Cancer Immunology, Immunotherapy (2020) 69:2425–2439
https://doi.org/10.1007/s00262-020-02630-8

ORIGINAL ARTICLE

Current trends in cancer immunotherapy: a literature-mining analysis

Stamatia Pouliliou[1] · **Christos Nikolaidis**[2] · **George Drosatos**[3]

Received: 8 May 2020 / Accepted: 28 May 2020 / Published online: 15 June 2020

Abstract
Cancer immunotherapy is a rapidly growing field that is completely transforming oncology care. Mining this knowledge base for biomedically important information is becoming increasingly challenging, due to the expanding number of scientific publications, and the dynamic evolution of this subject with time. In this study, we have employed a literature-mining approach that was used to analyze the cancer immunotherapy-related publications listed in PubMed and quantify emerging trends. A total of 93,033 publications published in 5055 journals have been retrieved, and 141 meaningful topics have been identified, which were further classified into eight distinct categories. Statistical analysis indicates a mean annual increase in the number of published papers of approximately 8% in the last 20 years. The research topics that exhibited the highest trends included "immune checkpoint inhibitors," "tumor microenvironment," "HPV vaccination," "CAR T-cells," and "gene mutations/tumor profiling." The top identified cancer types included "lung," "colorectal," and "breast cancer," and a shift in popularity from hematological to solid tumors was observed. As regards clinical research, a transition from early phase clinical trials to randomized control trials was recorded, indicating that the field is entering a more advanced phase of development. Overall, this mining approach provided an unbiased analysis of the cancer immunotherapy literature in a time-conserving and scale-efficient manner.

Keywords Cancer immunotherapy · Literature mining · Trends analysis · Topic modeling · LDA

Introduction

Cancer immunotherapy is a rapidly growing field that is currently considered as the "fifth pillar" of cancer therapy, thus joining the ranks of surgery, cytotoxic chemotherapy, radiation and targeted therapy as a promising, innovative approach for combating cancer [1]. This stems from the fact that cancer cells express tumor antigens that can be detected and potentially eliminated by the immune system [2]. This can be achieved actively, hence by directly targeting tumor antigens, or passively by enhancing existing anti-tumor responses with the use of monoclonal antibodies, lymphocytes, or cytokines [3].

Our knowledge of the relationship between cancer and the immune system has increased considerably in the last two decades [4]. However, the concept of malignant disease manifestation and immune system interaction has been postulated a long time ago [5]. Hallmarks of this scientific tradition include the discovery of dendritic cells [6], the use of Bacillus Calmette–Guerin (BCG) in bladder cancer [7], the administration of interleukin-2 (IL-2), interferon-alpha (IFN-α), and tumor necrosis factor (TNF) against hematological and solid tumors [8–10], and the discovery of Toll-like receptors' function in immune system evasion, tumor growth and survival [11].

Schreiber and his colleagues [12] describe the cancer immunoediting concept where the tumor immune system balance shifts among tumor escape, equilibrium, and

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00262-020-02630-8) contains supplementary material, which is available to authorized users.

✉ Stamatia Pouliliou
tinapouliliou@gmail.com

Christos Nikolaidis
cnikolae@med.duth.gr

George Drosatos
gdrosato@athenarc.gr

[1] Department of Radiotherapy/Oncology, School of Medicine, Democritus University of Thrace, Alexandroupolis, Greece

[2] School of Medicine, Democritus University of Thrace, Alexandroupolis, Greece

[3] Institute for Language and Speech Processing, Athena Research Center, Xanthi, Greece

elimination. Poor antigenic expression, immunosuppressive cytokines, myeloid-derived suppressor cells (MDSC), and expression of negative regulatory receptors on T cells assist in tumor escape. The tumor and the adaptive immune system coexist in the equilibrium phase where the immune system creates a growth-inhibitory environment, and antigenic tumor outgrowths are kept in check. In tumor elimination, which often occurs in early tumor development, highly antigenic tumor clones are recognized and eliminated by both innate and adaptive immune systems.

Until today, several different types of cancer immunotherapies have been experimentally tested, are currently ongoing clinical trials, or have been approved by major regulatory organizations around the world [13]. The contemporary immunooncology landscape is teeming with an arsenal of active agents, molecular targets, novel therapy types, and cancer-specific applications [14–16]. This expanding universe of scientific information mirrors the complexity of the underlying cancer biology, the diversity of scientific disciplines and research methods used in the analysis, understanding, and manipulation of the immune system, and its ability to fight cancer [17].

Extracting and analyzing this information, in order to identify relevant subjects, explore topic dynamics, and generate new research hypotheses, is becoming increasingly challenging. It is nearly impossible to grasp the entire spectrum of the published literature and synthesize concepts in a systematic and unbiased way [18]. An alternative and more innovative approach for automatically retrieving information from a vast number of scientific publications is through literature mining [19]. Data mining methods have the unique potential of new knowledge discovery, enabling researchers to integrate various sources of information and capture vital scientific insights, in a time-conserving and scale-efficient manner [20, 21]. These methods have been successfully used in the past to inform current practice, update existing protocols, and drive research initiatives across many scientific disciplines [22–25].

In this study, we use a topic modeling approach to understand the historical evolution of cancer immunotherapy, recapture known facts, and identify new subjects and trends that are associated with this rapidly expanding field. One of the most popular topic modeling algorithms, latent Dirichlet allocation (LDA), has been used to pursue text mining of the large cancer immunotherapy corpus [26, 27], and resulting topics and trends are presented herein.

Materials and methods

The analysis of cancer immunotherapy literature is based on the following three-step approach proposed in Drosatos and Kaldoudi [28]:

1. A PubMed query was created to build the corpus of published literature on cancer immunotherapy.
2. The main topics in the area of cancer immunotherapy were identified via the latent Dirichlet allocation (LDA) unsupervised topics modeling algorithm [26].
3. The trends were derived based on the popularity of each topic per year.

These three steps are presented in detail in the following subsections.

Search strategy

The research strategy of our literature-mining analysis followed the core idea of a systematic review. We are differentiated, in our trying to identify a representative set of the literature (not necessarily a superset) that is relevant to the cancer immunotherapy field, and simultaneously minimize the number of irrelevant articles. The resulted articles of this research strategy were then used for deriving topics and their trends.

Given that LDA uses a probabilistic and unsupervised approach, we decided to limit our search in the PubMed search engine [29], the most popular and comprehensive database for biomedical scientific literature, in order to avoid irrelevant articles to the desired field. The syntax of our search query consists of two parts: the synonym terms of "cancer" and the synonym terms of "immunotherapy." This method is known as *query expansion*, and the synonym terms were extracted by the MeSH taxonomy, which is a comprehensive controlled vocabulary for the purpose of indexing articles in PubMed. The exact query was the following:

```
((neoplasms[MeSH] OR neoplasm[TIAB] OR neoplasms[TIAB] OR
   cancer[TIAB] OR cancers[TIAB] OR neoplasia[TIAB] OR
   neoplasias[TIAB] OR tumor[TIAB] OR tumors[TIAB] OR
  tumour[TIAB] OR tumours[TIAB] OR malignancy[TIAB] OR
                malignancies[TIAB])
                      AND
    (immunotherapy[MeSH] OR immunotherapy[TIAB] OR
immunotherapies[TIAB] OR immunological[TIAB] OR "immunological
    therapy"[TIAB] OR "immunological therapies"[TIAB] OR
 "immunologically directed therapy"[TIAB] OR "immunologically
              directed therapies"[TIAB]))
                      AND
        ("0001/01/01"[PDAT] : "2018/12/31"[PDAT])
```

Each part of the query was designed to retrieve articles classified under the specific MeSH term and articles including any of the synonym terms in the title (TI) or abstract (AB) field. The intersection of both parts was performed using the logical operator "AND" in order to retrieve the final corpus of articles published up to the end of year 2018. PubMed database was searched on October 3, 2019, and

the results were downloaded in an XML format using the provided export function.

Topic modeling

Topic modeling algorithms are probabilistic methods that automatically identify topics from a large and unstructured collection of documents. In this work, we used the LDA algorithm [26, 27], and especially its scalable implementation in the MALLET toolkit (v2.0.8) [30], as proposed in Drosatos and Kaldoudi [28]. In order to avoid any noise in the topic modeling from the free text of articles, we applied the following cleaning process: (1) removed all punctuation and non-Latin characters; (2) excluded all stop words using the list in the Text Categorization Project [31]; (3) converted all words to their lemmas by applying the Krovetz stemming procedure [32]; and (4) excluded articles with no words in their abstracts or with less than 3 characters in their titles.

There are several heuristic approaches to tune LDA algorithm for identifying a meaningful number of topics and iterations [33]. In order to determine the appropriate number of topics and iterations to be used as parameters in the LDA algorithm in this study, we followed the approach proposed in Drosatos and Kaldoudi [28]. Thus, we performed a series of exploration experiments using different number of topics (from 50 to 250, with a step of 10) at different iterations (i.e., 8000 and 10000 iterations). For each number of topics and iterations, we repeated the experiment 10 times and calculated the similarity between successive repetitions using Jaccard distance [34]. This similarity distance was calculated between the list of the 10 top words defining each topic with the respective list of words in the different repetition. The final number of topics and iterations were selected when the percentage of topics with a similarity distance less than or equal to 0.57143 achieved a local maximum [28]. This practically means that the topics have more than or equal to 6 words out of the first 10 words (i.e., $\geq 60\%$) in common.

After the tuning process of the LDA algorithm, the artificially generated topics, where each topic consists of a weighted list of words, were screened, manually labeled giving a short label (title), and organized in conceptual categories. This process was independently performed by the authors of this paper for the whole list of topics, taking into account the top 20 words of each topic. Then, the researchers discussed their findings and agreed on a unified topics list with their respective categories.

Trend analysis

The trend analysis of topics was based on the approach proposed in Drosatos and Kaldoudi [28]. First, the weight of each topic for each document was calculated as the percentage of the document words belonging to a topic. Then, the popularity of the topic defined as the yearly topic contribution estimate $P(t,y)$ of the topic (t) for each year (y) was calculated as the average weight of this topic for all documents published that year D_y:

$$P(t,y) = \frac{1}{|D_y|} \sum_{d \in D_y} \frac{\left|\{w \in d : topic(w) = t\}\right|}{|d|} \quad (1)$$

where t represents a topic and w is a word in document d of the documents' collection D_y for year y. Accordingly, the overall popularity of a topic was defined as the overall topic contribution estimate, calculated as the average weight of this topic for all documents included in the corpus.

Finally, we applied moving averaging (over 3 years interval) to smooth out short-term fluctuations. Furthermore, we used the linear regression coefficient to identify the positive or negative trend for each topic. The overall topic modeling, labeling, and trend analysis process that was used on this paper can be also performed via a web-based platform (TM-Toolkit), which allows biomedical researchers with no experience in data modeling and programming to execute topic modeling and trends analysis of the literature using the PubMed database [35].

Results

Search results

The PubMed query (performed on October 3, 2019) returned 108,435 publications (total XML file size of 1.37 GB). Preprocessing excluded 15,402 publications with no abstracts (14.2% of all retrieved records). The final corpus included title, abstract, and keywords of 93,033 publications, corresponding to a total of 11,034,162 words and a vocabulary of 193,497 unique words.

The first record that was retrieved and was included in the final corpus dates to 1922 was published in the Journal of Experimental Medicine. The total retrieved records have been published in 5055 journals, which correspond to 15.5% of the total number of journals that are currently indexed in PubMed (as retrieved on 14-03-2020 from the online PubMed journal list available at https://www.nlm.nih.gov/bsd/serfile_addedinfo.html. The top 5 journal titles were: Cancer Research, AACR; Cancer Immunology, Immunotherapy, Springer; Journal of Immunology, AAI; Bone Marrow Transplantation, Nature; and Blood; ASH. As represented, the cancer immunotherapy domain corresponds to approximately 0.37% of the entire PubMed corpus (as of the end of 2018).

Figure 1 shows the distribution of publications, per year, that were included in the corpus as an absolute value and a percentage of the total number of articles indexed in

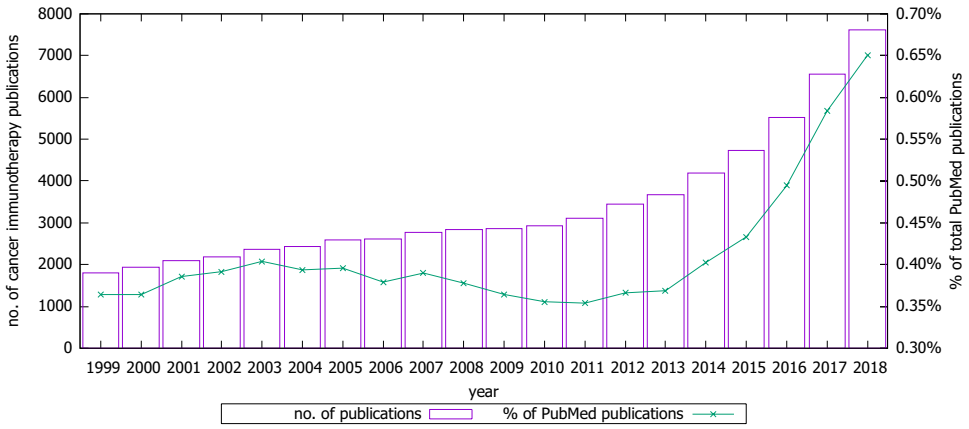


Fig. 1 Cancer immunotherapy publications per year in PubMed

PubMed each year. The cancer immunotherapy field shows a mean increase in the number of published papers of approximately 8% per annum during the last 20 years, and almost 16% in the last 5 years.

Topic modeling: resulting topics and categories

Screening of the 150 topics led to the identification of 141 meaningful topics (94% of all topics), which were organized into eight categories and five subcategories as follows:

1. *Targeted immunotherapy:* Twenty-two topics discussing key concepts of cancer immunotherapy such as "immune checkpoint inhibitors," "PD-1/PD-L1," "CAR T-cells," "bispecific monoclonal antibodies," "invariant natural killer T cells (iNKT)," and "dendritic cell-based vaccines."
2. *Cancer type:* Twenty-one topics corresponding to specific cancer types ("breast cancer," "colorectal cancer," "melanoma," "lymphoma," "lung cancer," etc.).
3. *Diagnosis:* Four topics discussing certain aspects of cancer diagnostics, namely "imaging, "immunohistochemistry," "carcinoembryonic antigen (CEA)," and "leukemia-associated phenotypic markers."
4. *Cancer therapies:* Sixteen topics that include classical cancer therapies (e.g., "chemotherapy," radiotherapy," "bone marrow transplantation," etc.), and modern applications (e.g., "gene therapy," "biological response modifiers," etc.).
5. *Clinical research:* Eighteen topics addressing various aspects of clinical research such as "early phase clinical studies," "randomized control trials," "survival prognostic factors," "clinical study outcomes," and "survival prognostic factors."
6. *Mechanisms of carcinogenesis:* Twelve topics related to the pathogenesis and/or progression of cancer (e.g., "cancer stem cells," "cell adhesion," "lymphatic metastasis," "cell signaling pathways," "tumor microenvironment," "and adaptive immune response").
7. *Translational research:* a broad topic further classified into five subcategories, namely (i) *animal research* including topics such as "murine tumor models" and "rat studies and experiments," (ii) *cell type* including "T regulatory cells," "cytotoxic T lymphocytes," "dendritic cells," "TH1/TH2 response," etc., (iii) *methodology* including "gene expression studies," "cell lines/in vitro culture," "flow cytometry," "nanoparticles," etc., (iv) *pathways and physiology* including "cell metabolism," "apoptosis," "microbiome," "epigenetic regulation," etc., and (v) *protein-based* including "tumor antigens," "chemokine receptors and ligands," "cytotoxic T lymphocytes epitopes," etc.
8. *General:* Seven topics that include reviews on the development of new cancer therapies, including "systematic reviews and meta-analyses," "clinical guidelines," etc.

Table 1 (Appendix) contains the list of identified topics organized in the above categories. The overall popularity of each topic is presented as percentage of the overall topic contribution and is used to calculate the rank of the topic in the entire list. (Most popular topic is ranked first.) Within each category, topics are organized in two groups, corresponding to positive and negative trends, respectively. Within each group, topics are listed in descending order using the absolute value of the corresponding regression coefficient.

Synthesis of results

Research topics with the highest rank (contribution) in the cancer immunotherapy literature included: (i) "tumor microenvironment" (2.17%), (ii) "in vivo intratumoral immunotherapy" (2.15%), (iii) "role of regulatory immune cells" (1.88%), (iv) "adaptive immune responses" (1.74%), (v) "cancer treatment reports" (1.62%), (vi) "syngeneic

mouse models" (1.30%), (vii) "cell lines/in vitro cultures" (1.29%), (viii) "immune checkpoint inhibitors" (1.26%), (ix) "clinical studies—outcomes" (1.20%), and (x) "T-cell memory (CD4+, CD8+)" (1.20%). Representative world clouds of some of the most popular topics are shown in Fig. 2.

The top five research topics with the highest positive trend were: (i) "immune checkpoint inhibitors"; (ii) "tumor microenvironment"; (iii) "HPV vaccination"; (iv) "CAR T-cells"; and (v) "gene mutations/tumor profiling." Likewise the top five research topics with the highest negative trend were: (i) "cytotoxic T lymphocytes (CTL) epitopes"; (ii) "murine tumor models"; (iii) "radioimmunotherapy"; (iv) "cytotoxic T lymphocytes"; and (v) "dendritic cell-based vaccines" (Fig. 3).

In the *Targeted immunotherapy* category, topics with increasing popularity included "immune checkpoint inhibitors," "CAR T-cells," "PD-1/PD-L1," and "tyrosine kinase inhibitors," all showing a significant increase in the last decade. On the other hand, the popularity of "dendritic cell-based vaccines" peaked in 2003 and has since been declining. Likewise the popularity of "adoptive cell transfer" peaked in 2010, and afterward, it has gradually decreased (Fig. 4).

As regards the *Cancer-type* category, a remarkable increase was observed at the "lung cancer" topic, followed by similar increases in "colorectal cancer" and "breast cancer." An increase in "prostate cancer" was observed between 2009 and 2012, and its popularity has since been decreasing. On the contrary, the contribution of "melanoma," "central

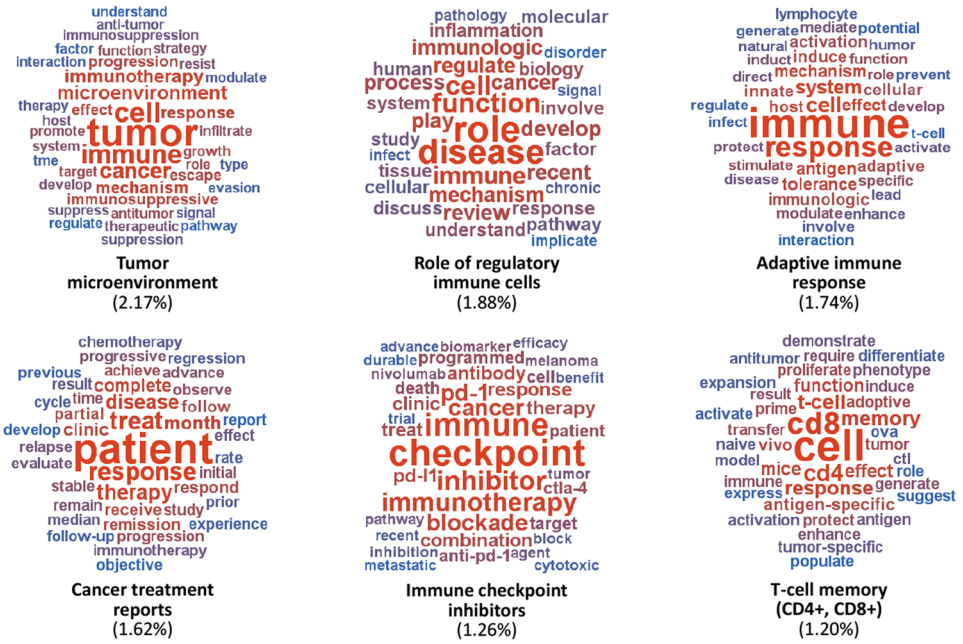


Fig. 2 Word clouds of popular research topics in the cancer immunotherapy corpus

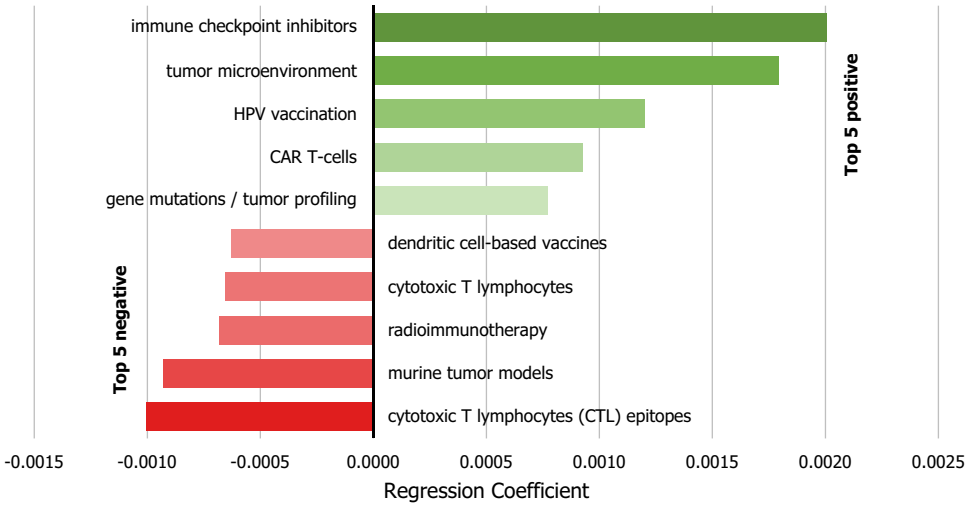


Fig. 3 Top five research topics with the higher positive and negative trends

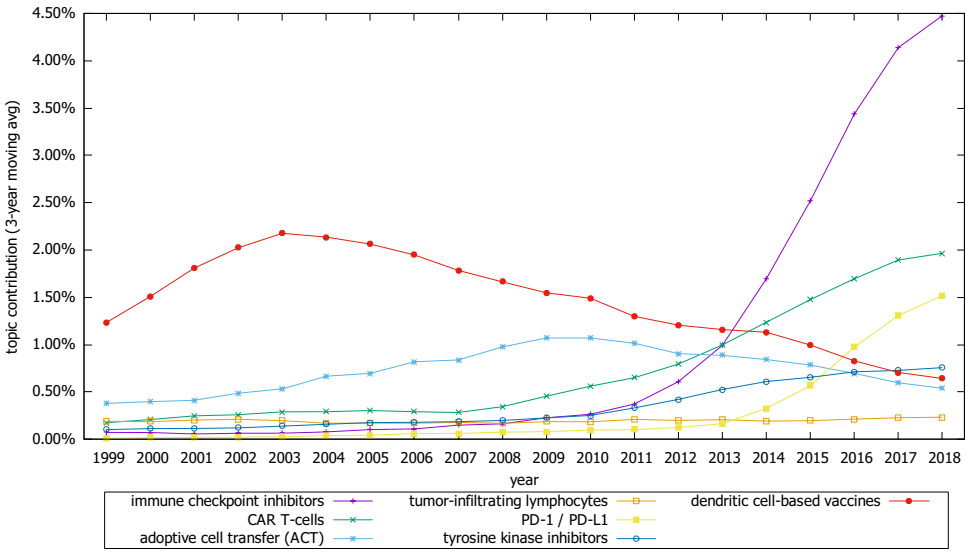


Fig. 4 Trends of selective topics related to targeted immunotherapy category

nervous system cancers," and "chronic lymphocytic leukemia (CLL)" has remained relatively stable over the last 15 years (Fig. 5).

In the *Clinical research* category, a steady decrease was observed in "early phase clinical studies" and "case reports" topics, which was contradicted by a steady increase in "randomized control trials" and "survival prognostic factors." Likewise a sharp increase in "immune-related adverse events" was observed after 2010. The contribution and trend of "clinical studies—outcomes," on the other hand, were relatively stable throughout the study period (Fig. 6).

In the *Mechanisms of carcinogenesis* category, the most remarkable increase was observed in the topic "tumor microenvironment," followed to a lesser extent by "cell signaling pathways" and "cancer stem cells." The contribution of "adaptive immune response" peaked in 2007, and the "role of regulatory immune cells" in 2012, and then, their popularity gradually decreased. The topic "inflammatory mediators" on the other hand has been decreasing in popularity throughout the study period (Fig. 7).

Finally, in the *Translational research* subcategory (a) *pathways and physiology*, topics like "apoptosis," "tumor-draining lymph nodes (TDLN)," and "animal oncogenic viruses" have showed decreasing trends, whereas "epigenetic regulation," "oncolytic viruses," "exosomes," and "microbiome" have been increasing (Fig. 8). As regards subcategory (b) *methodology*, classical topics such as "cell lines/in vitro cultures" have been decreasing, whereas "recombinant DNA vaccines" and "predictive models" have been increasing in popularity. The most remarkable increase, however, was observed in "gene mutations/tumor profiling," which was more pronounced in the last 5 years (Fig. 8).

A more exhaustive presentation of all listed categories and topics is provided in the Supplementary material. The

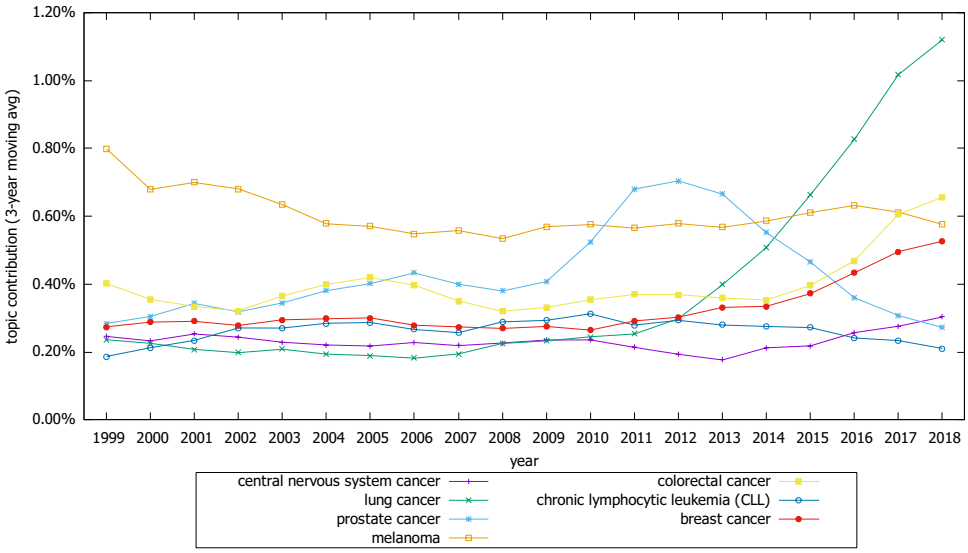


Fig. 5 Trends of selective topics related to cancer type category

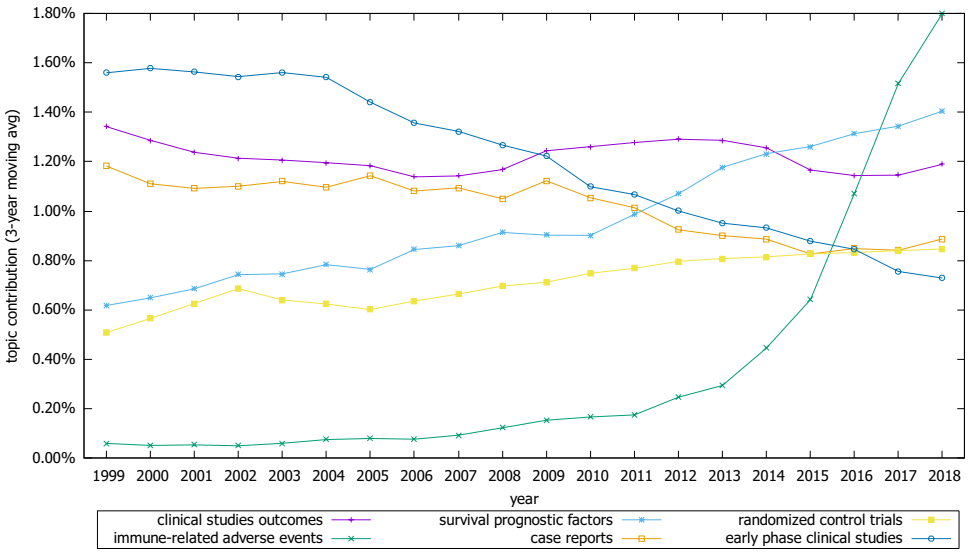


Fig. 6 Trends of selective topics related to clinical research category

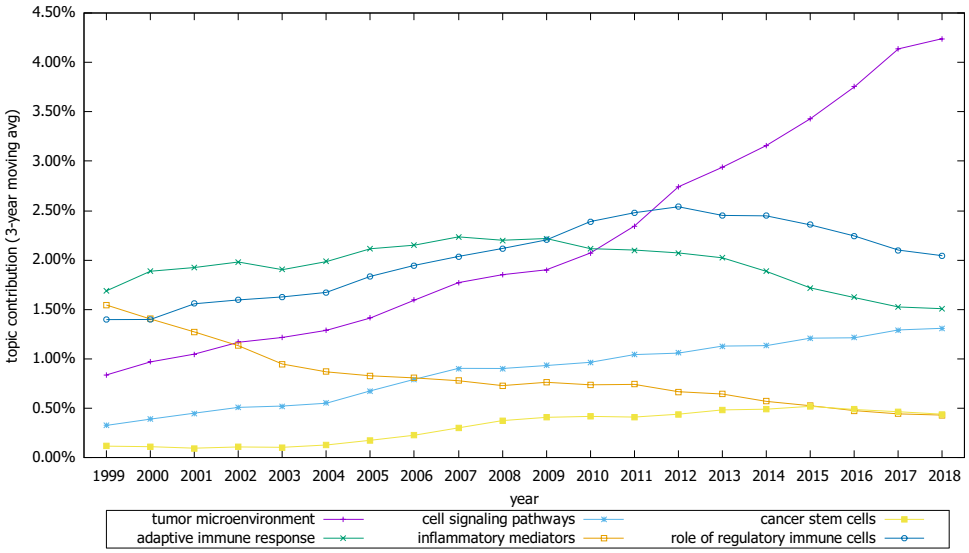


Fig. 7 Trends of selective topics related to mechanisms of carcinogenesis category

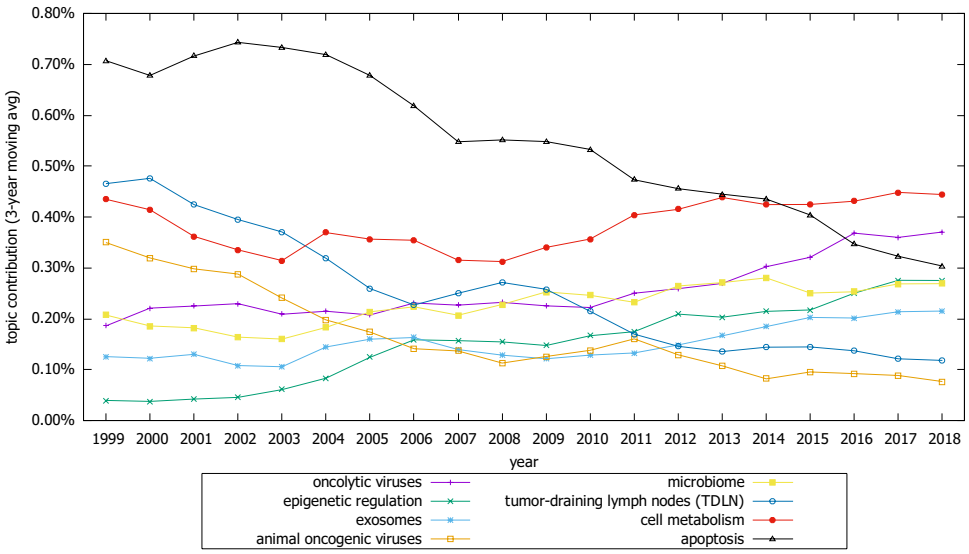
General category (Supp. Figure 12) provides proof-of-concept that cancer immunotherapy is an actively growing field of contemporary biomedical research with increasing trends being observed among seminal topics, such as "systematic reviews and meta-analyses" and "clinical guidelines."

Discussion

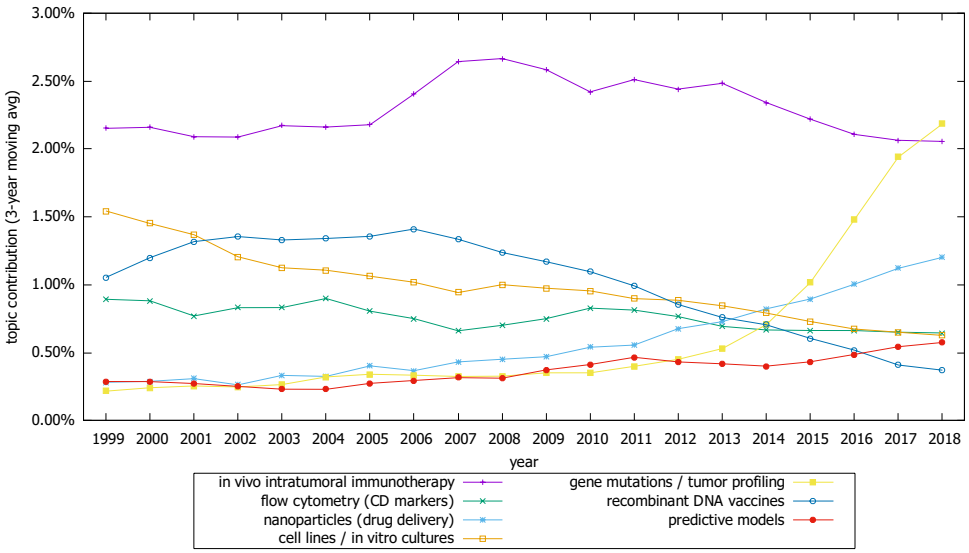
Literature analysis is fundamental for understanding the current state of a research field. It can provide new directions and guide further studies and experimentation. The exponential growth of scientific literature, however, makes this task increasingly challenging. The corpus of published papers is vast, and often not accessible due to copyright and other restrictions [36]. Moreover, this analysis can be extremely laborious, time-consuming, and subject to various types of errors [37].

(Semi)-automated methods, such as topic modeling, can be used to retrieve text-based information and apply it to generate meaningful insights in an efficient and unbiased manner [38]. In our study, we used the latent Dirichlet allocation (LDA) algorithm, a Bayesian hierarchical topic modeling algorithm, which can perform this task with minimal (< 5%) loss of relevant studies, while saving up to 70% of the workload of a classical systematic review [18].

Using this approach, we have analyzed over 90,000 publications in the cancer immunotherapy field, comprising more than 190,000 unique words. Topic modeling led to the identification of an abundance of meaningful topics, classified in conceptualized categories, and subcategories. The cancer immunotherapy domain is relative new and corresponds to a small fraction of the PubMed-listed



(a) Pathways and physiology topics.



(b) Methodology topics.

Fig. 8 Trends of selective topics related to translational research category

scientific literature (< 0.5%). However, its popularity has been increasing in the last 20 years, exhibiting an exponential growth averaging 8% per annum, and 16% in the last 5 years alone.

The category that had the largest number of topics in this study was *Translational research*, followed by *Targeted immunotherapy*, and *Cancer types*. The *General* category comprised mostly of topics related to reviews, clinical guidelines, and meta-analyses and had the highest overall contribution. As regards research topics, remarkable trends were seen among "immune checkpoint inhibitors," "tumor microenvironment," "gene mutations/tumor profiling," and "CAR T-cells." These topics have dominated the field and are currently contributing to approximately 10% of all cancer immunotherapy publications.

With respect to *Cancer types*, leading trends included "lung," "breast," and "colorectal" cancer. This probably reflects the increasing applicability of immunomodulatory checkpoint inhibitors (e.g., atezolizumab, nivolumab, and pembrolizumab) in these cancers, and the high prevalence of

the corresponding malignancies [39]. Among hematologic cancers, "myeloid leukemia" has been steadily decreasing as a topic, whereas "non-Hodgkin lymphoma" peaked in 2007 and then gradually decreased. This phenomenal shift in focus can be perhaps explained by the elaboration of cancer immunotherapy for solid tumors, which has been achieved in recent years [40].

In *Clinical research*, "early phase clinical studies" have been decreasing in popularity, whereas "randomized control trials" have been on the rise. This provides evidence that the field is maturing and is currently advancing to a later stage of development. Hundreds of ongoing Phase III clinical trials that are listed under cancer immunotherapy provide proof-of-concept (https://clinicaltrials.gov). Moreover, this is confirmed by the sharp increase in the interest of immune-related adverse effects. In the *Translational research* domain, prevailing topics included "in vivo intratumoral immunotherapy," "gene mutations/tumor profiling," "nanoparticles (drug delivery)," "epigenetic regulation," and "predictive models." Collectively these trends show a growing interest in personalized cancer therapies, where individual patient characteristics and biomarkers are being used to determine the optimum treatment on a case-by-case basis [41].

Cancer immunotherapy is transforming modern cancer care in an unprecedented way. In this study, we have shown that several ideas have been developed and evolved, and others have been abandoned through years. Immune system factors, such as antibodies, cytokines, CD4+ and CD8+ T cells, dendritic cells, and macrophages, have all been tested for their capacity to fight cancer with varied effectiveness. As of today, targeted immunotherapies (e.g., monoclonal antibodies, antibody–drug conjugates, and bispecific antibodies), immunomodulators (e.g., checkpoint inhibitors, cytokines, and adjuvants), preventive and therapeutic cancer vaccines (e.g., HPV, HBV, BCG, and Sipuleucel-T), oncolytic viruses (e.g., T-VEC using modified HSV), and adoptive cell therapies (e.g., CAR T cells) have been approved and marketed as novel anti-neoplastic medications [42].

Currently, there is a growing interest for dissemination, and real-world effectiveness of these medications among patient populations. This is realized as a T3–T4 transition in the transnational medicine continuum, suggesting that the field is entering its final, and most important stage [43]. From our analysis, we conclude that it is possible to identify these insights, using a data-driven approach. LDA-mediated topic modeling provides several advantages over traditional methods and is emerging as an effective, unbiased method for conducting this type of research. The limitation of our analysis is that we only use one database (PubMed) as a source of the entire cancer immunotherapy literature. However, the exclusive use of this database has been shown to achieve a precision of over 80%, when compared to combinations of PubMed, Embase, Web of Science, and Google Scholar [44]. The quantitative trends that have been displayed herein can be, thus, used as a good starting point for further experimentation and guide new research initiatives. This dynamically evolving field has the capacity to transform evidence generation and will be used more frequently in the future.

Author contributions All authors have contributed equally. All authors read and approved the final manuscript.

Funding This research is partially co-financed by Greece and the European Union (European Social Fund - ESF) through the Operational Programme "Human Resources Development, Education and Lifelong Learning" in the context of the project "Reinforcement of Postdoctoral Researchers - 2nd Cycle" (MIS-5033021), implemented by the State Scholarships Foundation (IKY).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Appendix

List of identified topics

Table 1 List of topics organized in 8 categories (and 5 subcategories), showing the regression analysis results, the overall popularity metric and the respective rank of the topics

Topic label	Trend analysis		Topic popularity	
	Reg. coeff.	R^2 (%)	Contrib (%)	Rank
Category: Targeted Immunotherapy				
Positive trends				
Immune checkpoint inhibitors	0.002007	65.99	1.26	11
HPV vaccination	0.001203	78.23	1.08	17
CAR T cells*	0.000926	81.98	0.76	35
PD-1 / PD-L1	0.000581	57.73	0.37	86
Tyrosine kinase inhibitors*	0.000376	87.56	0.35	95
Toll-like receptor and CpG oligodeoxynucleotides	0.000244	77.61	0.34	97
Adoptive cell transfer (ACT)	0.000174	21.03	0.62	54
Indoleamine 2.3-dioxygenase (IDO)*	0.000112	90.57	0.15	138
Tumor-associated macrophages (TAMs)	0.000060	35.42	0.26	117
Invariant natural killer T cells (iNKT)	0.000052	26.03	0.16	137
Tumor-infiltrating lymphocytes	0.000017	27.36	0.21	128
Vascular endothelial growth factor (VEGF)†	0.000013	03.06	0.32	103
Bispecific monoclonal antibodies†	0.000004	00.08	0.64	51
Negative trends				
Dendritic cell-based vaccines	− 0.000628	60.15	1.01	20
Anti-idiotypic monoclonal antibodies*	− 0.000326	95.78	0.75	36
IL-2 immunotherapy*	− 0.000325	87.08	0.66	49
Carbohydrate tumor antigen*	− 0.000202	80.36	0.34	98
Cancer testis antigens (CTA)	− 0.000130	37.86	0.35	94
Granulocyte-macrophage colony-stimulating factor (GM-CSF)*	− 0.000128	94.38	0.21	130
HER2/neu	− 0.000072	47.76	0.20	133
Immunotoxins	− 0.000042	42.11	0.22	126
Heat shock protein†	− 0.000021	16.54	0.20	131
Category: Cancer Type				
Positive trends				
Lung cancer	0.000391	62.80	0.46	70
Pancreatic & head and neck cancer	0.000147	67.94	0.39	83
Cervical cancer (HPV)	0.000145	32.49	0.57	60
Glioblastoma*	0.000106	81.61	0.37	87
Breast cancer	0.000094	53.78	0.36	89
Colorectal cancer	0.000084	31.42	0.43	73
Prostate cancer†	0.000080	12.58	0.35	91
Ovarian cancer*	0.000063	82.44	0.25	121
Hepatocellular carcinoma	0.000059	64.52	0.26	119
Pediatric cancers	0.000050	70.32	0.26	116
Multiple myeloma	0.000019	42.33	0.25	120
Chronic lymphocytic leukemia (CLL)†	0.000009	02.74	0.26	115
Central nervous system cancer†	0.000006	01.62	0.26	118
Negative trends				
Myeloid leukemia*	− 0.000252	91.90	0.81	31
Non-Hodgkin lymphoma†	− 0.000149	15.61	0.53	66
Lymphoma*	− 0.000141	92.30	0.55	65
Renal cancer	− 0.000115	47.29	0.46	71
Melanoma	− 0.000058	28.25	0.61	55
Skin lesions of the face	− 0.000030	22.08	0.38	84
Endocrine tumors†	− 0.000022	19.34	0.21	129
Malignant mesothelioma	− 0.000007	26.95	0.16	136

Table 1 (continued)

Topic label	Trend analysis		Topic popularity	
	Reg. coeff.	R^2 (%)	Contrib (%)	Rank
Category: Diagnosis				
Positive trends				
Imaging*	0.000188	89.09	0.28	111
Negative trends				
Immunohistochemistry*	− 0.000314	91.44	0.94	23
Leukemia-associated phenotypic markers*	− 0.000234	88.79	0.71	43
Carcinoembryonic antigen (CEA)*	− 0.000083	94.53	0.13	141
Category: Cancer Therapies				
Positive trends				
Allogeneic stem cell transplantation	0.000680	61.32	0.92	24
Radiotherapy	0.000108	46.44	0.35	92
Photoimmunotherapy*	0.000105	91.27	0.19	135
Biological response modifiers (BRM)	0.000044	23.49	0.34	96
Anti-TNF therapies†	0.000042	17.38	0.32	102
Negative trends				
Radioimmunotherapy*	− 0.000682	93.95	0.72	40
Gene therapy*	− 0.000543	98.60	0.57	61
Bone marrow transplantation*	− 0.000437	81.62	0.44	72
Radioimmunotherapy*	− 0.000399	94.02	0.33	101
Adjuvant treatments	− 0.000301	77.00	1.14	15
Vaccine adjuvants	− 0.000148	27.12	0.91	25
Chemotherapy	− 0.000135	79.58	0.64	52
Interferon-based treatments*	− 0.000128	91.49	0.24	122
Bladder cancer intravesical therapy	− 0.000096	55.92	0.56	64
Clinical nutrition in surgery	− 0.000054	52.19	0.31	107
Vaccination†	− 0.000019	06.64	0.35	90
Category: Clinical Research				
Positive trends				
Immune-related adverse events	0.000670	60.00	0.48	68
Anti-NMDA receptor encephalitis	0.000433	76.72	0.63	53
Survival prognostic factors*	0.000412	95.72	0.94	22
Progression-free survival (PFS)*	0.000256	93.65	0.71	44
Randomized control trials*	0.000165	90.96	0.66	48
Hypersensitivity reactions†	0.000009	01.42	0.22	125
Quality of life†	0.000009	05.21	0.35	93
Cancer risk factors†	0.000003	00.10	0.61	56
Negative trends				
Early phase clinical studies*	− 0.000501	97.15	1.16	14
Cancer treatment reports*	− 0.000473	91.49	1.62	8
CD34+ hematopoietic stem cells*	− 0.000450	87.27	0.52	67
Serum biomarkers*	− 0.000275	83.10	0.79	32
Lymphocyte stimulation*	− 0.000237	95.47	0.90	26
Peripheral blood mononuclear cells (PBMC)*	− 0.000232	86.80	0.70	45
Serum immunoglobulins*	− 0.000211	99.37	0.56	63
Case reports*	− 0.000178	81.88	1.14	16
Infections and immunodeficiency	− 0.000077	69.14	0.33	100
Clinical studies outcomes†	− 0.000031	09.20	1.20	12

Table 1 (continued)

Topic label	Trend analysis		Topic popularity	
	Reg. coeff.	R^2 (%)	Contrib (%)	Rank
Category: Mechanisms of Carcinogenesis				
Positive trends				
Tumor microenvironment*	0.001795	94.88	2.17	4
Cell signaling pathways*	0.000531	97.70	0.78	33
Role of regulatory immune cells	0.000520	67.07	1.88	6
Cancer stem cells*	0.000255	87.71	0.29	109
Negative trends				
Inflammatory mediators*	− 0.000489	87.02	0.84	29
Epstein–Barr virus and cytomegalovirus*	− 0.000141	88.63	0.33	99
Adaptive immune response†	− 0.000137	12.90	1.74	7
Tumor metastasis*	− 0.000125	80.53	0.42	77
UV DNA damage (skin cancer)	− 0.000089	67.46	0.23	124
Lymphatic metastasis*	− 0.000053	80.47	0.27	112
Cell adhesion	− 0.000053	36.70	0.42	76
Viral hepatitis	− 0.000033	29.08	0.31	105
Category: Translational Research				
Subcategory: Pathways and Physiology				
Positive trends				
Epigenetic regulation*	0.000130	95.89	0.15	139
Oncolytic viruses	0.000084	76.42	0.27	113
Microbiome	0.000056	76.61	0.23	123
Exosomes	0.000048	65.82	0.14	140
Cell metabolism	0.000042	27.30	0.42	79
Negative trends				
Apoptosis*	− 0.000239	93.85	0.42	75
Tumor-draining lymph nodes (TDLN)*	− 0.000192	90.82	0.27	114
Animal oncogenic viruses*	− 0.000131	82.34	0.41	80
Subcategory: Methodology				
Positive trends				
Gene mutations / tumor profiling	0.000771	61.31	0.70	46
Nanoparticles (drug delivery)*	0.000470	88.94	0.57	62
Predictive models*	0.000160	83.11	0.36	88
Test assays	0.000042	19.89	0.75	37
In vivo intratumoral immunotherapy†	0.000037	01.13	2.15	5
Negative trends				
Recombinant DNA vaccines	− 0.000501	73.01	0.72	39
Cell lines / in vitro cultures*	− 0.000417	92.98	1.29	10
Gene expression studies	− 0.000222	66.06	0.74	38
Protein purification (chromatography)*	− 0.000221	88.45	0.83	30
Protein/gene sequencing*	− 0.000204	91.74	0.40	82
Flow cytometry (CD markers)	− 0.000117	63.64	0.77	34
Bacteria-based immunotherapies	− 0.000038	41.03	0.20	132
BCG-based immunotherapies	− 0.000035	49.22	0.31	106
Gene polymporphisms†	− 0.000017	08.55	0.21	127
Subcategory: Protein-based				
Positive trends				
Tumor necrosis factor receptor (TNFR) superfamily	0.000090	75.75	0.43	74
Chemokine receptors and ligands	0.000056	37.71	0.19	134

Table 1 (continued)

Topic label	Trend analysis		Topic popularity	
	Reg. coeff.	R^2 (%)	Contrib (%)	Rank
Negative trends				
Cytotoxic T lymphocytes (CTL) epitopes*	− 0.001005	97.16	0.98	21
Tumor antigens*	− 0.000420	83.00	0.88	28
Single-chain variable fragment (scFv) antibodies	− 0.000174	76.30	0.72	42
T cell receptor (TCR)†	− 0.000002	00.16	0.32	104
Subcategory: Cell Type				
Positive trends				
Natural killer cells*	0.000398	90.08	0.66	50
T regulatory cells	0.000394	47.25	0.47	69
Myeloid-derived suppressor cells*	0.000365	83.64	0.31	108
T cell memory (CD4+, CD8+)†	0.000052	01.26	1.20	13
Negative trends				
Cytotoxic T lymphocytes*	− 0.000655	98.64	1.03	19
TH1/TH2 response*	− 0.000351	90.10	0.72	41
Dendritic cells	− 0.000326	57.70	0.58	58
Cytokine-induced killer cells (CIK)	− 0.000037	25.48	0.42	78
Subcategory: Animal Models				
Positive trends				
n/a				
Negative trends				
Murine tumor models*	− 0.000928	88.31	0.68	47
Mouse studies and experiments*	− 0.000455	94.31	1.08	18
Syngeneic mouse models*	− 0.000292	80.53	1.30	9
Rat studies and experiments*	− 0.000159	86.47	0.38	85
Immunodeficient mouse models	− 0.000071	38.50	0.58	59
Category: General				
Positive trends				
Targeted therapies (review)*	0.001418	88.93	3.46	1
Development of new cancer therapies (reviews)*	0.001266	91.81	2.64	3
Cancer immunotherapy (reviews)*	0.000851	88.63	3.24	2
Clinical guidelines*	0.000333	87.83	0.60	57
Systematic review and meta-analysis*	0.000262	92.18	0.29	110
Significant statistical results*	0.000260	88.79	0.89	27
Negative trends				
Anti-tumor immunity*	− 0.000153	87.05	0.40	81

Within each category, topics are organized in two groups corresponding to positive and negative trends, respectively; within each group, topics are listed with descending order of regression coefficient

* Topics with a good linear regression fit, R-squared > 80%

† Topics with nonsignificant reg. coefficient, p value > 0.05

References

1. Oiseth SJ, Aziz MS (2017) Cancer immunotherapy: a brief review of the history, possibilities, and challenges ahead. J Cancer Metastasis Treat 3(10):250. https://doi.org/10.20517/2394-4722.2017.41
2. Marabelle A, Tselikas L, Baere Td, Houot R (2017) Intratumoral immunotherapy: using the tumor as the remedy. Ann Oncol 28(Suppl. 12):xii33–xii43. https://doi.org/10.1093/annonc/mdx683
3. Mellman I, Coukos G, Dranoff G (2011) Cancer immunotherapy comes of age. Nature 480(7378):480–489. https://doi.org/10.1038/nature10673
4. Zhang Y, Quan L, Du L (2019) The 100 top-cited studies in cancer immunotherapy. Artif Cells Nanomed Biotechnol 47(1):2282–2292. https://doi.org/10.1080/21691401.2019.1623234
5. Dobosz P, Dzieciatkowski T (2019) The intriguing history of cancer immunotherapy. Front Immunol 10:2965. https://doi.org/10.3389/fimmu.2019.02965

6. Whiteside TL, Odoux C (2004) Dendritic cell biology and cancer therapy. Cancer Immunol Immunother 53(3):240–248. https://doi.org/10.1007/s00262-003-0468-6
7. Fuge O, Vasdev N, Allchorne P, Green JS (2015) Immunotherapy for bladder cancer. Res Rep Urol 7:65–79. https://doi.org/10.2147/RRU.S63447
8. Jiang T, Zhou C, Ren S (2016) Role of IL-2 in cancer immunotherapy. OncoImmunology 5(6):e1163462. https://doi.org/10.1080/2162402X.2016.1163462
9. Kirkwood J (2002) Cancer immunotherapy: the interferon-α experience. Semin Oncol 29(3, Suppl. 7):18–26. https://doi.org/10.1053/sonc.2002.33078
10. Waters JP, Pober JS, Bradley JR (2013) Tumour necrosis factor and cancer. J Pathol 230(3):241–248. https://doi.org/10.1002/path.4188
11. Rakoff-Nahoum S, Medzhitov R (2009) Toll-like receptors and cancer. Nat Rev Cancer 9(1):57–63. https://doi.org/10.1038/nrc2541
12. Dunn GP, Old LJ, Schreiber RD (2004) The immunobiology of cancer immunosurveillance and immunoediting. Immunity 21(2):137–148. https://doi.org/10.1016/j.immuni.2004.07.017
13. Zhang H, Chen J (2018) Current status and future directions of cancer immunotherapy. J Cancer 9(10):1773–1781. https://doi.org/10.7150/jca.24577
14. Kohrt HE, Tumeh PC, Benson D, Bhardwaj N, Brody J, Formenti S, Fox BA, Galon J, June CH, Kalos M, Kirsch I, Kleen T, Kroemer G, Lanier L, Levy R, Lyerly HK, Maecker H, Marabelle A, Melenhorst J, Miller J, Melero I, Odunsi K, Palucka K, Peoples G, Ribas A, Robins H, Robinson W, Serafini T, Sondel P, Vivier E, Weber J, Wolchok J, Zitvogel L, Disis ML, Cheever MA, on behalf of the Cancer Immunotherapy Trials Network (CITN) (2016) Immunodynamics: a cancer immunotherapy trials network review of immune monitoring in immuno-oncology clinical trials. J Immunother Cancer 4(1):15. https://doi.org/10.1186/s40425-016-0118-0
15. Marshall HT, Djamgoz MBA (2018) Immuno-oncology: emerging targets and combination therapies. Front Oncol 8:315. https://doi.org/10.3389/fonc.2018.00315
16. Farkona S, Diamandis EP, Blasutig IM (2016) Cancer immunotherapy: the beginning of the end of cancer? BMC Med 14(1):73. https://doi.org/10.1186/s12916-016-0623-5
17. Subramanian N, Torabi-Parizi P, Gottschalk RA, Germain RN, Dutta B (2015) Network representations of immune system complexity: immune networks. Wiley Interdiscip Rev Syst Biol Med 7(1):13–38. https://doi.org/10.1002/wsbm.1288
18. O'Mara-Eves A, Thomas J, McNaught J, Miwa M, Ananiadou S (2015) Using text mining for study identification in systematic reviews: a systematic review of current approaches. Syst Rev 4(1):5. https://doi.org/10.1186/2046-4053-4-5
19. Mo Y, Kontonatsios G, Ananiadou S (2015) Supporting systematic reviews using LDA-based document representations. Syst Rev 4(1):172. https://doi.org/10.1186/s13643-015-0117-0
20. Zou C (2018) Analyzing research trends on drug safety using topic modeling. Expert Opin Drug Saf 17(6):629–636. https://doi.org/10.1080/14740338.2018.1458838
21. Bisgin H, Liu Z, Fang H, Xu X, Tong W (2011) Mining FDA drug labels using an unsupervised learning technique—topic modeling. BMC Bioinform 12(Suppl. 10):S11. https://doi.org/10.1186/1471-2105-12-S10-S11
22. Andronis C, Sharma A, Virvilis V, Deftereos S, Persidis A (2011) Literature mining, ontologies and information visualization for drug repurposing. Brief Bioinform 12(4):357–368. https://doi.org/10.1093/bib/bbr005
23. Wang SH, Ding Y, Zhao W, Huang YH, Perkins R, Zou W, Chen JJ (2016) Text mining for identifying topics in the literatures about adolescent substance use and depression. BMC Public Health 16(1):279. https://doi.org/10.1186/s12889-016-2932-1
24. Zhou X, Peng Y, Liu B (2010) Text mining for traditional Chinese medical knowledge discovery: a survey. J Biomed Inform 43(4):650–660. https://doi.org/10.1016/j.jbi.2010.01.002
25. Faro A, Giordano D, Spampinato C (2012) Combining literature text mining with microarray data: advances for system biology modeling. Brief Bioinform 13(1):61–82. https://doi.org/10.1093/bib/bbr018
26. Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. J Mach Learn Res 3(Jan):993–1022
27. Blei DM (2012) Probabilistic topic models. Commun ACM 55(4):77–84. https://doi.org/10.1145/2133806.2133826
28. Drosatos G, Kaldoudi E (2019) A probabilistic semantic analysis of ehealth scientific literature. J Telemed Telecare 00:1–19. https://doi.org/10.1177/1357633X19846252
29. PubMed, US National Library of Medicine (2019) PubMed—biomedical literature from MEDLINE. https://www.ncbi.nlm.nih.gov/pubmed/. Accessed 29 Dec 2019
30. McCallum AK (2002) Mallet: a machine learning for language toolkit. http://mallet.cs.umass.edu. Accessed 20 Feb 2019
31. Text Categorization Project (2011) Lists of stopwords. http://code.google.com/p/text-categorization/. Accessed 20 Feb 2019
32. Krovetz R (1993) Viewing morphology as an inference process. In: 16th Annual international ACM SIGIR conference on research and development in information retrieval. ACM, New York, NY, SIGIR '93, pp 191–202. https://doi.org/10.1145/160688.160718
33. Agrawal A, Fu W, Menzies T (2018) What is wrong with topic modeling? And how to fix it using search-based software engineering. Inf Softw Technol 98:74–88. https://doi.org/10.1016/j.infsof.2018.02.005
34. Jaccard P (1912) The distribution of the flora in the alpine zone. New Phytol 11(2):37–50. https://doi.org/10.1111/j.1469-8137.1912.tb05611.x
35. Kavvadias S, Drosatos G, Kaldoudi E (2018) An online service for topics and trends analysis in medical literature. In: Lhotska L, Sukupova L, Lacković I, Ibbott GS (eds) World congress on medical physics and biomedical engineering, 3–8 June 2018, Prague, Czech Republic, IFMBE proceedings, vol 68/3. Springer Singapore
36. Reichman JH, Okediji RL (2012) When copyright law and science collide: empowering digitally integrated research methods on a global scale. Minn Law Rev 96(4):1362–1480
37. Ahmed I, Sutton AJ, Riley RD (2012) Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. BMJ 344(1 jan03):d7762–d7762. https://doi.org/10.1136/bmj.d7762
38. Zweigenbaum P, Demner-Fushman D, Yu H, Cohen KB (2007) Frontiers of biomedical text mining: current progress. Brief Bioinform 8(5):358–375. https://doi.org/10.1093/bib/bbm045
39. Raju S, Joseph R, Sehgal S (2018) Review of checkpoint immunotherapy for the management of non-small cell lung cancer. ImmunoTargets Ther 7:63–75. https://doi.org/10.2147/ITT.S125070
40. Kather JN, Berghoff AS, Ferber D, Suarez-Carmona M, Reyes-Aldasoro CC, Valous NA, Rojas-Moraleda R, Jäger D, Halama N (2018) Large-scale database mining reveals hidden trends and future directions for cancer immunotherapy. OncoImmunology 7(7):e1444412. https://doi.org/10.1080/2162402X.2018.1444412
41. Dumbrava EI, Meric-Bernstam F (2018) Personalized cancer therapy—leveraging a knowledge base for clinical decision-making. Mol Case Stud 4(2):a001578. https://doi.org/10.1101/mcs.a001578
42. Tang J, Pearce L, O'Donnell-Tormey J, Hubbard-Lucey VM (2018) Trends in the global immuno-oncology landscape. Nat Rev Drug Discov 17(11):783–784. https://doi.org/10.1038/nrd.2018.167

43. Klevorn LE, Teague RM (2016) Adapting cancer immunotherapy models for the real world. Trends Immunol 37(6):354–363. https://doi.org/10.1016/j.it.2016.03.010
44. Bramer WM, Rethlefsen ML, Kleijnen J, Franco OH (2017) Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. Syst Rev 6(1):245. https://doi.org/10.1186/s13643-017-0644-y