delimited networks to identify biomarkers of greater potential biologic relevance based on the emergence over time of term co-occurrence.

The development process is described using an associated interactive open-access research tool with examples of the



**Table 1.** Search terms used to identify specific cancer sites.

CANCER SITE	CANCER TERMS
Bladder	Bladder cancer, bladder carcinoma, urinary bladder neoplasms
Breast	Breast cancer, breast carcinoma, breast neoplasms
Colorectal	Bowel cancer, colon cancer, colonic neoplasms, colorectal cancer, rectal cancer
Lung	Lung cancer, lung neoplasms, NSCLC, SCLC, small cell lung carcinoma
Prostate	Prostate cancer, prostatic neoplasms
Renal	Kidney cancer, kidney neoplasms, renal cell carcinoma, RCC

Abbreviations: NSCLC, non-small cell lung carcinoma; RCC, renal cell carcinoma; SCLC, small cell lung carcinoma.

application of this approach to evaluate biomarkers of potential emerging scientific interest in cancer.<sup>10</sup>

## Methods

### *Identifying relevant publications mentioning biomarkers and cancer*

A publicly available data set comprising 726 cancer biomarkers was obtained from the Early Detection Research Network (EDRN),<sup>11</sup> an initiative of the National Cancer Institute focusing on the clinical application of early cancer detection strategies, and was used to seed our literature searches.

To identify an initial set of publications of interest, we performed co-occurrence searches for each biomarker in the EDRN data set in 20-word proximity to terms relating to 6 cancer sites (bladder, breast, colorectal, lung, prostate and renal [Table 1]) using the Dimensions-linked scholarly information platform. We extracted biomarker relationships to cancer terms from searches of full-text publications in English, including proceedings and preprints, from 1 January 2015 to 31 December 2020.

This search methodology identified articles of interest, defined as those with at least 1 biomarker and at least 1 term relating to a given cancer site within a 20-word proximity (denoting relevance to one of the specified cancer sites). To focus our study on biomarkers of emerging research interest, those with less than 5 or more than 1000 unique publication mentions were excluded.

### *Identifying relevant publications co-mentioning biomarkers by text mining*

Following the initial identification of publications of interest, we text-mined the resulting corpus. Each publication was pre-processed through tokenization (to simple unigrams), removal of punctuation and stop words, and conversion of uppercase text to lowercase.

To identify biomarkers that are likely to be mentioned in a shared biological context, we searched for the co-occurrence of 1 biomarker with another within a 20-word proximity; these were defined as co-occurring biomarker pairs. Co-occurrence proximity windows of 10 words and 30 words were also

experimented with but *ad hoc* analysis indicated that a 10-word window is slightly too short and often misses potentially meaningful co-occurrences. On the other hand, 30-word windows are too liberal and produce a network too dense to be meaningful. However, to identify further which biomarker pairs were more likely to represent biologically relevant relationships, only biomarker pairs that co-occurred more than once in the same publication and that also had co-occurred in at least 2 publications were included.

### *Network analysis*

We generated 7 networks, 1 for each cancer site and a ‘cancer-agnostic’ network that included all publications identified in the search for each of the 6 cancer sites. To help us to understand the extent of co-occurrence between the selected biomarkers, we then generated undirected co-occurrence networks. Using the NetworkX Python package, 1 network was produced for each cancer site, in addition to 1 network containing links found across all cancer sites we investigated. Each node in the resultant networks represents a biomarker and an edge between 2 biomarkers represents co-occurrence. A given co-occurrence is considered significant if it appears at least twice in a given publication. An edge is formed between 2 biomarkers if a significant co-occurrence is discovered in at least 2 separate publications.

Edge weight was used to calculate the betweenness centrality of the nodes and the cluster structure of the network. The weight was calculated as  $n_{ij} / \min(n_i, n_j)$ , where  $n_{ij}$  is the number of publications in which a significant co-occurrence was discovered between biomarkers  $i$  and  $j$ , and  $n_i$  and  $n_j$  are the number of publications in which biomarkers  $i$  and  $j$  were found (respectively, and for the specific cancer). This formulation of weight was chosen because of 2 desirable properties. First, and rather naturally, a higher number of publications containing co-occurrence results in a higher weight, and hence strength of relationship between the 2 nodes. Second, because of the  $\min(n_i, n_j)$  in the denominator, edge weight remains strong if the association between the 2 biomarkers is asymmetric (ie, A is strongly associated to B, but B is not strongly associated to A).

The converse being that the weight remains low in the case in which you are considering 2 very commonly mentioned biomarkers, that just by chance will inevitably co-occur with each other in some publications. For each biomarker node, only the 4 strongest edges were included in the network.

The biomarker networks, constructed as described above, are published on the Network Data Exchange (NDEx) platform,<sup>12</sup> and the nodes and edges are enriched with a variety of metadata.

On the assumption that, compared with the entire network, clusters of co-occurring biomarkers are more likely to represent biologically relevant relationships, highly connected clusters were built using the Leiden algorithm in the `leidenalg` Python package.<sup>13</sup> Specifically, we used the `ModularityVertexPartition` modularity implementation and optimized over 1000 iterations using the `Optimiser().optimise_partition` function. To prepare the graph data for compatibility with the `leidenalg` package, the network was first converted to the `igraph` format.

To sharpen further our focus on biomarkers with emerging interest, publication growth rate was determined by calculating a linear fit of normalized publication number over time (in years) for all biomarkers in each cancer site. To identify clusters with the fastest-growing scientific interest, we calculated the mean publication growth rate across all biomarkers in each cluster.

### Contextual analysis

We reviewed the textual context for selected biomarker publications to provide insights into the effectiveness of the methodology for identifying meaningful connections between biomarkers. This allowed us to extract biological processes and pathways linked to the biomarkers and cancer biology. Search results were classified as ‘successful’ if one of the co-occurring biomarker pairs was found in proximity to the desired cancer site and the biomarker co-occurrence was biologically meaningful. We classified identified articles as an ‘unsuccessful’ hit if biomarkers were found in proximity to a specific cancer but only in the reference section of the article or were incorrectly associated with the target cancer site. To gather data on Mendeley captures, we used the Altmetric platform and performed the analysis in R. We selected 3 different sets of publications using different approaches (detailed below), each seeking to identify biomarkers or publications of highest interest. As a starting point for each, we identified biomarker clusters that exhibited high publication growth rates. The fastest-growing clusters contained few biomarkers and were, therefore, susceptible to skew from a single fast-growing biomarker. For this reason, we selected from clusters containing at least 10 biomarkers.

- (1) Biological processes related to prominent biomarker pair. From the fastest-growing cluster across all 6 cancer networks, we identified the biomarker pair with

the highest number of co-occurrences, either internal or external to the cluster, and examined all their connections with other biomarkers. First, we looked for these pairings in biological pathways databases, namely the Biological General Repository for Interaction Datasets (BioGRID),<sup>14</sup> Reactome,<sup>15</sup> HumanNet v3<sup>16</sup> and (HIPPIE).<sup>17</sup> Next, we examined the textual context within all publications that mentioned the target biomarker pair.

- (2) Biological context of biomarker mentions (cancer-specific). We then identified the cluster with the second-fastest growth rate. Instead of reviewing in depth a single pair, we examined the 20 publications within the cluster with the highest number of Mendeley captures, taking Mendeley saves as a proxy for scholarly interest.
- (3) Biological context (cancer-agnostic). Lastly, we identified the fastest-growing cluster in the cancer-agnostic network, and examined the top 20 papers by Mendeley captures.

In all cases, biological processes associated with biomarkers were mapped to the National Cancer Institute Thesaurus (NCIt) ontology.<sup>18</sup>

### Gene set enrichment analysis

Gene set enrichment analysis on biomarkers contained in the 3 clusters identified above was carried out using the R package `enrichR`, an interface to the `Enrichr` database.<sup>19</sup> Biomarkers contained within each cluster defined a gene set used to query against terms in the Kyoto Encyclopedia of Genes and Genomes (KEGG) molecular pathways and the Gene Ontology (GO) Biological Process libraries. Enrichment terms were ranked by *P* value and genes overlapping with the annotated genes sets identified. *P* values were computed using the hypergeometric test (Fisher’s exact test), which assumes a binomial distribution and independence of any gene belonging to any gene set. The null hypothesis was that proportion of genes in the cluster gene set annotated with a given pathway or process term did not differ from the overall proportion of genes in the genome annotated with the same pathway or process.

## Results

### Biomarker co-occurrence

The Dimensions search identified 255 942 unique full-text publications. Many of these publications were relevant to more than 1 cancer site (Table 2). Of these publications, 92 395 contained at least 1 biomarker pair. The results of our searches and network analysis are summarized in (Figure 1).

The set of pairwise biomarker co-occurrences spanned 31 550 unique pairs across all cancer sites; the most commonly co-occurring biomarker pairs were matrix metalloproteinase

(MMP)1-MMP3, microRNA (*miR*)21-*miR*210 and *miR*126-*miR*21, with co-occurrences in 820, 632 and 510 publications, respectively.

### Biomarker networks

*Overview of networks.* We generated biomarker co-occurrence networks for each of the 6 cancer sites and the cancer-agnostic data set, accessible on the NDEx platform. To take forward our results for validation and further analysis, we identified the clusters with the highest mean publication growth rate for each network (Figure 2).

*Biological processes related to prominent biomarker pair.* Based on publication growth rate, we selected renal cancer cluster 1 (Figure 3). Renal cancer cluster 1 comprised 354 unique publications: 311 associated with its nodes, reflecting

publications co-mentioning the biomarker in this cancer site, and 140 associated with its edges, representative of biomarker co-occurrences.

The most mentioned biomarker in renal cancer cluster 1 was C-X-C motif chemokine ligand (CXCL)5 with 74 publications, whereas the biomarker pair with the most co-occurrences either internal or external to the cluster was CXCL5-CXCL2 with 122 co-mentions in 34 publications (Figure 4). Twenty of the 42 biomarker pairs were already annotated in known biological pathways databases (Supplemental Table 1).

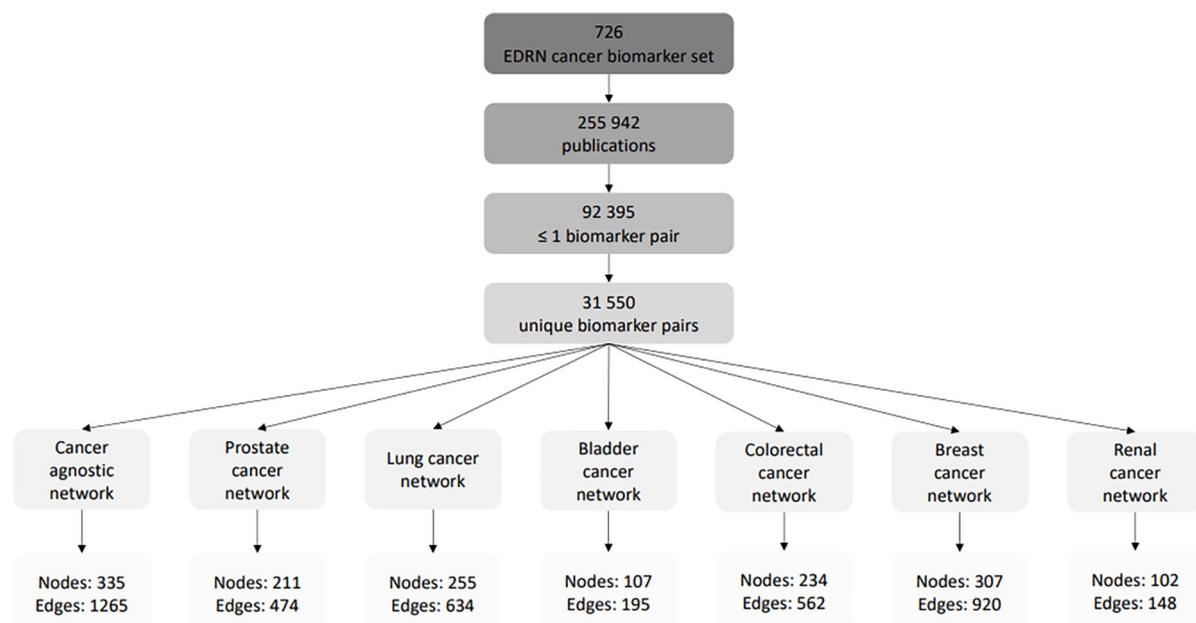
To assess whether our methodology successfully identified literature references describing biologically relevant biomarker relationships, we manually reviewed each publication. All 34 publications were valid in terms of their relevance to cancer biology, with 16 being specific to renal cancer, 3 not being specific to a cancer site and 15 having incorrect identifications of cancer site. The majority of the papers (19/34) were narrative reviews, with the remainder being preclinical reports (13/34), a phase 1 clinical trial (1/34) and a cohort study (1/34). Exploration of these papers using the number of Mendeley captures as a proxy for academic interest revealed that articles discussing chemokines as therapeutic targets were of most interest.

Identified biological processes were mapped to the NCI ontology<sup>14</sup> and were consistent with a proinflammatory role for CXCL5 and CXCL2 acting through their common receptor CXCR2 on neutrophils in the tumour microenvironment, influencing angiogenesis, myeloid cell infiltration and metastasis (Supplemental Table 2). Evaluation of remaining biomarker pairs in this renal cluster revealed the prevalence of chemokines, matrix metalloproteinases and other regulators of cell-matrix interactions.

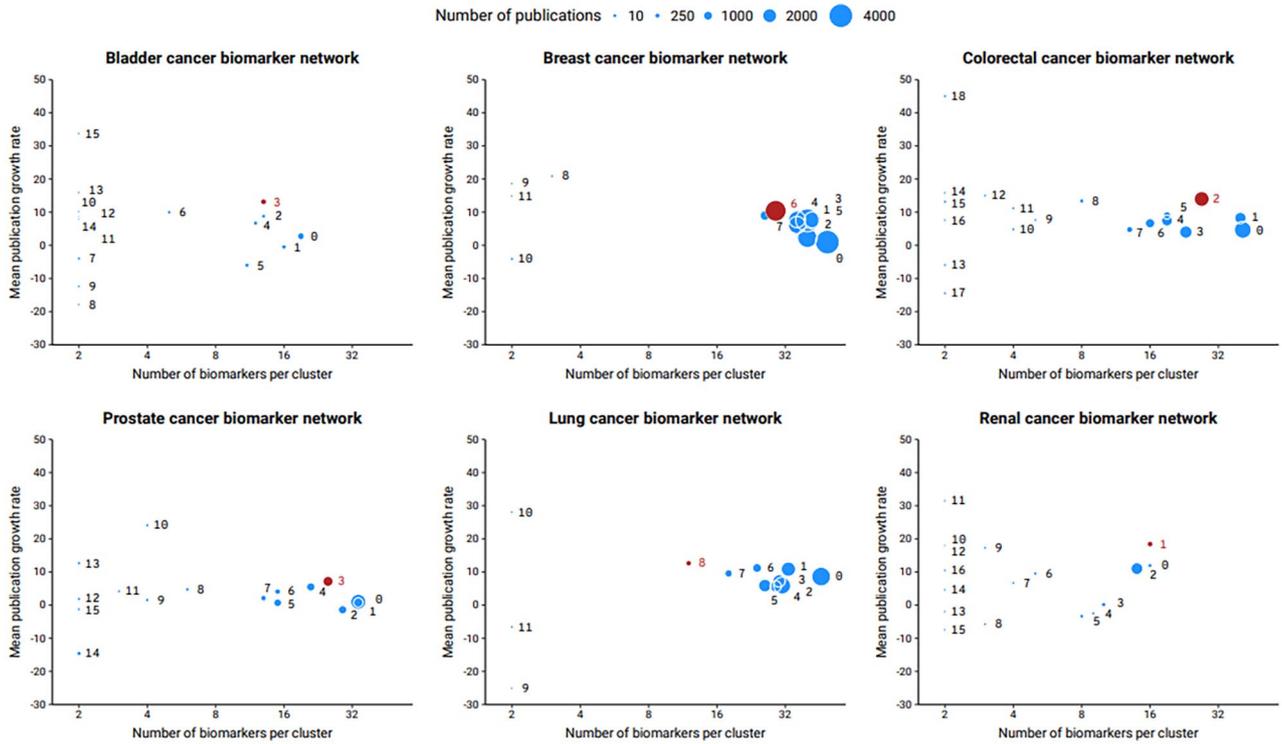
**Table 2.** Number of publications identified for each cancer site.

CANCER SITE	NUMBER OF PUBLICATIONS
Breast	108 134
Lung	88 874
Colorectal	69 284
Prostate	60 644
Renal	13 727
Bladder	13 591
Total	255 942

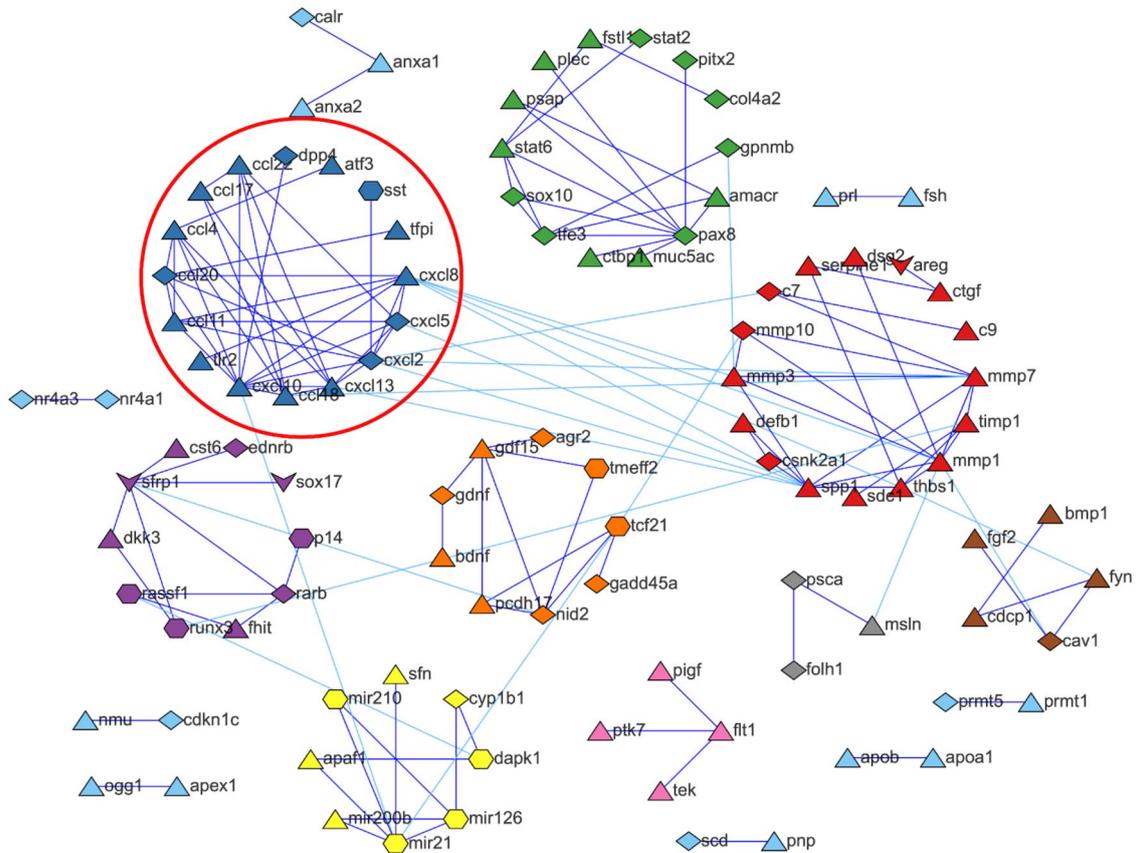
Many of these publications were relevant to more than 1 cancer site.



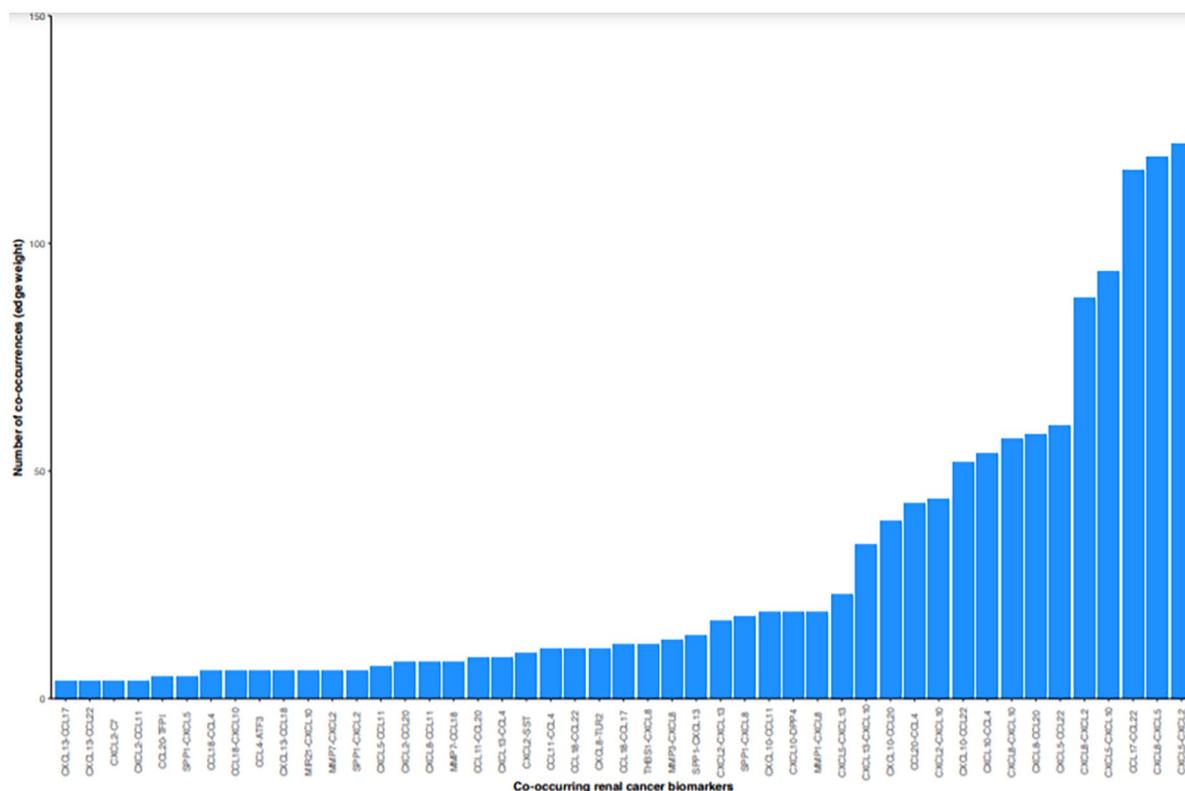
**Figure 1.** Analysis workflow and results.



**Figure 2.** Publication growth rate by cluster for each cancer site. The clusters with the highest publication growth rate and at least 10 biomarkers are highlighted red.



**Figure 3.** Renal cancer biomarker network. Cluster 1 (circled) had the highest publication growth rate among clusters with at least 10 biomarkers. Node colour represents cluster membership. Node shape represents biomarker type. Diamond, gene; triangle, protein; hexagon, genomic; chevron, proteomic.



**Figure 4.** Number of biomarker co-occurrences in renal cancer cluster one. CXCL5-CXCL2 had the most co-occurrences. Abbreviation: CXCL, C-X-C motif chemokine ligand.

### Biological Context of Biomarker Mentions Within the Colorectal Biomarker Network

The cluster with the second highest publication growth rate was colorectal cancer cluster 2 (Figure 5). This cluster contained 139 edges in total, of which 89 were within the cluster. The most common pair by co-occurrence was protein arginine *N*-methyltransferase (PRMT)5-PRMT1, with 361 co-mentions in 47 unique publications in which the pairing was mentioned more than once (Figure 6). There were 700 publications associated with the within-cluster edges.

To identify a subset of publications for analysis (instead of identifying the biomarker pair with the most co-occurrences, as done previously), we took the top 20 publications based on Mendeley captures for the entire cluster. Of these 20 publications, 15 were narrative reviews and 5 were preclinical research papers. These 20 publications contained 90 (51 unique) co-mentioned biomarker pairs, of which 21 (20 unique) were mentioned in the context of colorectal cancer, 60 were not specific to a cancer site, and 9 (7 unique) were incorrectly identified as being associated with colorectal cancer (Supplemental Table 3). Of the 90 biomarker pairs, there was a direct mechanistic link between 67 of them. Twenty-three of the 51 unique biomarker pairs were already annotated in known biological pathways databases (Supplemental Table 4). The most common biomarker pair was C-C motif chemokine ligand (CCL)17-CCL22, appearing in 9 publications.

Biomarker pairs in this colorectal cluster were mostly chemokines (50/51 unique pairs) and, when mapped to the

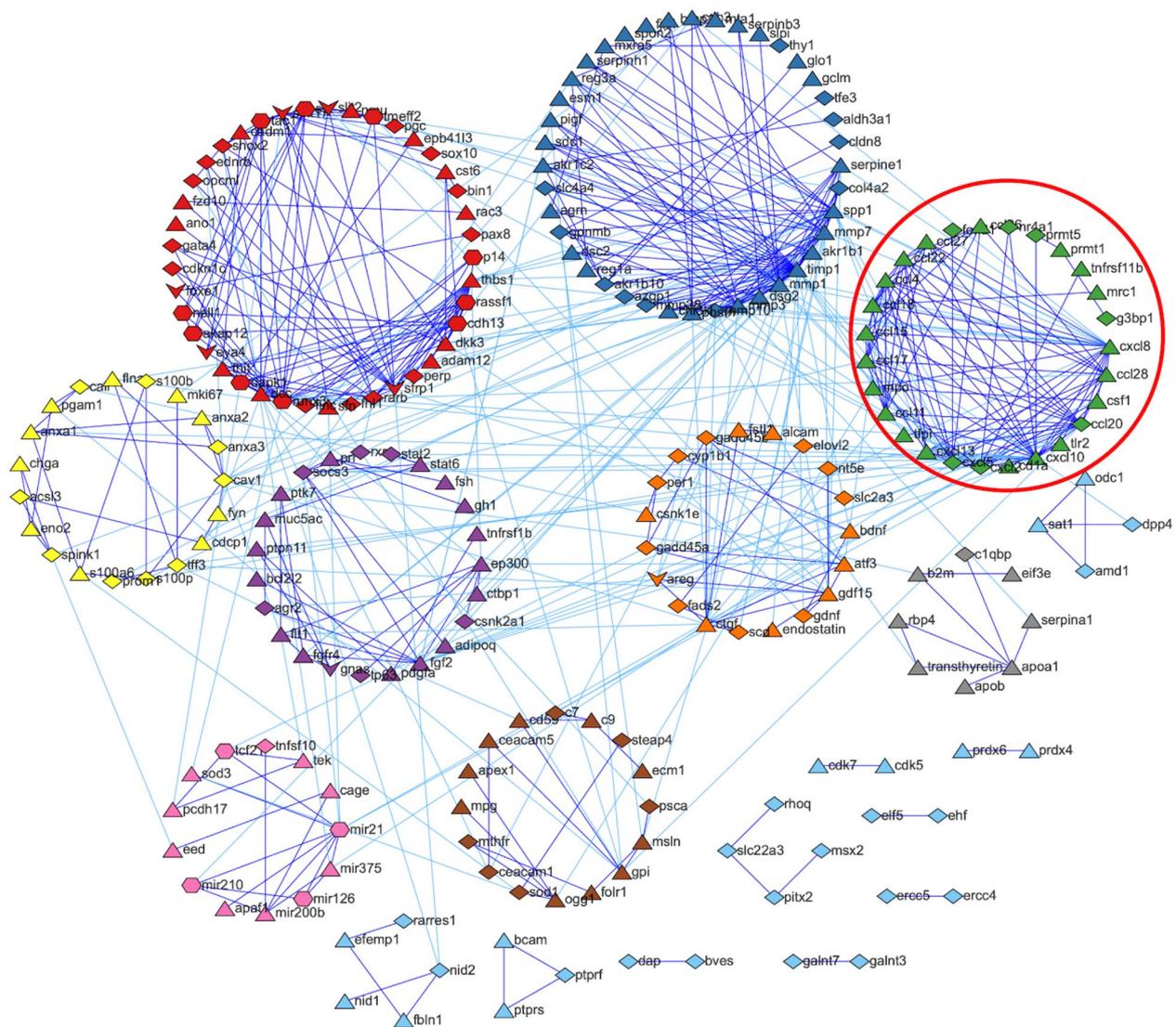
NCIt ontology, were shown to be associated with processes such as cellular infiltration and chemotaxis and to have a notable emphasis on chemokines that characterize M1 and M2 macrophages (Supplemental Table 5).

### Biological Context Within the Cancer-Agnostic Network

The cancer-agnostic network contained 12 clusters comprising 335 nodes with 1265 edges (Figure 7).

The cluster with the highest publication growth rate and at least 10 biomarkers was cluster 8, with 418 publications associated with its nodes (Figure 8). This cluster contained 26 edges in total, of which 11 were within the cluster. Five of the 11 biomarker pairs were already annotated in known biological pathways databases (Supplemental Table 6). There were 55 publications associated with the within-cluster edges so, to identify a subset of publications for analysis, we took the top 20 publications based on Mendeley captures for the entire cluster (Supplemental Table 7). The most commonly occurring biomarker pair was stearoyl-coenzyme A desaturase (SCD)-fatty acid desaturase 2 (FADS2) with 143 co-mentions (Figure 9).

Of these 20 publications, 11 were narrative reviews, 7 were preclinical research papers, 1 was a systematic review and meta-analysis and 1 was a booklet of congress poster abstracts. These 20 publications contained 29 (8 unique) co-mentioned biomarker pairs of which 2 (both unique), 12 (6 unique), 6 (5 unique), seven (4 unique), 14 (5 unique) and 0 were mentioned in the context of bladder, breast, colorectal, lung,



**Figure 5.** Colorectal cancer biomarker network. Cluster 2 (circled) had the highest publication growth rate and contained at least 10 biomarkers that were chosen for further study. Node colour represents cluster membership. Node shape represents biomarker type. Diamond, gene; triangle, protein; hexagon, genomic; chevron, proteomic.

prostate and renal cancer, respectively. Twenty-five (25/29) of the biomarker pairs were correctly identified as being associated with the 6 cancer sites included in this study. Five of the unique biomarker pairs (*scd-fabp5*, *sat1-odc1*, *sat1-amd1*, *fads2-evolv12*, *odc1-amd1*) were already annotated in known biological pathways databases.

Biomarker pairs in this cancer-agnostic network cluster were mostly related to biogenic amine metabolism (14/29) and fatty acid metabolism (13/29); 1 pair (1/29) was related to suicide gene therapy and 1 pair (1/29) was not relevant because the co-mention was incorrectly identified in a congress poster abstract booklet (Supplemental Table 8).

## Gene Set Enrichment Analysis

### Renal cancer cluster 1

The top 10 enriched KEGG pathways terms showed that many of the biomarkers are known to be involved in cytokine and chemokine signalling pathways, including interleukin

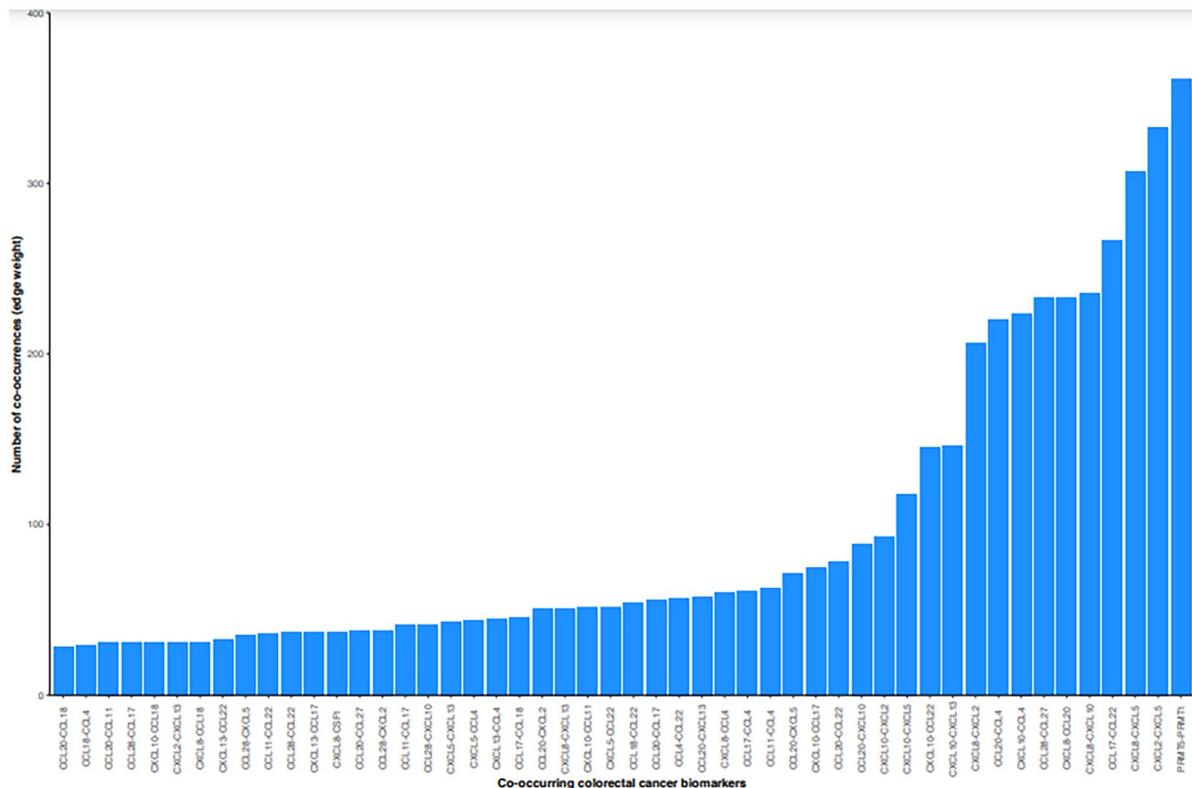
(IL)-17, tumour necrosis factor (TNF), toll-like receptor (TLR) and nuclear factor (NF)-kappa B signalling pathways (Table 3). GO biological process term enrichment highlighted the role of the biomarkers in chemotaxis, and cellular response to interferon gamma and IL-1 (Table 4).

### Colorectal cancer cluster 2

Analysis of biomarkers in colorectal cancer cluster 2 showed that, although not all biomarkers overlapped, the same KEGG pathways were enriched as for renal cancer cluster 1 (Table 5). Similarly, GO pathways were similar, although a response to interferon-gamma was not identified for the colorectal cancer biomarker set (Table 6).

### Cancer agnostic cluster 8

For the cancer-agnostic network cluster 8, KEGG pathway enrichment showed that few pathways were associated with



**Figure 6.** Number of biomarker co-occurrences in colorectal cancer cluster 2 (top 50 biomarker pairs). PRMT5-PRMT1 had the most co-occurrences. Abbreviation: PRMT, protein arginine *N*-methyltransferase.

multiple biomarkers; however, the involvement of several biomarkers in both the biosynthesis of fatty acids and peroxisome proliferator-activated receptors (PPAR) nuclear hormone receptors, which are activated by fatty acids and a potential role for ferroptosis were highlighted (Table 7). Few GO biological pathways were also associated with multiple biomarkers but polyamine and fatty acid biosynthesis and metabolism were notably enriched (Table 8).

## Discussion

In this study, we developed a novel full-text literature search and network analytics methodology to identify cancer biomarker relationships of emerging scientific interest; however, this approach is not limited to oncology. The tool presents emerging biomarkers in relational context to other biomarkers and oncology sites of interest and enables users to identify rapidly publications describing these biomarker relationships. It is freely accessible at <https://reports.dimensions.ai/mined-oncology-biomarkers/>

The initial corpus of literature from which the network was built was identified by selecting publications in which biomarker terms occurred in proximity to specific cancer terms.

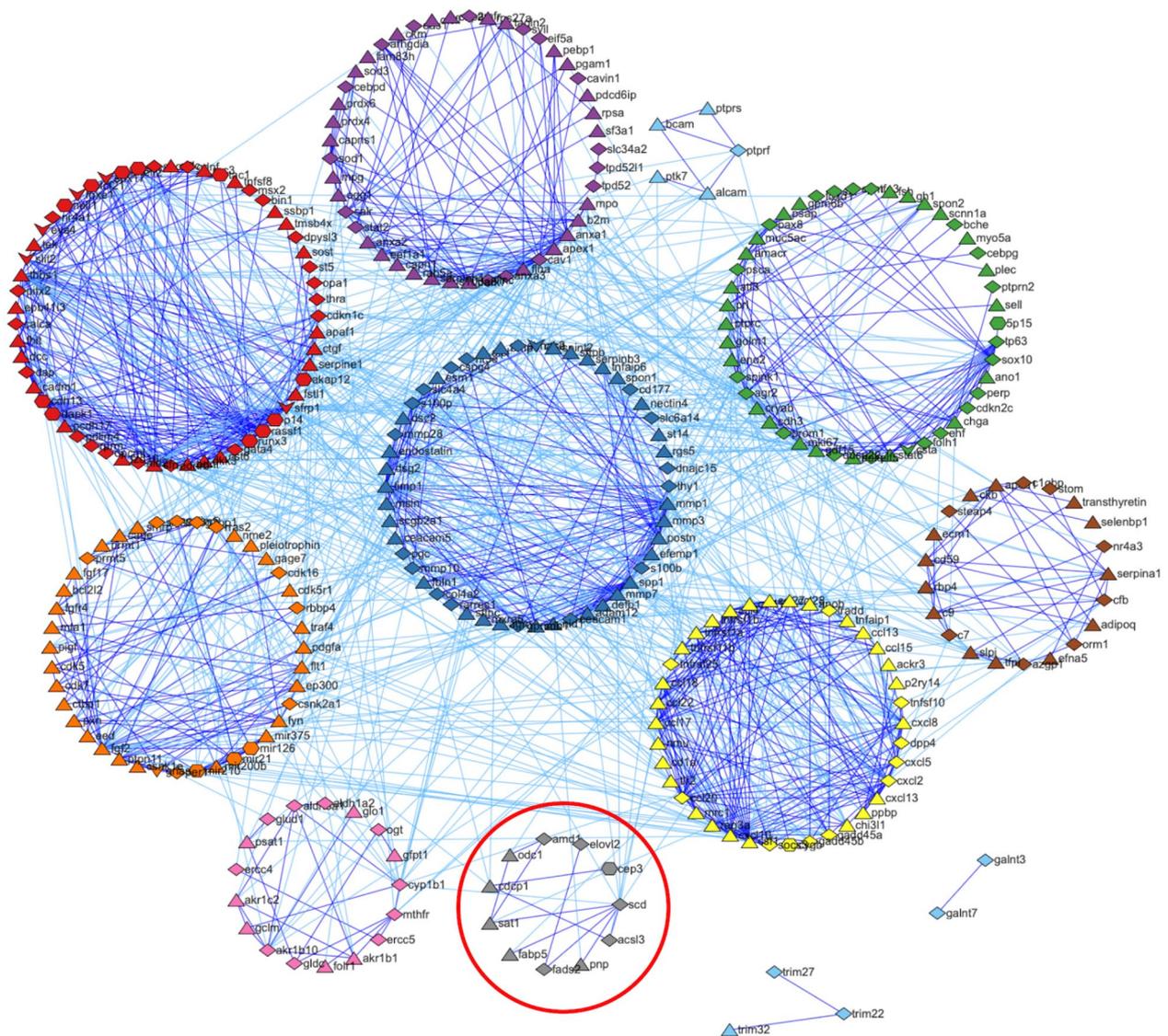
To enrich the contextual information on these biomarkers, the corpus of publications was text-mined to identify biomarkers that co-occurred, on the expectation that these paired biomarkers would be likely to share biological context. To sharpen further the focus on related biomarkers of emerging interest,

we focussed our manual validation on biomarkers and networks with higher publication velocity (ie, an increasing volume of literature attention over our time period of interest).

To test if the biomarker pairings were biologically meaningful, we focussed on 3 different approaches. For each, we identified the fastest-growing clusters because we were interested in the fields of interest of related biomarkers.

The textual analysis confirmed that the text-mining strategy was mostly successful in identifying networks and pairs of related biomarkers. In the renal cancer biomarker cluster selected for review, the CXCL2 and CXCL5 pair occurred most commonly. Not only do they both signal through the same receptor, C-X-C motif chemokine receptor (CXCR)2, but they are differentially expressed in multiple cancer sites, including renal cancer.<sup>20,21</sup> This direct and mechanistic link between the biomarkers was described in each of the 34 publications (although not always in the context of renal cancer) and is annotated in the HumanNet, BioGRID and Reactome databases.

The KEGG pathway enrichment of the selected renal cancer biomarker cluster revealed that the identified biomarkers are largely involved in cytokine and chemokine signalling, in particular the IL-1, IL-17, TNF, TLR and NF-kappa B pathways. Thus, our method identified biomarkers linked to 2 important, known renal cancer pathways and 3 pathways that are less understood but of emerging interest.

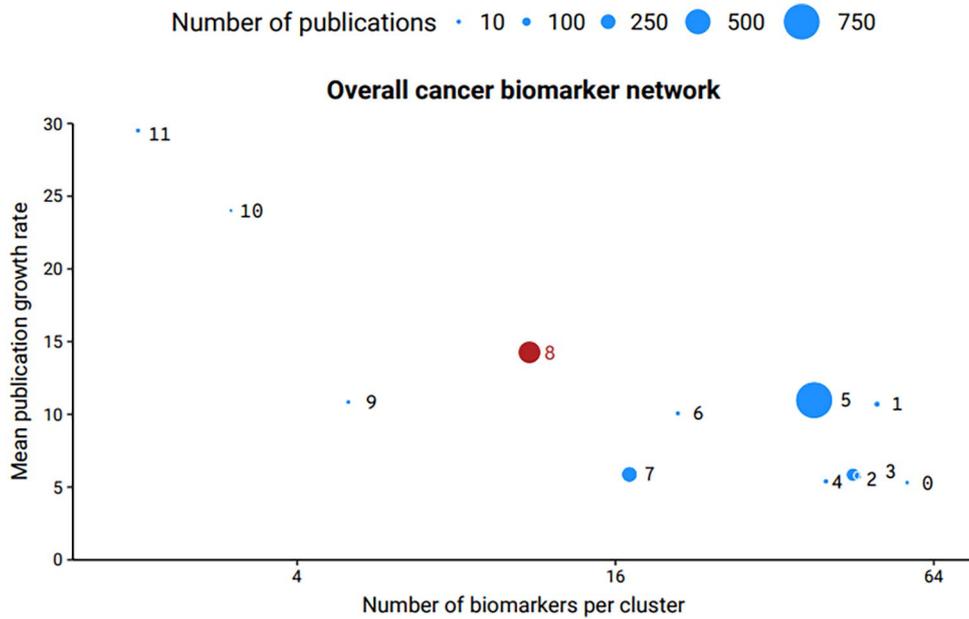


**Figure 7.** Cancer-agnostic biomarker network. Cluster 8 (circled) had the highest publication growth rate and contained at least 10 biomarkers that were chosen for further study. Node colour represents cluster membership. Node shape represents biomarker type. Diamond, gene; triangle, protein; hexagon, genomic; chevron, proteomic).

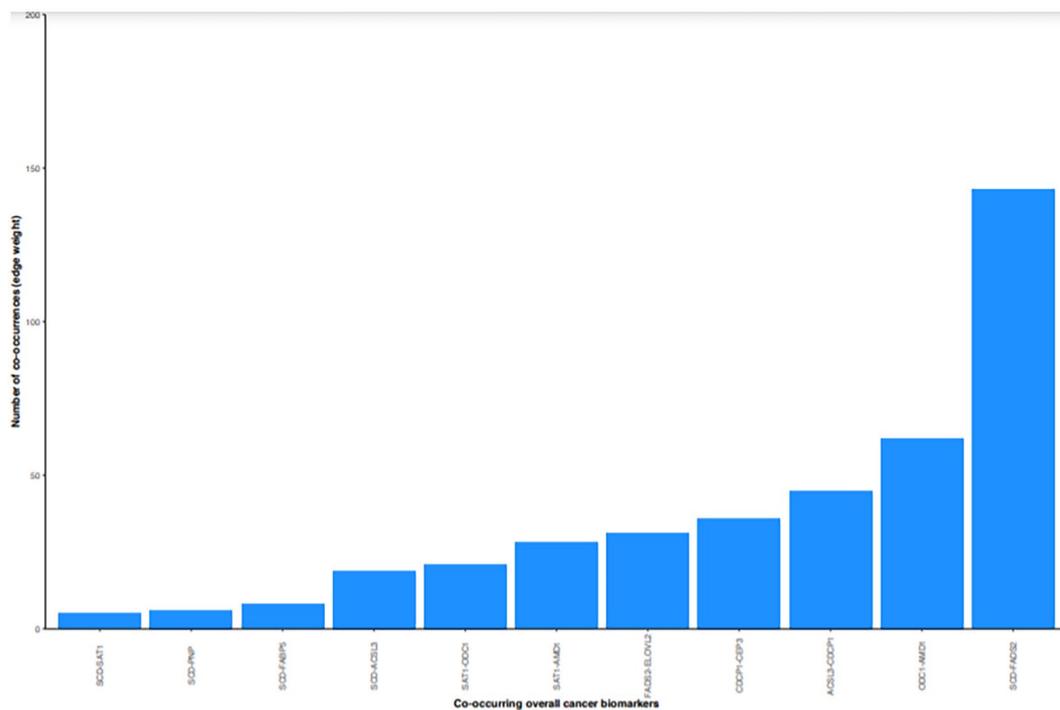
IL-1 is a pro-inflammatory cytokine associated with tumour invasiveness and metastasis that suppresses anti-tumour immunity through proliferation of polymorphonuclear myeloid-derived suppressive cells (PMN-MDSCs).<sup>22</sup> Moreover, IL-1 expression is induced by immunotherapy.<sup>22</sup> It is proposed that IL-1 blockade may be a suitable monotherapy or as a combination therapy with other immunotherapies.<sup>22,23</sup> Similarly, the IL-17 axis could be an attractive target for immunotherapy,<sup>24</sup> which demonstrates the potential utility of our technique. Emerging evidence associates IL-17 with tumour growth during early oncogenesis in multiple cancer types. Indicative of the pleiotropy of many cytokines, IL-17 expression may also be protective, relating to cancer cell apoptosis and antitumoural immune cell activation.<sup>24</sup>

Pathways requiring deeper understanding are TNF, TLRs and NF-kappa B. The role of TNF in cancer has been controversial, however it has been shown to inhibit anti-tumour immune response and to alter the phenotype of cancer cells,

making them less visible to T cells and to express immune inhibitory molecules: further research is underway.<sup>25</sup> Conversely, in renal cancer, TNF may be pro-tumorigenic and could be a target for immunotherapy.<sup>26</sup> TLRs activate several downstream pathways, and their involvement in cancer has resulted in the investigation of both TLR agonists and antagonists; however, understanding how these molecules might be incorporated into cancer treatment protocols is not fully understood.<sup>27</sup> NF-kappa B inhibition has been explored with little success, nevertheless, increased understanding of the NF-kappa B pathway has instigated renewed interest in the potential of NF-kappa B inhibitors in some cancers, including renal cancer.<sup>28</sup> Furthermore, demonstrating the importance of context in immunoregulation, upregulation of NF-kappa B is proposed as a potential mediator of the anti-tumour properties of current immunotherapies such as checkpoint inhibitors and chimeric-antigen receptor T cells (CAR-T)-cell-based therapies, and other therapies like TLR agonists.<sup>28</sup>



**Figure 8.** Publication growth rate by cluster for the cancer-agnostic network. Cluster 8 had the highest growth rate and contained at least 10 biomarkers.



**Figure 9.** Number of biomarker co-occurrences in cancer-agnostic network cluster 8. SCD-FADS2 had the most co-occurrences. Abbreviations: FACS2, fatty acid desaturase 2; SCD, stearyl-coenzyme A desaturase.

In the colorectal cancer biomarker cluster selected for review, the most common co-occurrence was that of PRMT5 with PRMT1, both of which have been associated with premature cellular ageing and cellular senescence.<sup>29</sup> PRMT1 methylates the epidermal growth factor receptor (EGFR), and PRMT1-mediated increased methylation, as well as the consequent overactivation of EGFR signalling, leads to sustained cell proliferation.<sup>29</sup> Methylation-defective EGFR reduced colorectal tumour growth in mice.<sup>29</sup> Importantly,

after treatment with the therapeutic EGFR monoclonal antibody cetuximab, EGFR methylation levels correlated with higher cancer reapppearance rates and reduced survival.<sup>29</sup> PRMTs are therefore attractive cancer targets for small molecule inhibition.

The majority of the remaining biomarker pairs in the colorectal cluster and all those that were chosen on the basis of Mendeley saves for validation were chemokine pairings and were shown to be associated with processes such as cellular

**Table 3.** Gene set enrichment for KEGG pathways, renal cancer cluster 1.

TERM	CLUSTER GENES	P VALUE
Viral protein interaction with cytokine and cytokine receptor	CXCL10;CCL11;CCL22;CXCL8;CCL20;CCL4;CXCL13;CCL18;CXCL2;CCL17;CXCL5	$1.18 \times 10^{-22}$
Chemokine signalling pathway	CXCL10;CCL11;CCL22;CXCL8;CCL20;CCL4;CXCL13;CCL18;CXCL2;CCL17;CXCL5	$2.00 \times 10^{-19}$
Cytokine-cytokine receptor interaction	CXCL10;CCL11;CCL22;CXCL8;CCL20;CCL4;CXCL13;CCL18;CXCL2;CCL17;CXCL5	$2.44 \times 10^{-17}$
IL-17 signalling pathway	CXCL10;CCL11;CXCL8;CCL20;CXCL2;CCL17;CXCL5	$4.46 \times 10^{-13}$
Rheumatoid arthritis	CXCL8;CCL20;CXCL2;CXCL5;TLR2	$8.18 \times 10^{-9}$
TNF signalling pathway	ATF2;CXCL10;CCL20;CXCL2;CXCL5	$2.09 \times 10^{-8}$
TLR signalling pathway	CXCL10;CXCL8;CCL4;TLR2	$1.20 \times 10^{-6}$
Legionellosis	CXCL8;CXCL2;TLR2	$1.20 \times 10^{-5}$
Amoebiasis	CXCL8;CXCL2;TLR2	$6.87 \times 10^{-5}$
NF-kappa B signalling pathway	CXCL8;CCL4;CXCL2	$7.28 \times 10^{-5}$

Abbreviations: ATF, activating transcription factor; CCL, C-C motif chemokine ligand; CXCL, C-X-C motif chemokine ligand; KEGG, Kyoto Encyclopedia of Genes and Genomes; TLR, toll-like receptor.

**Table 4.** Gene set enrichment for GO biological processes, renal cancer cluster 1.

TERM	CLUSTER GENES	P VALUE
Chemokine-mediated signalling pathway (GO:0070098)	CXCL10;CCL11;CCL22;CXCL8;CCL20;CCL4;CXCL13;CCL18;CXCL2;CCL17;CXCL5	$1.26 \times 10^{-25}$
Cellular response to chemokine (GO:1990869)	CXCL10;CCL11;CCL22;CXCL8;CCL20;CCL4;CXCL13;CCL18;CXCL2;CCL17;CXCL5	$2.89 \times 10^{-25}$
Neutrophil chemotaxis (GO:0030593)	CXCL10;CCL11;CCL22;CXCL8;CCL20;CCL4;CXCL13;CCL18;CXCL2;CCL17;CXCL5	$1.82 \times 10^{-24}$
Granulocyte chemotaxis (GO:0071621)	CXCL10;CCL11;CCL22;CXCL8;CCL20;CCL4;CXCL13;CCL18;CXCL2;CCL17;CXCL5	$2.99 \times 10^{-24}$
Neutrophil migration (GO:1990266)	CXCL10;CCL11;CCL22;CXCL8;CCL20;CCL4;CXCL13;CCL18;CXCL2;CCL17;CXCL5	$5.61 \times 10^{-24}$
Inflammatory response (GO:0006954)	CXCL10;CCL11;CCL22;CXCL8;CCL20;CCL4;CXCL13;CCL18;CXCL2;CCL17;CXCL5;TLR2	$7.00 \times 10^{-21}$
Lymphocyte chemotaxis (GO:0048247)	CXCL10;CCL11;CCL22;CCL20;CCL4;CXCL13;CCL18;CCL17	$3.55 \times 10^{-18}$
Cytokine-mediated signalling pathway (GO:0019221)	CXCL10;CCL11;CCL22;CXCL8;CCL20;CCL4;CXCL13;CCL18;CXCL2;CCL17;CXCL5	$8.99 \times 10^{-14}$
Response to interferon-gamma (GO:0034341)	CCL11;CCL22;CCL20;CCL4;CCL18;CCL17;TLR2	$1.39 \times 10^{-13}$
Response to IL-1 (GO:0070555)	CCL11;CCL22;CXCL8;CCL20;CCL4;CCL18;CCL17	$2.35 \times 10^{-13}$

Abbreviations: CCL, C-C motif chemokine ligand; CXCL, C-X-C motif chemokine ligand; GO, Gene Ontology; TLR, toll-like receptor.

infiltration and chemotaxis and to have a notable emphasis on chemokines that characterize M1 and M2 macrophages. Of further note was the pairing of colony stimulating factor 1 (CSF1) with CXCL8; CSF1 receptor (CSF1R) inhibition alters chemokine secretion by cancer-associated fibroblasts,

thereby attracting pro-tumour, (PMN-MDSCs)<sup>30</sup> Combined inhibition of CSF1R and CXCR2 (the receptor for CXCL8) blocks MDSC recruitment and reduces tumour growth, which is further improved by the addition of anti-programmed cell death protein 1 (PD-1) drugs.<sup>30</sup> The most common biomarker

**Table 5.** Gene set enrichment for KEGG pathways, colorectal cancer cluster 2.

TERM	CLUSTER GENES	P VALUE
Viral protein interaction with cytokine and cytokine receptor	CCL11;CXCL8;CCL22;CSF1;CCL20;CXCL13;CXCL2;CXCL5;CXCL10;CCL4;CCL18;CCL17;CCL28;CCL27;CCL15;CCL26	$5.38 \times 10^{-31}$
Cytokine-cytokine receptor interaction	CCL11;CXCL8;CCL22;CSF1;CCL20;TNFRSF11B;CXCL13;CXCL2;CXCL5;CXCL10;CCL4;CCL18;CCL17;CCL28;CCL27;CCL15;CCL26	$3.44 \times 10^{-25}$
Chemokine signalling pathway	CCL11;CXCL8;CCL22;CCL20;CXCL13;CXCL2;CXCL5;CXCL10;CCL4;CCL18;CCL17;CCL28;CCL27;CCL15;CCL26	$4.88 \times 10^{-24}$
IL-17 signalling pathway	CXCL10;CCL11;CXCL8;CCL20;CXCL2;CCL17;CXCL5	$3.32 \times 10^{-11}$
Rheumatoid arthritis	CXCL8;CSF1;CCL20;CXCL2;CXCL5;TLR2	$2.35 \times 10^{-9}$
TNF signalling pathway	CXCL10;CSF1;CCL20;CXCL2;CXCL5	$3.68 \times 10^{-7}$
Amoebiasis	CXCL8;CXCL2;CD1A;TLR2	$1.02 \times 10^{-5}$
TLR receptor signalling pathway	CXCL10;CXCL8;CCL4;TLR2	$1.10 \times 10^{-5}$
Legionellosis	CXCL8;CXCL2;TLR2	$6.12 \times 10^{-5}$
NF-kappa B signalling pathway	CXCL8;CCL4;CXCL2	$3.65 \times 10^{-4}$

Abbreviations: CCL, C-C motif chemokine ligand; CD, cluster of differentiation; CSF, colony stimulating factor; CXCL, C-X-C motif chemokine ligand; KEGG, Kyoto Encyclopedia of Genes and Genomes; TLR, toll-like receptor; TNFRSF, tumor necrosis factor receptor superfamily member

**Table 6.** Gene set enrichment for GO biological processes, colorectal cancer cluster 2.

TERM	CLUSTER GENES	P VALUE
Chemokine-mediated signalling pathway (GO:0070098)	CCL11;CXCL8;CCL22;CCL20;CXCL13;CXCL2;CXCL5;CXCL10;CCL4;CCL18;CCL17;CCL15;CCL26	$2.81 \times 10^{-27}$
Cellular response to chemokine (GO:1990869)	CCL11;CXCL8;CCL22;CCL20;CXCL13;CXCL2;CXCL5;CXCL10;CCL4;CCL18;CCL17;CCL15;CCL26	$7.66 \times 10^{-27}$
Neutrophil chemotaxis (GO:0030593)	CCL11;CXCL8;CCL22;CCL20;CXCL13;CXCL2;CXCL5;CXCL10;CCL4;CCL18;CCL17;CCL15;CCL26	$6.99 \times 10^{-26}$
Granulocyte chemotaxis (GO:0071621)	CCL11;CXCL8;CCL22;CCL20;CXCL13;CXCL2;CXCL5;CXCL10;CCL4;CCL18;CCL17;CCL15;CCL26	$1.27 \times 10^{-25}$
Neutrophil migration (GO:1990266)	CCL11;CXCL8;CCL22;CCL20;CXCL13;CXCL2;CXCL5;CXCL10;CCL4;CCL18;CCL17;CCL15;CCL26	$2.69 \times 10^{-25}$
Lymphocyte chemotaxis (GO:0048247)	CXCL10;CCL11;CCL22;CCL20;CCL4;CXCL13;CCL18;CCL17;CCL15;CCL26	$7.23 \times 10^{-21}$
Inflammatory response (GO:0006954)	CCL11;CXCL8;CCL22;CCL20;CXCL13;CXCL2;CXCL5;CXCL10;CCL4;CCL18;CCL17;CCL15;CCL26;TLR2	$8.34 \times 10^{-21}$
Cytokine-mediated signalling pathway (GO:0019221)	CCL11;CXCL8;CCL22;CSF1;CCL20;TNFRSF11B;CXCL13;CXCL2;CXCL5;CXCL10;CCL4;CCL18;CCL17;CCL15;CCL26	$2.51 \times 10^{-16}$
Lymphocyte migration (GO:0072676)	CCL11;CCL22;CCL20;CCL4;CCL18;CCL17;CCL15;CCL26	$2.62 \times 10^{-16}$
Monocyte chemotaxis (GO:0002548)	CCL11;CCL22;CCL20;CCL4;CCL18;CCL17;CCL15;CCL26	$4.01 \times 10^{-16}$

Abbreviations: CCL, C-C motif chemokine ligand; CXCL, C-X-C motif chemokine ligand; GO, Gene Ontology; TLR, toll-like receptor; TNFRSF, tumor necrosis factor ligand superfamily member.

**Table 7.** Gene set enrichment for KEGG pathways, cancer-agnostic network, cluster 8.

TERM	CLUSTER GENES	P VALUE
PPAR signalling pathway	FADS2;FABP5;SCD;ACSL3	$5.59 \times 10^{-8}$
Biosynthesis of unsaturated fatty acids	FADS2;SCD;ELOVL2	$3.59 \times 10^{-7}$
Arginine and proline metabolism	AMD1;ODC1;SAT1	$2.39 \times 10^{-6}$
Ferroptosis	ACSL3;SAT1	$2.23 \times 10^{-4}$
Fatty acid biosynthesis	ACSL3	$9.86 \times 10^{-3}$
Alpha-Linolenic acid metabolism	FADS2	$1.37 \times 10^{-2}$
Fatty acid elongation	ELOVL2	$1.48 \times 10^{-2}$
Nicotinate and nicotinamide metabolism	PNP	$1.91 \times 10^{-2}$
Fatty acid degradation	ACSL3	$2.34 \times 10^{-2}$
Cysteine and methionine metabolism	AMD1	$2.72 \times 10^{-2}$

Abbreviations: ACSL, acyl-CoA synthetase long chain family member ; AMD, adenosylmethionine decarboxylase ; EVOVL, elongation of very-long-chain fatty acids-like 2; FABP, fatty acid binding protein; FADS2, fatty acid desaturase, KEGG; Kyoto Encyclopedia of Genes and Genomes; ODC1, ornithine decarboxylase ; PNP, purine nucleoside phosphorylase; SAT, spermidine/spermine N1-acetyltransferase; SCD, stearoyl-CoA desaturase.

**Table 8.** Gene set enrichment for GO biological processes, cancer-agnostic network, cluster 8.

TERM	CLUSTER GENES	P VALUE
Polyamine metabolic process (GO:0006595)	AMD1;ODC1;SAT1	$3.53 \times 10^{-8}$
Fatty-acyl-coa biosynthetic process (GO:0046949)	SCD;ELOVL2;ACSL3	$4.98 \times 10^{-7}$
Unsaturated fatty acid metabolic process (GO:0033559)	FADS2;SCD;ELOVL2	$3.02 \times 10^{-6}$
Polyamine biosynthetic process (GO:0006596)	AMD1;SAT1	$5.77 \times 10^{-6}$
Spermidine metabolic process (GO:0008216)	AMD1;SAT1	$7.69 \times 10^{-6}$
Long-chain fatty acid metabolic process (GO:0001676)	FADS2;ELOVL2;ACSL3	$1.11 \times 10^{-5}$
Alpha-linolenic acid metabolic process (GO:0036109)	FADS2;ELOVL2	$2.14 \times 10^{-5}$
Cellular biogenic amine metabolic process (GO:0006576)	AMD1;ODC1	$2.14 \times 10^{-5}$
Long-chain fatty-acyl-coa biosynthetic process (GO:0035338)	ELOVL2;ACSL3	$4.19 \times 10^{-5}$
Linoleic acid metabolic process (GO:0043651)	FADS2;ELOVL2	$5.74 \times 10^{-5}$

Abbreviations: ACSL, acyl-CoA synthetase long chain family member ; AMD1, adenosylmethionine decarboxylase 1; EVOVL, elongation of very-long-chain fatty acids-like; FABP, fatty acid binding protein ; FADS, fatty acid desaturase , KEGG; Kyoto Encyclopedia of Genes and Genomes; ODC1, ornithine decarboxylase 1; SAT, spermidine/spermine N1-acetyltransferase ; SCD, stearoyl-CoA desaturase.

pair was CCL17-CCL22, appearing in 9 publications. This pair is known to the HIPPIE database, confirming that our strategy can identify functional biomarker relationships. It is interesting to note, and a strength of our approach, that we identified biomarker pairs that are functionally related but not currently annotated in interaction databases. For example, CXCL8 and CCL15 were identified by our approach and both have a role in recruitment of monocytes, neutrophil, and myeloid-derived suppressor cells to the tumour site.

Similarly, we identified CCL11 and CCL15, both of which interact with CCR3 but are not present in known interaction networks.

KEGG pathway enrichment of the selected colorectal cancer biomarker cluster identified the same pathways as for renal cancer. Indeed, IL-17, TNF, TLR and NF kappa B pathways are all associated with colorectal cancer, with IL-1 being highlighted in a recent systematic review as a high interest candidate for treatment of patients with colorectal cancer.<sup>31-35</sup>

For the cancer-agnostic network cluster that was selected for further study, 86% (25/29) of the publications we validated were correctly identified as being associated with the 6 cancer sites included in this study. Interestingly, 14 of the biomarker pairings in these publications were related to fatty acid metabolism, 15 to biogenic amine metabolism, and 1 to suicide gene therapy. The most co-mentioned biomarker pair was SCD-FADS2, with 143 co-mentions. Fatty acid metabolism is altered in cancer: fatty acids can mediate cancer progression and metastasis, and cancer cells obtain fatty acids from *de novo* synthesis and exogenous uptake. Sapeinate is generated from palmitate desaturation by FADS2, and monounsaturated fatty acids can be generated from palmitate by SCD.<sup>36</sup> Importantly, FADS2 appears to compensate for SCD, so, although SCD inhibitors are becoming available, the compensatory activity of FADS2 may be important to consider therapeutically.<sup>36</sup>

KEGG pathway enrichment of the selected cancer-agnostic biomarker cluster identified fatty acid biosynthesis, PPAR signalling and ferroptosis as pathways of interest, each of which could provide a novel strategy for cancer therapy. Fatty acids are not merely components of the cell membrane but are secondary messengers and sources of energy production, and could play a role in oncogenic signalling.<sup>36</sup> PPAR receptors are ligand-activated transcription factors that have a role in the modulation of inflammation, cell proliferation and differentiation, known to impact several cancer types.<sup>37</sup> Ferroptosis, an iron-dependent type of cell death triggered by extra-mitochondrial lipid peroxidation that has been observed in multiple cancer types, has a pivotal role in cancer cell destruction.<sup>38</sup>

Across our 3 example analyses, 40/74 papers were narrative reviews of the preclinical literature and 25/74 were preclinical studies. Only 1 clinical trial was identified, and it was at phase 1. This supports the notion that, by filtering out biomarkers that are already well known or with very little research volume and by using publication velocity as a metric, we successfully identified biomarker pairs that may be clinically important in the future.

Of note is the fact that, across the example clusters we analysed, 48/104 biomarker pairs are not annotated in the HumanNet, HIPPIE or Reactome interaction databases. This is important because it highlights the ability of text-mining approaches to identify potential relationships between entities that may not have been demonstrated in the laboratory or through computational prediction models based on protein sequence or structural data. Researchers adopting similar approaches could, as in the above example for the functional relationship between SCD-FADS2, use those biomarker pairs not in interaction databases to generate novel hypotheses. Relationships between biomarkers based on term co-occurrence mechanistically linked to cancer were shown in 126/153 cases, showing that noise is minimal, and that text-mining can be a useful adjunctive approach to the identification of meaningful, biologically relevant relationships.

Perhaps the main limitation of this approach is that it is difficult to summarize the optimal parameters for term co-occurrence. Potential solutions to this are to use multiple word proximity distances or, prior to proximity detection, separation of the text into semantic analysis units before processing; a suitable context window may be sentences. Decomposition of the corpus into sentences may reduce noise, in that terms co-occurring in the same sentence are highly likely to be related. However, this reduces sensitivity and so paragraphs may prove superior contextual units. Another approach could be to extract co-occurrence statistics not only from full publications, paragraphs or even sentences but over the entire corpus and then calculate the 'importance' of the co-occurring terms in relation to the corpus, similarly to term frequency inverse document frequency statistics (TF-IDF). It may also be useful in future analyses to differentiate between co-mentions in the introduction, results or discussion sections (for non-narrative publications).

A further limitation is that our method does not identify the types of associations, for example physical protein-protein, transcription factor-protein or pathway interactions, the molecular nature of the associations (protein or mRNA expression level, somatic mutation or copy number variation), nor does it identify negations. However, it is likely sufficient to represent a biological relationship without distinguishing the analyte. To enable identification of association types, a context aware system would need to be developed. At present, the most powerful framework for developing such capability would be fine tuning a deep learning language model for a Name Entity Recognition task, in which the entity types correspond to the desired nature of associations, an important aspect of signal transduction pathways; pattern-based approaches could be developed to infer this. The heuristic approach we have described, while perhaps not optimal, is practical and may allow analysis to proceed more quickly than machine learning-based approaches.

Further work could look to identify those biomarker pairs that appear in the preclinical literature and then at a later time point, to see if these same pairs emerge in the clinical literature, thus validating the approach as useful in the identification of 'up and coming' biomarkers. Similarly, pairs could be analysed in non-review papers and at a later time window to see if the pairs reach review publications. Finally, this approach could be tested retrospectively by analysing publications up until a designated time point and then, at a later date, investigate if identified molecules later became validated biomarkers.

## Conclusion

Our approach, which enables us to find publications based on biomarker relationships, identified biomarker relationships not known to existing interaction networks. This search method finds relevant literature that could be missed with keyword searches, even if full text is available. It enables users to focus on emergent research, extract relevant biological information and

may provide new biological insights that could not be achieved by individual review of papers.

### Authors Contributions

Methodology: BS, OP, DC, SH, TR and KW. Data collection: BS, OP and KW. Analysis: BC, OP and KW. Interpretation: BS, OP, DC, SH, TR and KW. Manuscript preparation: KW. All authors contributed to the critical revision and approval of the final manuscript.

### ORCID iDs

Kim Wager  <https://orcid.org/0000-0001-8107-9600>

Tomas Rees  <https://orcid.org/0000-0003-0221-0098>

### Supplemental Material

Supplemental material for this article is available online.

### REFERENCES

1. FDA-NIH Biomarker Working Group. *BEST (Biomarkers, Endpoints, and Other Tools) Resource*. Food and Drug Administration (US); 2021.
2. Dingerdissen HM, Bastian F, Vijay-Shanker K, et al. OncoMX: a knowledge-base for exploring cancer biomarkers in the context of related cancer and healthy data. *JCO Clin Cancer Inform*. 2020;4:210-220.
3. Bravo À, Cases M, Queralt-Rosinach N, Sanz F, Furlong LI. A knowledge-driven approach to extract disease-related biomarkers from the literature. *Biomed Res Int*. 2014;2014:253128.
4. Chen HO, Lin PC, Liu CR, Wang CS, Chiang JH. Contextualizing genes by using text-mined co-occurrence features for cancer gene panel discovery. *Front Genet*. 2021;12:771435.
5. Singhal A, Simmons M, Lu Z. Text mining for precision medicine: automating disease-mutation relationship extraction from biomedical literature. *J Am Med Inform Assoc*. 2016;23:766-772.
6. Younesi E, Toldo L, Müller B, et al. Mining biomarker information in biomedical literature. *BMC Med Inform Decis Mak*. 2012;12:148-23249606.
7. Saberian N, Shafi A, Peyvandipour A, Draghici S. MAGPEL: an autoMated pipeline for inferring vAriant-driven Gene panEls from the full-length biomedical literature. *Sci Rep*. 2020;10:12365.
8. Jurca G, Addam O, Aksac A, et al. Integrating text mining, data mining, and network analysis for identifying genetic breast cancer trends. *BMC Res Notes*. 2016;9:236.
9. Hook DW, Porter SJ, Herzog C. Dimensions: Building context for search and evaluation. *Front Res Metr Anal*. 2018;3:3.
10. Lever J, Jones MR, Danos AM, et al. Text-mining clinically relevant cancer biomarkers for curation into the civic database. *Genome Med*. 2019;11:78.
11. Srivastava S, Wagner PD. The early detection research network: A national infrastructure to support the discovery, development, and validation of cancer biomarkers. *Cancer Epidemiol Biomarkers Prev*. 2020;29:2401-2410.
12. Pillich RT, Chen J, Churas C, et al. NDEx: Accessing network models and streamlining network biology workflows. *Curr Protoc*. 2021;1:e258.
13. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep*. 2019;9:5233.
14. Oughtred R, Rust J, Chang C, et al. The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci*. 2021;30:187-200.
15. Fabregat A, Jupe S, Matthews L, et al. The reactome pathway knowledgebase. *Nucleic Acids Res*. 2018;46:D649-D655.
16. Hwang S, Kim CY, Yang S, et al. HumanNet v2: Human gene networks for disease research. *Nucleic Acids Res*. 2019;47:D573-D580.
17. Alanis-Lobato G, Andrade-Navarro MA, Schaefer MH. HIPPIE v2.0: enhancing meaningfulness and reliability of protein-protein interaction networks. *Nucleic Acids Res*. 2017;45:D408-D414.
18. Golbeck J, Frago G, Hartel F, Hendl J, Oberthaler J, Parsia B. The National Cancer Institute's Thesaurus and Ontology. *SSRN Electron J*. 2018;1:1-4. doi:10.2139/ssrn.3199007
19. Kuleshov MV, Jones MR, Rouillard AD, et al. Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016;44:W90-W97.
20. Mollica Poeta V, Massara M, Capucetti A, Bonocchi R. Chemokines and chemokine receptors: new targets for cancer immunotherapy. *Front Immunol*. 2019;10:379.
21. Zhang W, Wang H, Sun M, et al. CXCL5/CXCR2 axis in tumor microenvironment as potential diagnostic biomarker and therapeutic target. *Cancer Commun*. 2020;40:69-80.
22. Xiao Z, Singh S, Singh M. Improving cancer immunotherapy by targeting IL-1. *OncoImmunology*. 2021;10:2008111.
23. Boersma B, Jiskoot W, Lowe P, Bourquin C. The interleukin-1 cytokine family members: role in cancer pathogenesis and potential therapeutic applications in cancer immunotherapy. *Cytokine Growth Factor Rev*. 2021;62:1-14.
24. Vitiello GA, Miller G. Targeting the interleukin-17 immune axis for cancer immunotherapy. *J Exp Med*. 2020;217:e20190456.
25. Montfort A, Colacios C, Levade T, Andrieu-Abadie N, Meyer N, Ségui B. The TNF paradox in cancer progression and immunotherapy. *Front Immunol*. 2019;10:1818. doi:10.3389/fimmu.2019.01818
26. Harrison ML, Obermueller E, Maisey NR, et al. Tumor necrosis factor alpha as a new target for renal cell carcinoma: two sequential phase II trials of infliximab at standard and high dose. *J Clin Oncol*. 2007;25:4542-4549. Oct 10 2007; 25(29):4542-9.
27. Keshavarz A, Pourbagheri-Sigaroodi A, Zafari P, Bagheri N, Ghaffari SH, Bashash D. Toll-like receptors (TLRs) in cancer; with an extensive focus on TLR agonists and antagonists. *IUBMB Life*. 2021;73:10-25.
28. Lalle G, Twardowski J, Grinberg-Bleyer Y. NF-κB in cancer immunity: Friend or foe? *Cells*. 2021;10:9.
29. Blanc RS, Richard S. Arginine methylation: the coming of age. *Mol Cell*. 2017;65:8-24.
30. Kumar V, Donthireddy L, Marvel D, et al. Cancer-associated fibroblasts neutralize the anti-tumor effect of CSF1 receptor blockade by inducing PMN-MDSC infiltration of tumors. *Cancer Cell*. 2017;32:654-668.e5.
31. Cheng KJ, Mejia Mohammed EH, Khong TL, Mohd Zain S, Thavagnanam S, Ibrahim ZA. IL-1α and colorectal cancer pathogenesis: enthralling candidate for anti-cancer therapy. *Crit Rev Oncol Hematol*. 2021;163:103398.
32. Lind DS, Hochwald SN, Malaty J, et al. Nuclear factor-Kappa B is upregulated in colorectal cancer. *Surgery*. 2001;130:363-369.
33. Razi S, Baradaran Noveiry B, Keshavarz-Fathi M, Rezaei N. IL-17 and colorectal cancer: from carcinogenesis to treatment. *Cytokine*. 2019;116:7-12.
34. Stanilov N, Miteva L, Dobrova Z, Stanilova S. Colorectal cancer severity and survival in correlation with tumour necrosis factor-α. *Biotechnol Biotechnol Equip*. 2014;28:911-917.
35. Xu Y, Xu Q, Yang L, et al. Gene expression analysis of peripheral blood cells reveals toll-like receptor pathway deregulation in colorectal cancer. *PLoS One*. 2013;8:e62870.
36. Koundouros N, Pouligiannis G. Reprogramming of fatty acid metabolism in cancer. *Br J Cancer*. 2020;122:4-22.
37. Fucci A, Colangelo T, Votino C, Pancione M, Sabatino L, Colantuoni V. The role of peroxisome proliferator-activated receptors in the esophageal, gastric, and colorectal cancer. *PPAR Res*. 2012;2012:242498.
38. Mou Y, Wang J, Wu J, et al. Ferroptosis, a new form of cell death: opportunities and challenges in cancer. *J Hematol Oncol*. 2019;12:34.