Scientometrics (2019) 119:1545–1574
https://doi.org/10.1007/s11192-019-03095-y

# Overview of trends in global epigenetic research (2009–2017)

**Carlos Olmeda-Gómez**[1] · **Carlos Romá-Mateo**[2,3,4,5] · **Maria-Antonia Ovalle-Perandones**[1]

Received: 6 September 2018 / Published online: 6 April 2019


**Abstract**
Epigenetics, one of the most rapidly intensifying fields of biological research, explores inheritable gene expression not governed by alterations in the DNA sequence. This article analysed the literature on epigenetics published between 2009 and 2017 using qualitative and visualisation techniques. The data were drawn from Clarivate Analytics' Web of Science Core Collection in January 2018. CiteSpace V software was used to establish an intellectual overview, based on 13,295 scientific articles and review papers. Document co-citations were analysed and a variety of graphics was created. The aim was to define the scope of the field and identify its constituent specialities with automatic procedures based on the keywords and titles in the citing articles in the clusters identified. Ten subspecialities were identified. Field core papers were defined from the co-citation findings, Kleinberg's burst detection algorithm and the structural variation theory metrics calculated for all the articles selected. The findings can be used to acquire an in-depth view of the patterns and trends in place in the domain and identify innovative and potentially transformational studies.

**Keywords** Epigenetics · Co-citation analysis · Structural variation theory · Core documents · Knowledge domain · CiteSpace

## Introduction

Biomedical research in the field of epigenetics has intensified substantially over the last 10 years and is currently a hot topic not only for understanding cellular and molecular processes, but also for biomarker design, disease prediction and the development of biotechnological tools for agro-ecology. The popularity of the domain is attributable both

✉ Carlos Olmeda-Gómez
olmeda@bib.uc3m.es

[1] Department of Library and Information Science, Carlos III University, Madrid, Spain

[2] Department Physiology, School of Medicine and Dentistry, University of Valencia, Valencia, Spain

[3] FIHCUV-INCLIVA, Valencia, Spain

[4] CIBERER, Centro de Investigación Biomédica en Red de Enfermedades Raras, Valencia, Spain

[5] Epigenetics Research Platform (CIBERER-UV-INCLIVA), Valencia, Spain

to those advances and to the publication of seminal studies that revealed promising links between genetic mechanisms and environmental factors, mediated by epigenetic processes. Nonetheless, the complexity of these mechanisms and their relationship with the broader laws of genetics has spawned misunderstandings and criticism around the use of the term 'epigenetics', or even the mechanisms so labelled.

The word 'epigenetics' appeared very early in the history of human science: Aristotle used the term 'epigenesis' to describe the development of a fertilised egg into a whole, differentiated organism. That idea stood in contrast to preformation, according to which all adult structures are present in the embryo, albeit in microscopic dimensions. Many centuries were needed to fully rule out preformation and establish epigenesis as the process governing embryonic growth. That may explain why it was in developmental biology, which studies the post-fertilisation biological changes involved in growth from the molecular to the whole organism level, where the term epigenetics acquired its modern meaning. In 1942 Conrad Hal Waddington coined the term 'epigenetics' to define the layers of regulation that link the cellular differentiation in a developing embryo to genetic changes dependent upon the environmental conditions prevailing in its cells (Waddington 2012). Further to that definition, irrespective of genomic information, external stimuli would govern cellular differentiation, a key process in determining the features that afford cells their specificity.

Epigenetics, which literally means 'over genetics', was soon used to define any change in genetic expression that could be inherited by cells, regardless of the information contained in the DNA sequence. That definition has since been restated, ranging from a 'stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence' (Berger et al. 2009) to 'heritable changes in gene activity and expression (in the progeny of cells or of individuals) and also stable, long-term alterations in the transcriptional potential of a cell that are not necessarily heritable' (Bernstein et al. 2006). The development of epigenetics as a discipline has met with a number of issues, however, and the claim that epigenetic marks can be inherited in humans has proven to be particularly controversial (Nagy and Turecki 2015). Hence, a clear definition of what is and what is not epigenetics, or with a methodological twist, which specific biological processes can be contextualized and clearly delimited within the discipline, is imperative to understanding its scope.

Biomedically speaking, research in epigenetics has shown that epigenetic marks and the activity of their regulators constitute very useful trackers of cellular, especially pathological, activity. Epigenetic parameters have prompted revolutionary progress in fields such as biomarkers (Garcia-Gimenez et al. 2017), disease prediction (Feinberg 2018) and drug design, and have proven particularly significant for the study of cancer biology (Agrawal et al. 2018; Bai et al. 2018), ageing (Gangisetty et al. 2018), and neuronal behaviour (Holmes 2018). Epigenetic regulation also informs promising strategies in ecology geared to understanding the effects of climate change on biodiversity and in agriculture to improving breeding systems (Jeremias et al. 2018; Ramirez-Prado et al. 2018). In this context, there is a clear need for a comprehensive, chronological and conceptual bibliographic study of the use of epigenetics and the relationships among key studies in which epigenetic terminology is present. This article aims to identify the hot topics in epigenetics research today and explain the role of the discipline in scientific advancement.

This study aims to provide an overview of the most recent core papers published on epigenetics in general, worldwide. The research questions (*RQ*) posed are as follows.

(1) *RQ*: Globally speaking, what areas and sub-areas comprise the research core in epigenetics?
(2) *RQ:* What is the overall citation structure exhibited by the papers published on epigenetics in the period studied?
(3) *RQ:* What research topics are addressed and how are they structured? Which papers form part of the intellectual base and which of the research front?
(4) *RQ:* Which papers have attracted intense researcher attention and for how long were they deemed 'hot topics'?
(5) *RQ:* Which recent papers contain transformational potential?

## Related studies

With the diversification and restructuring of scientific communication, the rigorous, transparent and reproducible tracking of scientific literature and findings in a given domain is becoming increasingly difficult. Traditionally, scoping or systematic reviews were conducted of the literature on a given area of knowledge. Scoping reviews entail broad exploratory searches of the literature relevant to a domain to establish its key characteristics (Arksey and O'Malley 2005). Their implementation is complex, however, among others because filtering the enormous and exponentially growing datasets and volumes of literature published is time consuming (Bell et al. 2009; van Noorden et al. 2014; Bornmann and Mutz 2015). The sub-fields of biology, where the number of papers has doubled every 15 years on average (Milojevic 2015), are no exception to that trend (Pautasso 2012). The time required to write up a report of scoping review findings may range from 6 to 12 months (Pham et al. 2014).

Science mapping is a generic process for analysing and visualising areas of knowledge by retrieving and processing a given corpus of scientific literature. The techniques involved are borrowed primarily from computer science, metric studies on science, data and information visualisation, visual analytics and network science (Börner et al. 2003). The findings are varied and depend on the scale of the analysis. They enable scientists to identify other groups of researchers, measure the impact of their own papers (Börner et al. 2005), recognise core topics and the references persistently invoked by domain scientists (Boyack et al. 2009; Glänzel and Thijs 2017) and find possible partners to collaborate in areas of mutual interest (Boyack 2009). Other dynamic features of knowledge production (Morris et al. 2003) are likewise analysed. By way of example, journal maps (Leydesdorff et al. 2017a), patents (Leydesdorff et al. 2017b), national trends in the literature on a given topic stored in bibliographic databases with citation indices (Darvish and Tonta 2016; Rainho Bras et al. 2017), ground-breaking papers (Zhu et al. 2017), the geographic dissemination of knowledge (Börner et al. 2006) can be detected and visualised, and global science maps drawn (Börner et al. 2012; Boyack and Klavans 2014).

With the use of citation analysis and visualisation techniques, researchers or anyone interested in a given domain can keep abreast of developments in the field of study, find interdisciplinary papers (Small 2010), synthesise what is already known and identify research lacunae. Such techniques also afford an overview of the types of authors engaging in the speciality (Zhao and Strotmann 2014) and a systematic review of an author's works (Chen 2017), helping identify priorities, focus research and clarify the terminology used (Vargas-Quesada et al. 2017). Researchers can thus describe and interpret new problems that may inform their future research strategies and endeavour (Foster et al. 2015). Such

computer techniques are designed to provide an additional point of view, but not to replace the tasks performed by domain experts, which are indispensable.

The paper is organised into sections. The one immediately following describes the data sources and explains the analytical method and technical procedures used. The findings geared to meeting the objectives proposed are presented in "Results" section and discussed in "Discussion" section. The paper ends with the conclusions drawn and perspectives envisaged.

## Materials and methods

The analysis performed here used CiteSpace V free software (Chen 2006), designed over the last 13 years as a comprehensive tool for the analysis, detection and visualisation of patterns and trends emerging in the scientific literature. Its design also draws from theories on the philosophy, history and sociology of science, visual analytics and structural hole (Burt 2004) and information foraging theory (Sandstrom 1994).

The research fronts and intellectual base for a given speciality (Price 1965) are identified primarily by drawing maps based on cluster analysis and reference co-citation (Chen 2004). Clustering is a procedure for classifying items into groups (also called classes or categories) in which the items in a given cluster are more similar to one another in terms of a pre-defined property than they are to the items in other clusters. Co-citation analysis has constituted a de facto standard for mapping science since the nineteen seventies (Boyack and Klavans 2010, p. 2390). In cluster analysis of reference co-citations, two papers are inter-connected when they are simultaneously cited in a third (Small and Griffith 1974; Griffith et al. 1974). Research fronts comprise the co-citing papers in a cluster. The collection of references cited by a given scientific community serves as the intellectual base for the research front. The connections between a research front and its intellectual base are established on the grounds of citations (Braam et al. 1991; Persson 1994).

### Data collection

The data for the study were downloaded on 25 January 2018 from Clarivate Analytics' Web of Science (WoS) Core Collection and more specifically its Science Citation Index Expanded (SCI-EXPANDED). The data were downloaded via online connection to the Web of Science website through the Carlos III University of Madrid's library, using the accesses furnished by the Spanish Foundation for Science and Technology (Spanish initials, Fecyt). WoS downloads are subject to a ceiling of 500 records. The total number was downloaded by successively selecting the 'full record and cited references' option, in 'plain text' format for subsequent analysis and visualisation.

The fields included in the records downloaded were as follows: FN File name; VR Version Number; PT Publication Type; AU: Authors; AF: Authors Full Name; TI Document Type; SO Publication Name; LA Language; DT Document Type; DE Author Keywords; ID Keywords Plus®; AB Abstract; C1 Author Address; RP Reprint Address; EM E-mail Address; RI ResearcherID Number; OI ORCID Identifier; CR Cited References; NR Cited Reference Count; TC Web of Science Core Collection Times Cited Count; Z9 Total Times Cited Count (Web of Science Core Collection, BIOSIS Citation Index, Chinese Science Citation Database, Data Citation Index, Russian Science Citation Index, SciELO Citation Index); U1 Usage Count (Last 180 Days); U2 Usage Count (Since 2013); PU: Publisher; PI

Publisher City; PA Publisher Address; SN International Standard Serial Number (ISSN); EI Electronic International Standard Serial Number (eISSN); J9 29-Character Source Abbreviation; JI ISO Source Abbreviation; PD Publication Date; PY Year Published; VL Volume; IS Issue; BP Beginning Page; EP: Ending Page.

This essential database of selected journals that meet very demanding editorial standards can be used to obtain citation-based quantitative indicators. It presently covers around 12,700 publications (Testa 2009). Pubmed was ruled out as a data source because the lack of information on references cited renders the records unusable for some of the node types analysed here, such as references cited, authors cited, or journals cited.

Given that the aim of this study was to review the global intellectual structure of the most recent epigenetic research, the records were downloaded using a simplified search strategy in which the term 'epigenetics' was defined in the 'Topic' field. Term searches in that field performed simultaneously searches in the fields 'Title', 'Abstract', 'Author Keywords' and 'Keywords Plus' (fed by an algorithm that draws words or phrases frequently appearing in the titles of the articles included in the references but not necessarily in the title or the 'Author Keywords' of the article in question (Garfield and Sher 1993). This search strategy was broad enough to be used in science mapping with the method selected. Other parameters defined for the search strategy included the timeframe (2009–2017), chosen because a prior research note covered a period ending in December 2008, (Clarivate Analytics 2009); document type (articles or reviews only); and language (English).

## Procedures and visualisation

### Research questions 1 and 2

Two procedures were deployed in the generic visualisation of epigenetics categories. Co-occurrences of the topic categories assigned by Web of Science to journals and hence to the papers published in them were used to form a network. Those categories or topics are assigned by WoS based on a number of criteria, including journal relationships based on inter-citations calculated by undisclosed algorithms (Pudovkin and Garfield 2002). Each subject category covers journals addressing the same field of research. Subject categories are defined at the journal, not at the individual paper level. Co-occurrence links interconnect the subject categories of the journals publishing the papers concerned. Coincident inter-category links among journals publishing such papers is an empirical sign of topic affinity. The higher the frequency of coincidence, the greater is the weight of the link in the network of subjects generated. WoS subject categories can also be used to draw scalable topic maps (Leydesdorff and Rafols 2009; Leydesdorff et al. 2013a; Hu and Zhang 2017). This analysis used the categories to which the SCI-Expanded journals publishing papers on epigenetics in the period studied were assigned.

An overall map of citation patterns was drawn for the epigenetic papers compiled, i.e., for both the citing papers and for the papers cited in the dataset. Dual map overlays were used for this purpose (Chen and Leydesdorff 2014).

### Research question 3

The choice of the right method and best design to display representative data and reduce the complexity of the original dataset is a problem routinely faced in science

mapping. Here, as the procedure was intended to detect epigenetic specialities, the method consisted in analysing reference co-citation clusters.

The records were sub-divided sequentially by consecutive years. The criterion for selecting the papers to be represented was a *g*-index (Egghe 2006) threshold value in the yearly records, based on the distribution of citations received by each paper. Given a set of papers ranked in decreasing order of the number of citations received, the *g*-index divides the set into two groups. The index is the largest number such that the top $g$ articles together account for at least $g^2$ citations. The references published in a given year and meeting the *g*-index criterion after applying a constant correction factor, $k=5$, were used to calculate the reference co-citation values. The $k$ value was defined to select the largest number of highly cited papers for the analysis (Costas and Bordons 2008).

The co-citation values for each yearly network were normalised using the cosine index (Leydesdorff 2008). The clusters for all the years were then re-mixed to form a single network covering the entire period and compute the design. The network graphic consisted in nodes (the co-cited papers) and edges (co-citation links). To avoid edge overload, redundant edges were pruned from the network with the Pathfinder algorithm, which retained only the most prominent links (Schvaneveldt et al. 1989). The nodes in the resulting network were distributed spatially with the Kamada–Kawai algorithm (1989). The parameters used to divide the records by years, select records with the *g*-index and constant $k$, prune with Pathfinder and visualise with the Kamada–Kawai algorithm were the same as used for the subject category co-occurrence network.

Non-overlapping spectral clustering was used to generate the co-citation clusters (Ng et al. 2002). Cluster quality was measured with silhouettes (Rousseeuw 1987). Clusters were automatically labelled using the log-likelihood ratio (LLR) (Dunning 1993), applied to the universe of keywords in the citing articles in the respective clusters. Both the keywords defined by the authors and those assigned by the Web of Science were used. Latent semantic indexing (Deerwester et al. 1990), a second labelling method applied, retrieved terms from the titles of the citing articles.

The papers in the co-citation network assigned to each cluster had associated structural and time-based values. Betweenness centrality is a structural measure (Brandes 2001). As nodes with high betweenness values in a co-citation network tend to form bridges inter-connecting different topic communities, they hold the potential to expand the domain's intellectual limits (Chen 2005). A second measure associated with nodes is sigma, a metric that combines the value of a structural measure, betweenness in this case, with a time-related node property, burst strength or inter-pair recognition. It quantifies the co-cited references in terms of scientific originality (Chen et al. 2010).

### Research question 4

Burst detection, the technique used, establishes the frequency with which a given element grows relative to a similar element in a short period of time (Mane and Börner 2004). Here it was used to determine the maximum duration and degree of citation intensity for a given reference, thereby identifying the papers with a sudden increase in the number of citations over a 4-year period. Those references were defined as the landmark papers for the hot topics comprising the intellectual base of the domain studied. Kleinberg's (2002) algorithm was used to detect such bursts.

### Research question 5

One intrinsic difficulty in co-citation analysis is that it fails to detect recently published papers in the co-citation clusters. The potential of a recently published article to alter the intellectual structure of the discipline existing prior to its publication can be determined, however. It is computed from the degree of structural change introduced by the article in the pre-existing intellectual domain by establishing new and unexpected connections that expand its borders (Chen 2012).

In this case, articles with such potential value were identified by calculating structural variation measures: modularity change rate (MCR), cluster linkage (CL) and the change in the distribution of betweenness values relative to the pre-existing distribution in the network in which the new paper is included ($CL_{KL}$). The structural variation values were calculated based on reference networks consisting in a moving window for the 2 years prior to publication of the new article.

## Results

### Basic data statistics

A total of 13,295 bibliographic records (articles and reviews) and 421,342 associated references were retrieved. These two datasets comprised the knowledge space analysed. The number of records downloaded by year of publication is graphed in Fig. 1. Output in papers on epigenetics rose from 649 in 2009 to 1980 in 2017, with 60% of the publications appearing in the last 4 years of the period. Steep inter-annual growth was clearly perceptible throughout the period studied ($R^2 = 0.982$).

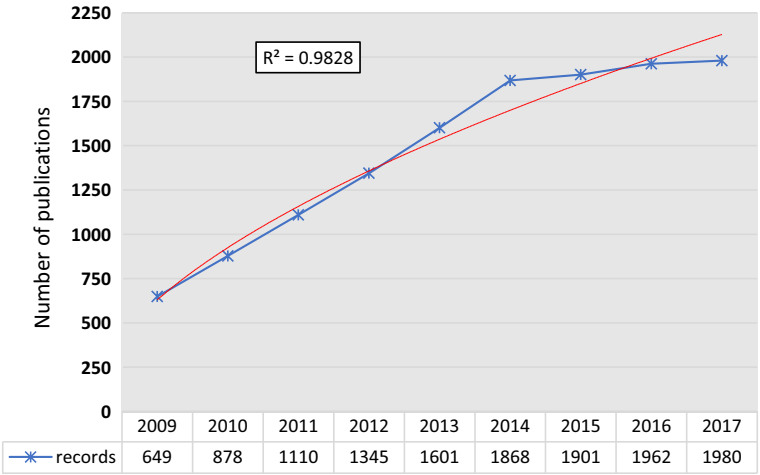


**Fig. 1** Papers on epigenetics retrieved from the Web of Science, 2009–2017 (red line = power trendline). (Color figure online)

### RQ 1 and 2: overview of the core corpus in epigenetics

The topics addressed in epigenetics papers from 2009 to 2017 can be visualised in the form of a pruned network of co-occurring subject categories (Fig. 2). Each category refers to a topic associated with the journals in the WoS database. The analysis identified 178 topics and 250 edges, constituting the largest inter-connected network. The labels are written over the circles in the map. The size of each circle is proportional to the number of papers attributed to the respective category. Node colours denote the range of years of occurrence while purple outlining is an indication of prominent betweenness centrality.

Subject categories represented the highest level of the speciality. Table 1 lists the first 20 Web of Science categories associated with the dataset. The table also gives the cumulative number of papers under the category, the betweenness value and the year of first occurrence in the dataset. The six highest ranking categories were: biochemistry and molecular biology, genetics and heredity, oncology, cell biology, neurosciences and neurology and neurosciences.

Of the top 20 categories, chemistry had the most significant betweenness value (0.18), denoting its influence on the dissemination of ideas in epigenetics.

The map in Fig. 2 and the data in Table 1 identify the most significant topics and elementary relationships in epigenetics, affording non-experts a preliminary view of the subjects generally addressed in the field.

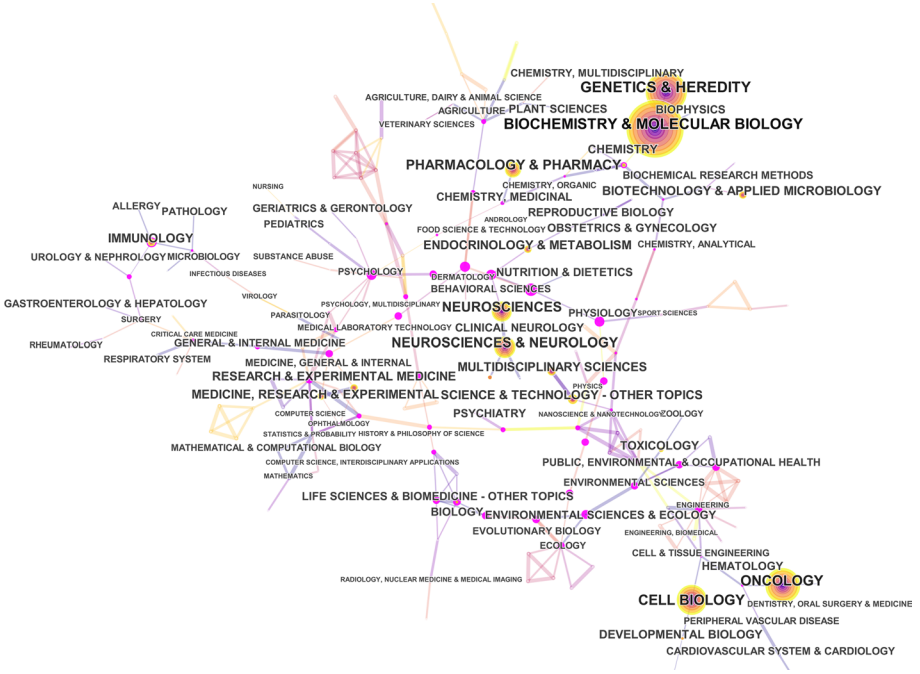


**Fig. 2** Pathfinder-pruned network of co-occurring Web of Science subject categories for epigenetics (nodes = subject categories; edges = co-occurrence of subject categories in papers; label size proportional to node size; only categories with ≥ 24 counts shown)

**Table 1** Top 20 subject categories in the dataset

| Frequency | Centrality | First occurrence | WoS category |
|---|---|---|---|
| 2509 | 0.05 | 2009 | Biochemistry and molecular biology |
| 1864 | 0 | 2009 | Genetics and heredity |
| 1629 | 0.02 | 2009 | Oncology |
| 1483 | 0.04 | 2009 | Cell biology |
| 1140 | 0.05 | 2009 | Neurosciences and neurology |
| 1002 | 0.07 | 2009 | Neurosciences |
| 935 | 0 | 2009 | Pharmacology and pharmacy |
| 626 | 0.02 | 2009 | Science and technology—other topics |
| 590 | 0 | 2009 | Multidisciplinary sciences |
| 577 | 0.07 | 2009 | Endocrinology and metabolism |
| 555 | 0 | 2009 | Biotechnology and applied microbiology |
| 516 | 0.01 | 2009 | Research and experimental medicine |
| 500 | 0.17 | 2009 | Immunology |
| 402 | 0.02 | 2009 | Developmental biology |
| 393 | 0.18 | 2009 | Chemistry |
| 381 | 0.02 | 2009 | Toxicology |
| 337 | 0 | 2009 | Psychiatry |
| 326 | 0 | 2009 | Biophysics |
| 320 | 0.02 | 2009 | Reproductive biology |

The dual map overlay in Fig. 3 shows the papers published from 2009 to 2017 based on an epigenetic topic-based search. It depicts the citation trajectories of the 2536 journals publishing the epigenetic papers analysed in the dataset. It shows the location of the citing journals (citing view, left) and the pathways toward the communities of cited journals (cited view, right). The origins, pathways and destinations are overlain on a basic global map representing the communities of journals listed in the Web of Science (Leydesdorff

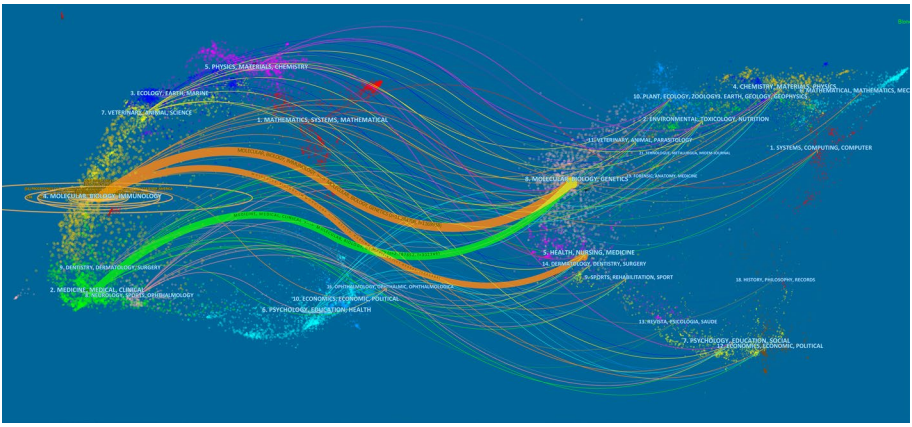


**Fig. 3** Dual map overlay of epigenetics literature based on 2536 journals

et al. 2013b; Blondel et al. 2008). The > 10,000 points on the map that represent the journal communities provide the context for perceiving the citation pattern in epigenetic journals (Chen and Leydesdorff 2014).

In epigenetics, the citing literature was found in primarily two communities of journals. The first was labelled 'Molecular biology, Immunology' (orange line on the left side of the map) and the second 'Medicine, Medical, Clinical' (lime green curve on lower left). The citation curves springing from these two areas on the left flow toward areas on the right, where the cited topics are represented. The orange line is sub-divided in two, indicating that some of the molecular biology and immunology papers cite different groups of journals: one dealing with genetics and the other with heath, nursing and medicine. The main trajectories between source and target citation communities are given in Table 2, sorted by z-score value in descending order. Z-score is distance from the mean of a data point, expressed as the number of standard deviations. Line thickness (Fig. 3) is proportional to the z-score value.

### RQ 3: landscape of epigenetic specialities

The network of co-cited paper clusters, representing inter-related specialities from 2009 to 2017, is shown in Fig. 4. The 1135 references meeting the modified ($k=5$) $g$-index criterion were retrieved to create a network for each year, which was subsequently condensed. The condensed network contained 425 references and after pruning, 858 edges. The component with the largest number of edges contained 414 references. The network had ten clusters and at 0.704, a high modularity, indicating that the epigenetic specialities described in this study were well defined in terms of the features of the co-citation clusters. The colours in Fig. 4 signify the date when the co-citation links were established. The yellow areas were older (#0, #1, #2, #3) than the pink, which were older than the red (#4 and #5). The cluster labels in Fig. 4 were drawn from the keywords in the citing articles using log likelihood ratio techniques.

### RQ 3: timeline

A timeline of yearly fluctuations in research specialities is reproduced in Fig. 5. Here each cluster of references is positioned on a horizontal line, labelled on the right. The clusters are ranked by number of references in descending order. The coloured curves are citation links appearing in the year represented by the colour at issue (more recent = ochre and yellow, less recent = purple and violet). The references are shown as circles, with the colours identifying structural properties. Red nodes contain references with high burst values while violet outlining denotes high betweenness values.

**Table 2** Subject areas of citing and cited journals in epigenetics

| Areas, citing journals | Areas, cited journals | z-score |
|---|---|---|
| Molecular biology, immunology | Molecular biology, genetics | 11.492 |
| Medicine, medical, clinical | Molecular biology, genetics | 3.326 |
| Molecular biology, immunology | Health, nursing, medicine | 1.978 |

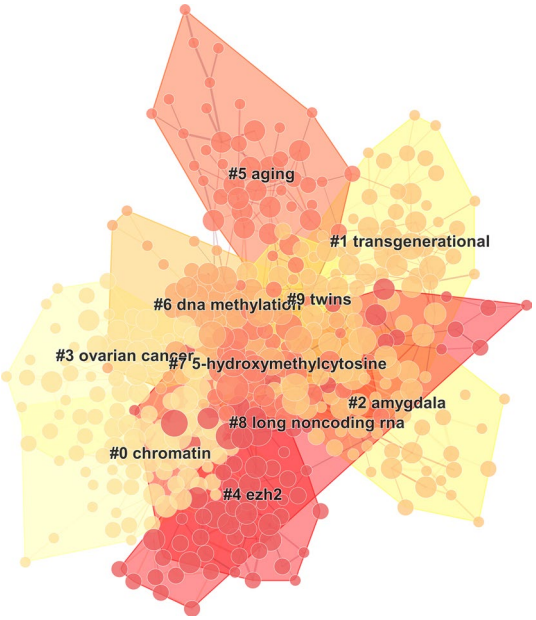


**Fig. 4** *g*-index-generated landscape of the core network of co-cited references in epigenetics, 2009–2017 (labels drawn from keywords in citing papers using the log-likelihood ratio)

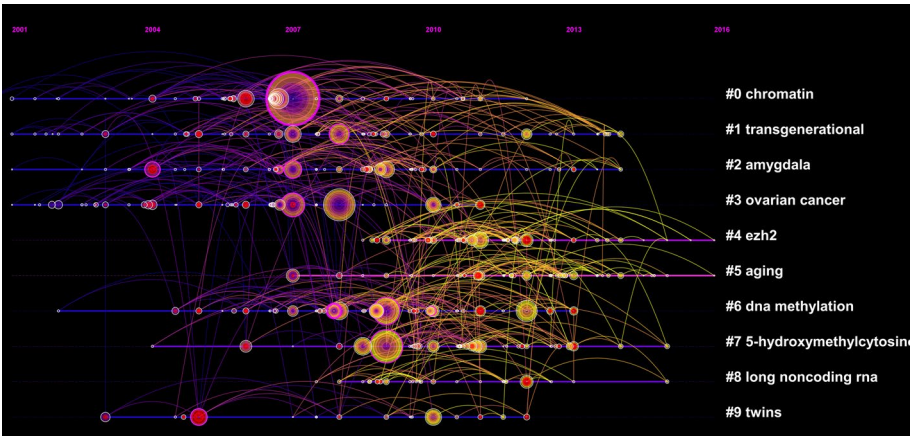


**Fig. 5** Horizontal timeline of research specialities (labels drawn from keywords in citing papers using the log-likelihood ratio). (Color figure online)

The nodes combining the two features are deemed to be of particular interest. The most prominent papers have a larger diameter, in keeping with their higher co-occurrence rate.

Two characteristics are instrumental to understanding the clusters. The first is duration and the second ongoing activity. The graph identified the longest lasting specialities to be #0 chromatin, #1 transgenerational, #2 amygdala and #3 ovarian cancer, all of which are presently inactive, after 11 (#0), 13 (1# and #2) or 10 (#3) years of activity. Specialities #4 (ezh2), #5 (ageing), #7 (5-hydroxymethylcytosine) and #8 (long

noncoding RNA) were found to be the most recent and active, as attested to by the inter- and intra-citation arcs connecting the clusters and the existence of red circles, denoting citation bursts.

Speciality details are listed in Table 3, where cluster quality is measured in terms of the mean silhouette value. The higher that value, the more distinctly defined is the cluster. Cluster #1, for instance, had a silhouette value of 0.845, denoting high definition, although all the clusters exhibited sufficiently high values. Two procedures were used to label each cluster: (a) retrieval of keywords from the citing articles using the log likelihood ratio (LLR); and (b) retrieval of terms contained in the citing articles with latent semantic indexing. The numerical value for each label denotes either the likelihood ratio or statistical significance. The mean year is the average of the publication dates for the references comprising the cluster.

## RQ 3: key specialities

Cluster #0, labelled 'chromatin' or 'chromatin organization', was the largest and had a silhouette value of 0.79. Its 56 papers dated from 2004 to 2011, with a peak in 2005–2007 (Table 4). All were published in major scientific journals, such as *Nature* (or its specialised titles dealing with cell and molecular biology) and *Cell*. This cluster consequently included the foremost reviews and seminal papers that have defined the modern epigenetic landscape.

The following clusters by size, which varied only narrowly among them, were #1 to #5. They addressed topics characteristic of epigenetic research, such as the effect of the environment on and the potential heritability of epigenetic changes (transgenerational epigenetics), cancer and the regulation of epigenetic marks involved in cancer growth and/or treatment (ovarian cancer/ezh2) and ageing. The remaining clusters dealt with more recent concepts such as the inclusion of hydroxymethylcytosine as a new epigenetic mark (see next section) or the role of non-coding RNA in epigenetic mechanisms. Clusters #1, #4 and #5 were of particular interest, given the topics addressed.

Cluster #1, as the set of core references that established the grounds for epigenetic mark heritability, contained frequently cited papers dating from early domain development. The paper by Anway (2005), which headed the list with a frequency of 139 and a $\sigma$ value of 6.11, merits special mention. It was one of the first articles on the alterations induced by toxic compounds in the epigenomes of not one but several generations of laboratory animals. Although like other M. K. Skinner group studies (some of which were withdrawn and many others found to be non-replicable) it was controversial, it was among the first to draw attention to the possible transgenerational effects of pesticides on animals. Given the controversy surrounding the paper, whether the high citation indexes can be attributed to the influence of the research itself or to the refutation of the results is difficult to ascertain.

Most of the other papers in the cluster also analysed epigenetic marks in the germline. A paper by Hammoud et al. (2009) containing a thorough description of methylation marks in histones located at specific development-related loci wielded substantial influence (Table 5). Earlier articles in the same cluster had already established the fundamentals for understanding what would years later be called 'transgenerational epigenetics', however: the two most prominent were authored by Morgan et al. (2005) and Waterland and Jirtle (2003), which also probed to be a high burst study. Several papers published in 2005–2010 dealt with the role of environmental modifications in mouse epigenetic reprogramming able to affect the observable phenotype or even drive metabolic change. A paper by Ng

**Table 3** Largest 10 clusters of co-cited references in a 1135 reference network

| Cluster ID | Size | Silhouette | Labels from keywords (LLR)[a] | Labels from titles (LSI)[b] | Year |
|---|---|---|---|---|---|
| 0 | 56 | 0.79 | Chromatin (179.59, 1.0E−4) | Chromatin organisation (10932.17, 1.0E−4) | 2006 |
| 1 | 50 | 0.845 | Transgenerational (131.92, 1.0E−4) | Environmental epigenetics (6243.32, 1.0E−4) | 2008 |
| 2 | 48 | 0.751 | Amygdala (177.76, 1.0E−4) | Epigenetic mechanism (26397.14, 1.0E−4) | 2008 |
| 3 | 47 | 0.806 | Ovarian cancer (122.72, 1.0E−4) | Cancer epigenetics (8215.7, 1.0E−4) | 2005 |
| 4 | 45 | 0.83 | ezh2 (120.26, 1.0E−4), | Cancer epigenome (9231.96, 1.0E−4) | 2011 |
| 5 | 45 | 0.825 | ageing (227.22, 1.0E−4), | age-related disease (8884.97, 1.0E−4) | 2012 |
| 6 | 41 | 0.81 | DNA methylation (52.57, 1.0E−4) | genome-wide DNA methylation analysis (8502.53, 1.0E−4) | 2008 |
| 7 | 39 | 0.728 | 5-Hydroxymethylcytosine (187.48, 1.0E−4) | Embryonic stem cell (10,794.37, 1.0E−4) | 2010 |
| 8 | 26 | 0.891 | Long noncoding RNA (164.3, 1.0E−4) | Noncoding RNA (12,359.05, 1.0E−4) | 2010 |
| 9 | 17 | 0.821 | Twins (32.85, 1.0E−4) | Inflammatory bowel disease (1754.28, 1.0E−4) | 2008 |

[a]Log-likelihood ratio

[b]Latent semantic indexing

**Table 4** Key constituents of cluster #0 by sigma value

| Frequency | Burst | Centrality | Sigma | Authors | Year | Source |
|---|---|---|---|---|---|---|
| 252 | 27.95 | 0.07 | 7.45 | Bernstein | 2006 | Cell |
| 119 | 30.94 | 0.07 | 7.06 | Shi | 2004 | Cell |
| 90 | 19.61 | 0.07 | 4.04 | Boyer | 2006 | Nature |
| 224 | 10.56 | 0.07 | 2.01 | Mikkelsen | 2007 | Nature |
| 144 | 17.23 | 0.04 | 1.99 | Reik | 2007 | Nature |

**Table 5** Key constituents of cluster #1 by sigma value

| Frequency | Burst | Centrality | Sigma | Authors | Year | Source |
|---|---|---|---|---|---|---|
| 139 | 30.36 | 0.06 | 6.11 | Anway | 2005 | Science |
| 110 | 43.62 | 0.03 | 3.28 | Waterland | 2003 | Mol Cell Biol |
| 72 | 18.68 | 0.06 | 2.82 | Morgan | 2005 | Hum Mol Genet |
| 65 | 16.23 | 0.02 | 1.31 | Hammoud | 2009 | Nature |
| 111 | 14.21 | 0.01 | 1.15 | Dolinoy | 2006 | Environ Health Persp |
| 89 | 9.72 | 0.01 | 1.1 | Ng | 2010 | Nature |
| 78 | 20.24 | 0 | 1.08 | Lillycrop | 2005 | J Nutr |
| 133 | 10.37 | 0 | 1.04 | Dolinoy | 2007 | P Natl Acad Sci Usa |

**Table 6** Citing papers in cluster #1

| Coverage | References |
|---|---|
| 0.16 | Bernal and Jirtle (2010). Epigenomic disruption: the effects of early developmental exposures |
| 0.16 | Li et al. (2010). Epigenetic programming by maternal nutrition: shaping future generations |
| 0.16 | Rosenfeld (2010). Animal models to study environmental epigenetics |
| 0.16 | Zama and Uzumcu (2010). Epigenetic effects of endocrine-disrupting chemicals on female reproduction: an ovarian perspective |
| 0.14 | LeBaron et al. (2010). Epigenetics and chemical safety assessment |

et al. (2010) had a considerable impact, serving as the grounds for a wealth of literature on metabolic impairment and obesity in offspring attributable to parental nutrition. This line of research with laboratory animals was supplemented with observational studies in human populations that are still under debate.

The relationship between germline epigenetic marks and the environmentally-mediated changes subsequently transmitted to the offspring is presently an epigenetic hot topic. Unsurprisingly, then, the citing articles constituting the research front in this cluster (Table 6) were reviews addressing the effect of environmental factors on embryonic development, in terms of both a healthy environment and nutrition. Such articles contribute to a deeper understanding of area dynamics. The first article listed in Table 6, for instance, cited 16% of the references in cluster #1.

Clusters #4 and #5, with 45 papers each, were particularly relevant since they included references on the global analysis of the epigenome from two perspectives: cancer development and treatment and ageing and age-related disease (Table 3). These continue to be hot topics in the field, for they are pivotal to many aspects of epigenetic

**Table 7** Key constituents of cluster #4 by sigma value

| Frequency | Burst | Centrality | Sigma | Authors | Year | Source |
|---|---|---|---|---|---|---|
| 210 | 27.46 | 0.08 | 8.99 | Dawson | 2012 | Cell |
| 96 | 16.32 | 0.06 | 2.69 | Xu | 2011 | Cancer Cell |
| 94 | 15.97 | 0.05 | 2.14 | Figueroa | 2010 | Cancer Cell |
| 114 | 4.06 | 0.09 | 1.42 | Chi | 2010 | Nat Rev Cancer |
| 78 | 13.25 | 0.03 | 1.39 | Kaelin | 2013 | Cell |
| 80 | 6.77 | 0.02 | 1.12 | Choudhary | 2009 | Science |

**Table 8** Citing papers in cluster #4

| Coverage | References |
|---|---|
| 0.2 | Roy et al. (2014). Driver mutations of cancer epigenomes |
| 0.13 | Fong et al. (2014). Epigenetics in the hematologic malignancies |
| 0.13 | Paluch et al. (2016). Epigenetics: a primer for clinicians |
| 0.11 | Baylin and Jones (2016). Epigenetic determinants of cancer |
| 0.11 | Hoeksema and de Winther (2016). Epigenetic regulation of monocyte and macrophage function |
| 0.11 | Pfeifer et al. (2014). The role of 5-hydroxymethylcytosine in human cancer |
| 0.11 | Schoofs et al. (2014). Origins of aberrant DNA methylation in acute myeloid leukemia |
| 0.11 | Yang et al. (2016). Natural compound-derived epigenetic regulators targeting epigenetic readers, writers and erasers |

research. Enhancer of zeste homolog 2 (ezh2), an enzyme directly involved in the methylation of histone proteins that has been a research target in the pharmacological treatment of cancer, has attracted sufficient scientific interest to serve as a label for an entire cluster.

Cluster #4 included studies specifically dealing with human cancer, as well as two more general papers: one by Choudhary et al. (2009) describing the extra-epigenetic functional significance of lysine acetylation, and the other by Kaelin and McKnight (2013), reviewing the relationships between metabolic impairment and disease development in general (Table 7). The presence of the latter in such a cancer-intensive cluster attested to the prominence of cancer in epigenetic research.

The 'burstiest' paper in the cluster was Dawson and Kouzarides's (2012), an indication that by 2012 a sizeable corpus of evidence had been gathered on the mechanisms underlying genetic and epigenetic aberrations in cancer. One of the most frequently cited reviews, it drew heavily from other papers in the cluster, such as the ones by Xu et al. (2011) and Figueroa et al. (2010). Both these authors described experimental evidence of epigenetic alterations in human cancer cells and their mechanistic relationship to the malignant properties of such cells.

All the citing articles in cluster #4 were reviews. While dealing very specifically with cancer genetics and epigenetics, given the breadth and depth of the field, they were very varied, with some confined to more technical issues. Pfeifer et al. (2014), for instance, discussed specific, recently described epigenetic marks, while Schoofs et al. (2014) reviewed the state of play of DNA methylation and acute myeloid leukaemia (Table 8).

**Table 9** Key constituents of cluster #5 by sigma value

| Frequency | Burst | Centrality | Sigma | Authors | Year | Source |
|---|---|---|---|---|---|---|
| 91 | 15.55 | 0.09 | 3.97 | Bjornsson | 2008 | Jama-J Am Med Assoc |
| 134 | 22.81 | 0.04 | 2.4 | Bibikova | 2011 | Genomics |
| 82 | 13.93 | 0.06 | 2.17 | Liu | 2013 | Nat Biotechnol |
| 59 | 14.73 | 0.05 | 2 | Bell | 2011 | Genome Biol |
| 144 | 12.73 | 0.05 | 1.89 | Rakyan | 2011 | Nat Rev Genet |
| 204 | 15.45 | 0.02 | 1.31 | Feinberg | 2007 | Nature |
| 74 | 18.49 | 0 | 1.05 | Joubert | 2012 | Environ Health Persp |

**Table 10** Citing papers in cluster #5

| Coverage | References |
|---|---|
| 0.2 | Gassen et al. (2017). Life stress, glucocorticoid signaling, and the aging epigenome: implications for aging-related diseases |
| 0.13 | Ali et al. (2015). An epigenetic map of age-associated autosomal loci in northern European families at high risk for the metabolic syndrome |
| 0.13 | Gensous et al. (2017). The epigenetic landscape of age-related diseases: the geroscience perspective |
| 0.13 | Hodgson et al. (2017). Epigenetic age acceleration assessed with human white-matter images |
| 0.13 | Morris and Beck (2015). Analysis pipelines and packages for infinium humanmethylation450 bead chip (450 k) data |
| 0.13 | Ong et al. (2015). Measuring epigenetics as the mediator of gene/environment interactions in Dohad |
| 0.13 | Tejedor and Fraga (2017). Interindividual epigenetic variability: sound or noise? |

Cluster #5, in turn, under the labels ‘ageing’ or ‘age-related disease’, contained a varied set of papers ranging from the technical analysis of epigenetic marks (a topic highly interconnected with all the other clusters), such as the article by Bibikova et al. (2011), to others with a narrower perspective, including a discussion of the effect of smoking during pregnancy (Joubert et al. 2012) or of the relationship between DNA methylation and the genetic risk of developing rheumatoid arthritis (Liu et al. 2013). This cluster also included a series of more general studies on the role of methylation and other states of the epigenome in disease development in general. The papers by Bell et al. (2011), Feinberg (2007) and Rakyan et al. (2011) lay in that group (Table 9).

The ‘burstiest’ of all these papers was Bibikova et al. (2011), in all likelihood because of the prominence of the technology described, which is requisite to any genome-wide analysis of the importance of the epigenome in disease. Joubert et al. (2012) authored the second highest burst paper and Bjornsson et al. (2008) the third, which was also the one most specifically concerned with the relationship between ageing and DNA methylation. It pioneered research on the correlation between the passage of time and epigenetic changes, analysed in DNA samples from several families over the course of 11 to 16 years.

The citing articles in this cluster were the most recent, denoting the keen research interest in ageing and age-related disorders attendant upon longer life expectancy in modern societies (Table 10). Since epigenetic marks are affected by environmental exposure and

their flexibility makes them susceptible to profound change over the years, these papers will continue to be cited in research on diseases differing in their particulars, but sharing an aetiology based on flawed epigenetic regulation or epigenetic deregulation. Those considerations may explain the greater heterogeneity of this than the other clusters. The citing articles identified dealt with topics such as the technical improvement of sequencing technologies (Morris and Beck 2015), nutritional and other types of environmental stress in connection with ageing (Gassen et al. 2017) or reviews, one of which even coined a new term for the epigenetic study of age-related diseases: geroscience (Gensous et al. 2017).

This discussion of the findings would be incomplete without a specific section on the highest burst papers in the clusters identified, which are summarised below.

## RQ 4: citation bursts

Burst is defined to mean a sudden and significant rise in the number of times a given article is cited in a short period of time. Burst analysis identifies waves of interest based on the number of citations received by a given paper within a dataset. The technique was deployed here to retrieve the references in the total dataset (421,342) exhibiting with a minimum duration of 4 year burst in the period studied (2009–2017), irrespective of the date of

**Table 11** Top 23 references with minimum duration 4-year bursts, by year of publication. (Color table online)

| References[a] | Year | Strength | From | To | 2009–2017 |
|---|---|---|---|---|---|
| Gardiner-Garden | 1987 | 22.8542 | 2009 | 2013 | |
| Frommer | 1992 | 16.0762 | 2009 | 2013 | |
| Wolff | 1998 | 15.2505 | 2009 | 2012 | |
| Reik | 2001 | 22.2148 | 2010 | 2013 | |
| Esteller | 2001 | 20.1715 | 2009 | 2013 | |
| Takai | 2002 | 20.2541 | 2009 | 2012 | |
| Gaudet | 2003 | 22.6103 | 2009 | 2013 | |
| Lillycrop | 2005 | 22.2569 | 2009 | 2012 | |
| Feinberg | 2006 | 28.845 | 2009 | 2012 | |
| Eckhardt | 2006 | 24.8067 | 2009 | 2013 | |
| Lee | 2006 | 24.2608 | 2009 | 2012 | |
| Boyer | 2006 | 21.8785 | 2009 | 2013 | |
| Dolinoy | 2006 | 17.0918 | 2009 | 2012 | |
| Li | 2007 | 16.8835 | 2009 | 2012 | |
| Métivier | 2008 | 25.6927 | 2009 | 2012 | |
| Guan | 2009 | 10.6792 | 2011 | 2014 | |
| Peleg | 2010 | 18.6388 | 2011 | 2014 | |
| Chi | 2010 | 18.4022 | 2011 | 2014 | |
| Ernst | 2011 | 17.3661 | 2013 | 2017 | |
| Hanahan | 2011 | 16.9844 | 2013 | 2017 | |
| Guo | 2011 | 16.7932 | 2012 | 2015 | |
| Rakyan | 2011 | 13.404 | 2012 | 2015 | |
| Dunham | 2012 | 22.6505 | 2013 | 2017 | |

[a]Only the first author's name is listed

publication. The 23 references with the highest burst values are listed in Table 11. The blue bars in the final column indicate the period that covers the burst analysis and the red bars the years when the strongest bursts occurred.

The highest burst papers were closely correlated to the landmark studies that defined epigenetics as a clearly identifiable discipline. An article by Gardiner-Garden and Frommer (1987) had a very high burst score for 5 years, unsurprisingly in light of the subject. The authors conducted an in-depth analysis of the presence of CpG islands in vertebrate genomes, suggesting that their specific location within the genome, the possibility of methylation, and the implications for gene expression were inter-related. At this writing, that seminal study has received 2248 citations. Just 5 years later, Frommer et al. (1992) headed research in which he and his colleagues proposed a specific sequencing protocol to analyse the methylation pattern of individual DNA strands. That article, likewise widely cited (1972 citations), constituted a technological breakthrough, describing a procedure that came into general use for much epigenetics-based research. It comes as no surprise, then, that these very early seminal works have accumulated the highest citation scores over long periods, or remained highly influential long after the decade when they were published, for they contain the background for further intellectual and technological developments in epigenetics.

The second highest burst papers, which appeared in 2006 and 2008, were all closely related to the biological significance of methylation patterns, genetic regulation and disease. In a decisive paper published in *Nature Reviews Genetics*, Feinberg et al. (2006) reviewed the relationship between the epigenetic alteration of genes and stem cell properties as a key to cancer progression. That review has received 1149 citations at this writing, 889 from papers on genetic heredity, and unsurprisingly has the highest strength burst value. It also directly connected to two other bursty studies identified in the present analysis. Eckhardt et al. (2006) described the high stability of DNA methylation marks, whilst Lee et al. (2006) explored a side issue: genetic control of embryonic stem cells and their developmental potential, knowledge of which is needed to correlate cell differentiation to methylation patterns in genetic regions. In the latter part of that early period of seminal publications, Metivier et al. (2008) broke the ground for research on the highly dynamic nature of epigenetic marks and their regulatory processes, with insights that led to the establishment of the close relationship between epigenetic gene regulation and environmental factors. The most recently published high burst paper was authored by Dunham and Kundaje (2012), who described the ENCODE project findings. One of the latest landmarks in genomic research, this review contains a comprehensive summary of the functional features of the human genome. Although not strictly confined to epigenetics, it nonetheless furnishes vital information for research in the field and has had a heavy impact on the papers published in 2013–2017. It has received 5631 citations to date. That statistic both shows that the definition of epigenetics as an entity separate from the rest of genetics and genomics is sometimes hazy and attests to the close interconnection among disciplines.

## RQ 5: articles with transformational potential

Structural variation theory is used to determine the degree of innovation of the ideas in scientific papers. Once a paper is published and assigned to a specific scientific domain, its innovation and transformational potential is measured against the pre-existing intellectual structure. Papers with transformational potential are understood to be studies containing innovations not previously established in the respective list of references and the network of

references cited contained therein, which constitute the knowledge domain. Those papers and the innovation-related links between the references they include connect the clusters comprising a domain and transform the former intellectual base at issue (Chen 2012).

The findings for the dataset used here are given in Table 12, which lists three articles per year with such potential. The selection criterion was the geometric mean of the values for three separate metrics calculated for all the articles published in each year. In 2017, for instance, the three most prominent articles in this respect were related to 5-hydroxymethylcytosine. The article with the highest score, published in 2011, reviewed the techniques for epigenetic studies and their applications in characterising human disease (Martin-Subero and Esteller 2011).

A number of studies published in 2011–2017 hold the potential to define the future direction of epigenetic research. The topic of most was epigenetics and disease, with an emphasis on the mechanisms specifically involved in methylation. The present survey also revealed a tendency to confine research to three major areas: cancer, ageing, and neurobiology, with a few sporadic studies on the relationship between epigenetic regulation and evolution (Duncan et al. 2014), physical performance during exercise and athletic potential (Ehlert et al. 2013) and rheumatic disease (Zufferey et al. 2014). Even those seemingly divergent topics nonetheless dealt with the epigenetic regulation of methylation, a subject that continued to be of keen interest even as late as 2017, as inferred by the studies by López et al. (2017), Mellen et al. (2017) and Skvortsova et al. (2017). Those very recent reviews constituted a turning point in the analysis of epigenetic marks, condensing the growing evidence that 5-hydroxymethylcytosine (5hmC), formerly assumed to be an oxidation product of widely studied 5-methylcytosine, should be regarded as a new epigenetic mark. Drawing from the existing evidence of the relationship between methylation patterns and disease, richly supplemented by papers in this same cluster (Herceg and Vaissiere 2011; Rodriguez-Paredes and Esteller 2011; Liu 2012; Jankovska et al. 2015; Roos et al. 2016), they provide a substantial body of knowledge for future research. They will serve as the grounds for evaluating the existence of further levels of epigenetic regulation with implications of great significance for neurobiology (Mellen et al. 2017) and age-related diseases (López et al. 2017). Skvortsova et al. (2017), in turn, called for caution when interpreting earlier findings on methylcytosine levels, contending that they may have led to underestimating the presence of 5-hydroxymethylcytosine. In light of its methodological contribution, this article will probably continue to be highly cited in the years to come, if the number of papers on 5hmC research grows at the present rate. Last but not least, Allis and Jenuwein (2016) authored one of the most accurate and thorough reviews of epigenetic processes to date, which may also continue to be widely cited.

## Discussion

Experts engaging in a given field of research can conduct systematic reviews of the literature published in the area. That endeavour can be enhanced by using advanced bibliometric methods to select references based on cumulative citation data indicative of the attention aroused in the scientific community. Such a procedure eludes the exclusive reliance on approaches inevitably biased (Lee et al. 2013) by the criteria applied by those engaging in evaluation. If references are selected using a mix of techniques based on scalable citation graphics and lexical procedures, they can be used to explore the knowledge structure of any

**Table 12** Potentially transformative papers published in recent years (2009–2017)

| Year | ΔM[a] | ΔCLw[b] | ΔCKL[c] | Geometric mean | References | Title |
|---|---|---|---|---|---|---|
| 2017 | 10.11 | 0.1 | 0.05 | 0.37 | López et al. (2017) | The role of 5-hydroxymethylcytosine in development, aging and age-related iseases. |
| 2017 | 4.28 | 0.14 | 0.04 | 0.29 | Mellen et al. (2017) | 5-Hydroxymethylcytosine accumulation in postmitotic neurons results in functional demethylation of expressed genes |
| 2017 | 3.98 | 0.09 | 0.05 | 0.26 | Skvortsova et al. (2017) | Comprehensive evaluation of genome-wide 5-hydroxymethylcytosine profiling approaches in human DNA |
| 2016 | 6 | 0.07 | 0.11 | 0.36 | Allis and Jenuwein (2016) | The molecular hallmarks of epigenetic control |
| 2016 | 4.15 | 0.13 | 0.02 | 0.22 | Roos et al. (2016) | Integrative DNA methylome analysis of pan-cancer biomarkers in cancer discordant monozygotic twin-pairs |
| 2016 | 4.63 | 0.05 | 0.04 | 0.21 | Brazel and Vernimmen (2016) | The complexity of epigenetic diseases |
| 2015 | 7.66 | 0.09 | 0.11 | 0.42 | Jones et al. (2015) | DNA methylation and healthy human aging |
| 2015 | 8.61 | 0.08 | 0.09 | 0.40 | He and Ecker (2015) | Non-CG Methylation in the Human Genome |
| 2015 | 8.81 | 0.08 | 0.04 | 0.30 | Jankovska et al. (2015) | The potential of DNA modifications as biomarkers and therapeutic targets in oncology |
| 2014 | 11.49 | 0.1 | 0.09 | 0.47 | Duncan et al. (2014) | Epigenetics, plasticity, and evolution: how do we link epigenetic change to phenotype |
| 2014 | 14.72 | 0.1 | 0.06 | 0.45 | Yao and Jin (2014) | Unlocking epigenetic codes in neurogenesis |
| 2014 | 8.8 | 0.15 | 0.06 | 0.43 | Zufferey et al. (2014) | Epigenetics and methylation in the rheumatic diseases |
| 2013 | 10.08 | 0.18 | 0.02 | 0.33 | Watanabe et al. (2013) | Epigenetic regulation in pluripotent stem cells: a key to breaking the epigenetic barrier |
| 2013 | 7 | 0.1 | 0.03 | 0.28 | Ehlert et al. (2013) | Epigenetics in sports |
| 2013 | 6.8 | 0.1 | 0.01 | 0.19 | Moore et al. (2013) | DNA methylation and its basic function |
| 2012 | 9.05 | 0.33 | 0.05 | 0.53 | Day and Sweatt (2012) | Epigenetic treatments for cognitive impairments |
| 2012 | 7.6 | 0.2 | 0.02 | 0.31 | Liu (2012) | Epigenetics advancing personalized nanomedicine in cancer therapy |
| 2012 | 4.26 | 0.13 | 0.04 | 0.28 | Kobow and Blümcke (2012) | The emerging role of DNA methylation in epileptogenesis |
| 2011 | 26.97 | 2.25 | 0.16 | 2.13 | Martin-Subero and Esteller (2011) | Profiling epigenetic alterations in disease |
| 2011 | 27.37 | 1.56 | 0.13 | 1.77 | Rodriguez-Paredes and Esteller (2011) | Cancer epigenetics reaches mainstream oncology |

**Table 12** (continued)

| Year | ΔM[a] | ΔCLw[b] | ΔCKL[c] | Geometric mean | References | Title |
|---|---|---|---|---|---|---|
| 2011 | 22.85 | 1 | 0.16 | 1.54 | Herceg and Vaissiere (2011) | Epigenetic mechanisms and cancer: an interface between the environment and the genome |

[a]Modularity change rate

[b]Cluster linkage

[c]Centrality divergence. Only the first author's name is listed

field of research. The purpose here was to review the recent literature on epigenetics with sufficient breadth and depth to approximately determine domain structure.

The study consisted in building and assembling a number of units. The first addressed scientific output in epigenetics. The papers identified and included in this stage of the analysis mirrored the growing interest of researchers and teams of researchers. The analytical elements applied are summarised below.

- RQ1 and RQ2 Analysis of the co-occurrence of the subject categories assigned to journals was followed by citation overlay mapping, both to delimit and visualise the field of research, initially to acquire what might be termed an 'overview'.
- RQ3 Domain specialities were identified through co-citation analysis to determine the core articles constituting the intellectual base of the discipline. Cluster quality was measured with silhouettes.
- RQ3 The most recent papers on the speciality that comprise the research front were found by identifying the main citing papers through co-citation analysis. The papers included in each cluster were characterised with four metrics, namely co-citation links, citation bursts, betweenness and sigma, a hybrid indicator combining betweenness and burst.
- RQ3 The specialities comprising the intellectual base were identified by selecting terms based on occurrence patterns in natural language (titles and keywords) using latent semantic indexing, a semantic distributivity technique, and the log-likelihood ratio, a non-parametric test.
- RQ4 The smallest unit defining granularity or degree of detail was the individual article. The first technique used detected papers with citation bursts lasting 4 consecutive years. That retrospective selection highlighted the papers of key interest to the research community that were the object of sudden, intense and sustained attention throughout the period studied.
- RQ5 The structural and time-related properties of articles with transformational potential in the field of epigenetics were also analysed.

Certain limitations to the methodology used here merit mention. Firstly, the findings are the outcome of a given search strategy applied to literature listed in a specific database. The paper must consequently be viewed as an exploratory study in which the results are applicable to a limited context. Establishing the bounds of a subject area, discipline or speciality is clearly challenging (Chen 2016) and any analysis aiming to identify the intellectual structure of a domain is sensitive to the values of the many variables involved. A different search strategy applied to the same database or the same strategy applied to a different database would obviously have yielded different results. No universal scientific database with all the information systematically organised is in place, nor has the community reached a consensus about which is the best. Nonetheless, studies comparing citation databases deem that the one chosen here, WoS, is applicable to any study based on network paradigms, given its consistency (Ŝubelj et al. 2015).

In this study the specialities comprising the field addressed were delimited using non-overlapping spectral clustering with CiteSpace co-citation analysis software. The authors are aware of other approaches for defining specialities using clustering techniques (Ŝubelj et al. 2016; Velden et al. 2017) or different inter-citation relationships, such as direct citation or bibliographic coupling (Klavans and Boyack 2017). No golden rule has been defined for clustering applicable to the data analysed. The authors therefore support the approach defined by Gläser et al. (2017), who contend that there is no single way to identify topics

and that more than one subject structure can be defined, depending on the perspective from which a given area of knowledge is viewed.

Analogous remarks can be applied to cluster composition. In this study, the clusters of co-cited papers contained both bursty and high betweenness value articles. The papers with the highest sigma value contained innovative discoveries of key significance for the intellectual base supporting epigenetics. These are not, however, the sole topological metrics for determining significance in a field of research. Shibata et al. (2009), for instance, used the density of the core paper clusters for that purpose.

## Conclusion

Several conclusions of interest can be drawn from the present analysis of the scientific literature on epigenetics. The very short number of studies on its biotechnological applications to agriculture and animals attests to the ongoing predominance of biomedical research, in terms of the number and impact of papers. The popularity of this discipline among scientists, attributable to the association it defines between environmental factors and the inheritance of health- and disease-related traits, is based primarily on a series of groundbreaking papers that vested it with a sound body of knowledge. More recent studies are targeting a deeper understanding of the molecular features of epigenetic marks in the context of biomarker design and disease monitoring. The focus is also shifting to the implications of epigenetics for ageing and age-related pathologies in humans. Of particular note is the fact that most of the research rests on the epigenetic process sensu stricto: DNA methylation and chromatin remodelling. The role of histone modifications or noncoding RNAs, although present in many of the studies identified in the analysis, constitute a minority perspective. Interestingly, whilst the definition of noncoding RNA as an epigenetic mark has been controversial, this exercise confirms that most authors still deem research based on noncoding RNA, especially long noncoding RNA (lncRNA) and microRNA (miRNA), to be related to the epigenetic regulation of gene expression. All-in-all, the present findings help understand and structure the complex landscape of epigenetics research over the last few decades.

Future research may target a number of horizons and the analyses conducted may use direct rather than the indirect or co-citation approach adopted here. Different clustering algorithms might also be applied. Another feature that might be addressed is co-word networks and fluctuations in domain topics.

**Acknowledgements** The authors wish to thank Margaret Clark, translator, for her linguistic support.

## References

Agrawal, K., Das, V., Vyas, P., & Hajduch, M. (2018). Nucleosidic DNA demethylating epigenetic drugs—A comprehensive review from discovery to clinic. *Pharmacology & therapeutics. pii: S0163, 7258*(18), 30031–30037.

Ali, O., Cerjak, D., Kent, J. W., Jr., James, R., Blangero, J., Carless, M. A., et al. (2015). An epigenetic map of age-associated autosomal loci in northern European families at high risk for the metabolic syndrome. *Clinical Epigenetics, 7*(12), 1–18.

Allis, C. D., & Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. *Nature Reviews Genetics, 18*(8), 487–500.

Anway, M. D. (2005). Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science, 308*(5727), 1466–1469.
Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology, 8*(1), 19–32.
Bai, Z. T., Bai, B., Zhu, J., Di, C. X., Li, X., & Zhou, W. C. (2018). Epigenetic actions of environmental factors and promising drugs for cancer therapy. *Oncology Letters, 15,* 2049–2056.
Baylin, S. B., & Jones, P. A. (2016). Epigenetic determinants of cancer. *Cold Spring Harbor Perspectives in Biology, 8*(9), a019505.
Bell, G., Hey, T., & Szalay, A. (2009). Beyond the data deluge. *Science, 323*(6), 1297–1298.
Bell, J. T., Pai, A. A., Pickrell, J. K., Gaffney, D. J., Pique-Regi, R., Degner, J. F., et al. (2011). DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biology, 12*(1), R10.
Berger, S. L., Kouzarides, T., Shiekhattar, R., & Shilatifard, A. (2009). An operational definition of epigenetics. *Genes and Development, 23,* 781–783.
Bernal, A. J., & Jirtle, R. L. (2010). Epigenomic disruption: The effects of early developmental exposures. *Clinical and Molecular Teratology, 88,* 938–944.
Bernstein, B. E., Mikkelsen, T. S., Xie, Xh, Kamai, M., Huebert, D. J., Cuff, J., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell, 125*(2), 315–326.
Bibikova, M., Barnet, B., Tsan, Ch., Ho, V., Klotzle, B., Le, J. M., et al. (2011). High density DNA methylation array with single CpG site resolution. *Genomics, 98*(4), 288–295.
Bjornsson, H. T., Sigurdsson, M. I., Fallin, M. D., Irizarry, R. A., Aspelund, A., Cui, H. M., et al. (2008). Intra-individual change over time in DNA methylation with familial clustering. *Journal of the American Medical Association, 299*(24), 2877–2883.
Blondel, V. D., Guillaume, J. L., Lambiote, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanic-Theory and Experiment, 8*(10), 10008.
Börner, K., Chen, C., & Boyack, K. W. (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology, 37,* 179–255.
Börner, K., Dall'Asta, L., Ke, W., & Vespignani, A. (2005). Studying the emerging global brain: Analyzing and visualizing the impact of co-authorship teams. *Complexity, 10*(4), 57–67.
Börner, K., Klavans, R., Patek, M., Zoss, A. M., Biberstine, J. R., Light, R. P., et al. (2012). Design and update of a classification system: The UCSD map of science. *PLOS ONE*. https://doi.org/10.1371/journal.pone.0039464.
Börner, K., Penumarty, S., Meiss, M., & Ke, W. (2006). Mapping the diffusion of scholarly knowledge among major U.S. research institutions. *Scientometrics, 68*(3), 415–426.
Bornmann, L., & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology, 66*(11), 2215–2222.
Boyack, K. W. (2009). Using detailed maps of science to identify potential collaborations. *Scientometrics, 79*(1), 27–44.
Boyack, K. W., Börner, K., & Klavans, R. (2009). Mapping the structure and evolution of chemistry research. *Scientometrics, 79*(1), 45–60.
Boyack, K. W., & Klavans, R. (2010). Co-citation analysis. Bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology, 61*(12), 2389–2404.
Boyack, K. W., & Klavans, R. (2014). Creation of a highly detailed, dynamic, global model and map of science. *Journal of the Association for Information Science and Technology, 65*(4), 670–685.
Boyer, L. A., Plath, K., Zeitlinger, J., Brambrinck, T., Medeiros, L. A., Lee, T. I., et al. (2006). Polycomb complexex repress developmental regulators in murine embryonic stem cells. *Nature, 441,* 349–353.
Braam, R. R., Moed, H. F., & van Raan, A. F. J. (1991). Mapping of science by combined co-citation and word analysis. I. Structural aspects. *Journal of the American Society for Information Science, 42*(2), 233–251.
Brandes, U. (2001). A faster algorithm for betweenness centrality. *The Journal of Mathematical Sociology, 25*(2), 163–177.
Brazel, A. J., & Vernimmen, D. (2016). The complexity of epigenetics diseases. *Journal of Pathology, 238*(2), 333–344.
Burt, R. S. (2004). Structural holes and good ideas. *American Journal of Sociology, 110*(2), 349–399.
Chen, C. (2004). Searching for intellectual turning points: Progressive knowledge domain visualization. *Proceedings of the National Academy of Sciences of the United States of America, 101*(Suppl. 1), 5303–5310.

Chen, C. (2005). The centrality of pivotal points in the evolution of scientific networks. In *IUI '05 proceedings of the 10th international conference on intelligent user interfaces* (pp. 98–105). New York: ACM.
Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology, 57*(3), 359–377.
Chen, C. (2012). Predictive effects of structural variation on citation counts. *Journal of the American Society for Information Science and Technology, 63*(3), 431–449.
Chen, C. (2016). Grand challenges in measuring and characterizing scholarly impact. *Frontiers in Research Metrics and Analytics*. https://doi.org/10.3389/frma.2016.00004.
Chen, C. (2017). Eugenes Garfield's scholarly impact: A scientometrics review. *Scientometrics, 114*(2), 489–516.
Chen, C., Ibekwe-SanJuan, F., & Hou, J. (2010). The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis. *Journal of the American Society for Information Science and Technology, 61*(7), 1386–1409.
Chen, C., & Leydesdorff, L. (2014). Patterns of connections and movements in dual-map overlays: A new method of publication portfolio analysis. *Journal of the Association for Information Science and Technology, 65*(2), 334–351.
Chi, P., Chen, Y., Zhang, L., Guo, J., Wongvipat, J., Shamu, T., et al. (2010). ETV1 is a linear survival factor that cooperates with KIT in gastrointestinal stromal tumours. *Nature Reviews Cancer, 467,* 849–853.
Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., et al. (2009). Lysine Acetylation targets protein complexex and co-regulates major cellular functions. *Science, 325*(5942), 834–840.
Clarivate Analytics. (2009). *Special topics. Epigenetics. Science Watch,* March. http://archive.sciencewatch.com/ana/st/epigen/. Accessed January 28, 2018.
Costas, R., & Bordons, M. (2008). Is g-index better than h-index? An exploratory study at the individual level. *Scientometrics, 77*(2), 267–288.
Darvish, H., & Tonta, Y. (2016). Diffusion of nanotechnology knowledge in Turkey and its network structure. *Scientometrics, 107*(2), 569–592.
Dawson, M. A., & Kouzarides, T. (2012). Cancer epigenetics: From mechanism to therapy. *Cell, 150*(1), 12–27.
Day, J. J., & Sweatt, J. D. (2012). Epigenetic treatments for cognitive impairments. *Neuropsychopharmacology, 37*(1), 247–260.
Deerwester, S., Dumais, S. T., Landauer, T. K., Furnas, G. W., & Harshman, R. A. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41*(6), 391–407.
Dolinoy, D. C., Huang, D., & Jirtle, R. L. (2007). Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. *Proceedings of the National Academy of Sciences of the United States of America, 134*(32), 13056–13061.
Dolinoy, D. C., Weidman, J. R., Waterland, R. A., & Jirtle, R. L. (2006). Maternal genistein alters coat color and protects Avy mouse offspring from obesity by modifying the fetal epigenome. *Environmental Health Perspectives, 114*(4), 567–572.
Duncan, E. J., Gluckman, P. D., & Dearden, P. K. (2014). Epigenetics, plasticity, and evolution: How do we link epigenetic change to phenotype. *Journal of Experimental Zoology, 322*(4), 208–220.
Dunham, I., & Kundaje, A. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature, 489,* 57–74.
Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics, 19*(1), 61–74.
Eckhardt, F., Lewin, J., Beck, S., et al. (2006). DNA methylation profiling of human chromosomes 6, 20 y 22. *Nature Genetics, 38,* 1378–1385.
Egghe, L. (2006). Theory and practise of the *g*-index. *Scientometrics, 69*(1), 131–152.
Ehlert, T., Simon, P., & Moser, D. A. (2013). Epigenetics in sports. *Sports Medicine, 43*(2), 93–110.
Ernst, J., Kherapoud, P., Mikkelsen, T. S., Shoresh, N., Ward, L. D., Epstein, Ch B, et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature, 473,* 43–49.
Esteller, M., Corn, P. G., Baylin, S. B., & Herman, J. G. (2001). A gene hypermethylation profile of human cancer. *Cancer Research, 61*(8), 3225–3229.
Feinberg, A. P. (2007). Phenotypic plasticity and the epigenetics of human disease. *Nature, 447*(7143), 433–440.
Feinberg, A. P. (2018). The key role of epigenetics in human disease prevention and mitigation. *The New England Journal of Medicine, 378,* 1323–1334.
Feinberg, A. P., Ohlsson, R., & Henikoff, S. (2006). The epigenetic progenitor origin of human cancer. *Nature Reviews Genetics, 7,* 21–33.

Figueroa, M. E., Abdel-Wahab, O., Lu, C., Ward, P. S., Patel, J., Shih, A., et al. (2010). Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell, 18*(6), 553–567.

Fong, C. Y., Morison, J., & Dawson, M. A. (2014). Epigenetics in the hematologic malignancies. *Haematologica, 99*(12), 1772–1783.

Foster, J. G., Rzhetsky, A., & Evans, J. A. (2015). Tradition and innovation in scientists' research strategies. *American Sociological Review, 80*(5), 875–908.

Frommer, M., MacDonald, L. E., Millar, D. S., Collis, C. S., Watt, F., Grigg, G. W., et al. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences of the United States of America, 89*(5), 1827–1831.

Gangisetty, O., Cabrera, M. A., & Murugan, S. (2018). Impact of epigenetics in aging and age related neurodegenerative diseases. *Frontiers in Bioscience, 23,* 1445–1464.

Garcia-Gimenez, J. L., Seco-Cervera, M., Tollefsbol, T. O., Roma-Mateo, C., Peiro-Chova, L., Lapunzina, P., et al. (2017). Epigenetic biomarkers: Current strategies and future challenges for their use in the clinical laboratory. *Critical Reviews in Clinical Laboratory Sciences, 54,* 529–550.

Gardiner-Garden, M., & Fommer, M. (1987). CpG Islands in vertebrate genomes. *Journal of Molecular Genomes, 196*(2), 261–282.

Garfield, E., & Sher, I. H. (1993). KeyWords Plus ™—Algorithmic derivative Indexing. *Journal of the American Society for Information Science, 44*(5), 298–299.

Gassen, N. C., Chrousos, G. P., Binder, E. B., & Zannas, A. S. (2017). Life stress, glucocorticoid signaling, and the aging epigenome: Implications for aging-related diseases. *Neuroscience and Biobehavioral Reviews, 74*(part B), 356–365.

Gaudet, F., Graeme Hodgson, J., Eden, A., Jackson-Grusby, L., Dausman, J., Gray, J. W., et al. (2003). Induction of tumors in mice by genomic hypomethylation. *Science, 300*(5618), 489–492.

Gensous, N., Bacalini, M. G., Pirazzini, Ch., Marasco, E., Giuliani, C., Ravaioli, F., et al. (2017). The epigenetic landscape of age-related diseases: The geroscience perspective. *Biogerontology, 18*(4), 549–559.

Glänzel, W., & Thijs, B. (2017). Using hybrid methods and `core documents' for the representation of clusters and topics: The astronomy dataset. *Scientometrics, 111*(2), 1071–1087.

Gläser, J., Glänzel, W., & Scharnhorst, A. (2017). Same data-different results? Towards a comparative approach to the identification of thematic structures in science. *Scientometrics, 111*(2), 981–998.

Griffith, B. C., Small, H. G., Stonehill, J. A., & Drey, S. (1974). The structure of scientific literatures II: Toward a macro and microstructure for science. *Science Studies, 4*(4), 339–365.

Guan, J.-S., Haggarty, S. J., Giacometti, E., et al. (2009). HDAC2 negatively regulates memory formation and synaptic plasticity. *Nature, 459,* 55–60.

Guo, J., Su, Y., Zhong, C., Ming, G. L., & Song, H. (2011). Hydroxilation of 5-methylcitosine by TET1 promotes active DNA demethylation in the adult brain. *Cell, 145*(3), 423–434.

Hammoud, S. S., Nix, D. A., Zhang, H., Purwar, J., Carrell, D. T., & Cairns, B. R. (2009). Distinctive chromatin in human sperm packages genes for embryo development. *Nature, 460*(7254), 473–478.

Hanahan, D., & Winberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell, 144*(5), 646–674.

He, Y., & Ecker, J. R. (2015). Non-CG methylation in the human genome. *Annual Review of Genomics and Human Genetics, 16,* 55–77.

Herceg, Z., & Vaissiere, T. (2011). Epigenetic mechanisms and cancer: An interface between the environment and the genome. *Epigenetics, 6*(7), 804–819.

Hodgson, K., Carless, M. A., Kulkarni, H., Curran, J. E., Sprooten, E., Knowles, E. E., et al. (2017). Epigenetic age acceleration assessed with human white-matter images. *Journal of Neurosciences, 37*(18), 4735–4743.

Hoeksema, M. A., & de Winther, M. P. (2016). Epigenetic regulation of monocyte and macrophage function. *Antioxidants and Redox Signal, 25*(14), 758–774.

Holmes, A. (2018). G2B reviews: Epigenetics, epitranscriptomics, microRNAs and more: Emerging approaches to the study of genes, brain and behavior. *Genes, Brain, and Behavior, 17,* e12453.

Hu, J., & Zhang, Y. (2017). Discovering the interdisciplinary nature of Big Data research through social network analysis and visualization. *Scientometrics, 112*(1), 91–109.

Jankovska, A. M., Millward, C. L., & Caldwell, Ch W. (2015). The potential of DNA modifications as biomarkers and therapeutic targets in oncology. *Expert Review of Molecular Diagnostics, 15*(10), 1325–1337.

Jeremias, G., Barbosa, J., Marques, S. M., Asselman, J., Gonçalves, F. J. M., & Pereira, J. L. (2018). Synthesizing the role of epigenetics in the response and adaptation of species to climate change in freshwater ecosystems. *Molecular Ecology, 27*(13), 2790–2806.

Jones, M. J., Goodman, S. J., & Kobor, M. S. (2015). DNA methylation and healthy human aging. *Aging Cell, 14*(6), 924–932.

Joubert, H., Haberg, S. E., Nilsen, R. M., Wang, X. T., Vollser, S. E., Murphy, S. K., et al. (2012). 450 epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. *Environmental Health Perspectives, 120*(10), 1425–1431.

Kaelin, W. G., & McKnight, S. L. (2013). Influence of metabolism on epigenetics and desease. *Cell, 153*(1), 56–69.

Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters, 31*(1), 7–15.

Klavans, R., & Boyack, K. W. (2017). Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge. *Journal of the American Society for Information Science and Technology, 68*(4), 984–998.

Kleinberg, J. (2002). Bursty and hierarchical structure in streams. In *Proceedings of of the 8th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 91–101). NewYork: ACM Press.

Kobow, K., & Blümcke, I. (2012). The emerging role of DNA methylation in epileptogenesis. *Epilepsia, 53*(Suppl. 9), 11–20.

LeBaron, M. J., Rasoulpour, R. J., Klapacz, J., Ellis-Hutchings, R. G., Hollnagel, H. M., & Gollapudi, B. B. (2010). Epigenetics and chemical safety assessment. *Mutation Research/Reviews in Mutation Research, 705*(2), 83–95.

Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology, 64*(1), 2–17.

Lee, T. I., et al. (2006). Control of developmental regulators by polycomb in human embryonic stem cells. *Cell, 125*(2), 301–313.

Leydesdorff, L. (2008). On the normalization and visualization of autor co-citation data: Salton's cosine versus the Jaccard Index. *Journal of the American Society for Information Science and Technology, 59*(1), 77–85.

Leydesdorff, L., Bornmann, L., & Wagner, C. S. (2017a). Generating clustered journal maps: An automated system for hierarchical classification. *Scientometrics, 110*(3), 1601–1614.

Leydesdorff, L., Carley, S., & Rafols, I. (2013a). Global maps of science based on the new Web-of-Science categories. *Scientometrics, 94*(2), 589–593.

Leydesdorff, L., Kogler, D. F., & Yan, B. (2017b). Mapping patent classifications: Portfolio and statistical analysis, and the comparison of strengths and weaknesses. *Scientometrics, 112*(3), 1573–1591.

Leydesdorff, L., & Rafols, I. (2009). A global map of science based on the ISI Subject categories. *Journal of the American Society for Information Science and Technology, 60*(2), 348–362.

Leydesdorff, L., Rafols, I., & Chen, C. (2013b). Interactive overlays of journal and the measurement of interdisciplinarity on the basis of aggregated Journal–Journal citations. *Journal of the American Society for Information Science and Technology, 64*(12), 2573–2586.

Li, B., Carey, M., & Workman, J. L. (2007). The role of chromatin during transcription. *Cell, 128*(4), 707–719.

Li, C. C. Y., Maloney, Ch A, Cropley, J. E., & Suter, C. M. (2010). Epigenetic programming by maternal nutrition: Shaping future generations. *Epigenomics, 2*(4), 539–549.

Lillycrop, K. A., Phillips, E. S., Jackson, A. A., Hanson, H. A., & Burdge, G. C. (2005). Dietary protein restriction of pregnant rats induces and folic acids supplementations prevents epigenetic modification of hepatic gene expression in the offspring. *The Journal Nutrition, 135*(6), 1382–1386.

Liu, S. (2012). Epigenetics advancing personalized nanomedicine in cancer therapy. *Advanced Drug Delivery Reviews, 64*(13), 1532–1543.

Liu, Y., Aryee, M. J., Padyukov, L., Fallin, M. D., Hesselberg, E., Runarsson, A., et al. (2013). Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nature Biotechnology, 31,* 142–147.

López, V., Fernández, A. F., & Fraga, M. F. (2017). The role of 5-hydroxymethylcytosine in development, aging and age-related diseases. *Ageing Research Reviews, 37*(August), 28–38.

Mane, K. K., & Börner, K. (2004). Mapping topics and topic burst in PNAS. *Proceedings of the National Academy of Sciences of the United States of America, 101*(Suppl. 1), 5287–5290.

Martin-Subero, J. I., & Esteller, M. (2011). Profiling epigenetic alterations in disease. In E. Ballestar (Ed.), *Epigenetic contributions in autoimmune disease. Advances in experimental medicine and biology* (pp. 162–178). Boston: Springer.

Mellen, M., Ayata, P., & Heintz, N. (2017). 5-hydroxymethylcytosine accumulation in postmitotic neurons results in functional demethylation of expressed genes. *Proceedings of the National Academy of Sciences of the United States of America, 114*(37), 7812–7821.

Metivier, R., Gallais, R., Tiffoche, Ch., Le Peron, Ch., Jurkowska, R. Z., Carmouche, R. P., et al. (2008). Cyclical methylation of a transcriptionally active promoter. *Nature, 452,* 45–50.
Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., et al. (2007). Genome-wide maps of chromatin state in pluripontent and lineage-committed cells. *Nature, 448,* 553–560.
Milojevic, S. (2015). Quantifying the cognitive extent of science. *Journal of Informetrics, 9*(4), 962–973.
Moore, L. D., Le, T., & Fan, G. (2013). DNA methylation and its basic function. *Neuropsychopharmacology, 38*(1), 23–38.
Morgan, H. D., Santos, F., Green, K., Dean, W., & Reik, W. (2005). Epigenetic reprogramming in mammals. *Human Molecular Genetics, 14*(1), R47–R58.
Morris, T. F., & Beck, S. (2015). Analysis pipelines and packages for Infinium human methylation450 bead chip (450 k) data. *Methods, 72,* 3–8.
Morris, S. A., Yen, G., Wu, Z., & Asnake, B. (2003). Time line visualization of research fronts. *Journal of the American Society for Information Science and Technology, 54*(5), 413–422.
Nagy, C., & Turecki, G. (2015). Transgenerational epigenetic inheritance: An open discussion. *Epigenomics, 7,* 781–790.
Ng, A. Y., Jordan, M. I., & Weiss, Y. (2002). On spectral clustering: Analysis and an algorithm. In *NIPS'01 Proceedings of the 14th international conference on neural information processing systems: Natural and synthetic* (pp. 849–856). Cambridge, MA: MIT Press.
Ng, S. F., Lin, R. C., Laybutt, D. R., Barres, R., Owens, J. A., & Morris, M. J. (2010). Chronic high-fat diet in fathers programs β-cell dysfunction in female rat offspring. *Nature, 467*(7318), 963–966.
Ong, M.-L., Lin, X., & Holbrook, J. D. (2015). Measuring epigenetics as the mediator of gene/environment interactions in DOHaD. *Journal of Developmental Origins of Health and Disease., 6*(1), 10–16.
Paluch, B. E., Naqash, A. R., Brumberger, Z., Nemeth, M. J., & Griffiths, E. A. (2016). Epigenetics: A primer for clinicians. *Blood Reviews, 30*(4), 285–295.
Pautasso, M. (2012). Publication growth in biological sub fields: Patterns, predictability and sustainability. *Sustainability, 4,* 3234–3247.
Peleg, S., Sananbenesi, F., Zovoilis, A., et al. (2010). Altered histone acetylation is associated with age-dependent memory impairment in mice. *Science, 328,* 753–756.
Persson, O. (1994). The intellectual base and research fronts of JASIS 1986-1990. *Journal of the American Society for Information Science, 45*(1), 31–38.
Pfeifer, G. P., Xiong, W., Hahn, M. A., & Jin, S.-G. (2014). The role of 5-hydroxymethylcytosine in human cancer. *Cell and Tissue Research, 356*(3), 631–641.
Pham, M. T., Rajic, A., Greig, J. D., Sargeant, J. M., Papadopoulus, A., & McEwen, S. A. (2014). A scoping review of scoping reviews: Advancing the approach and enhacing the consistency. *Research Synthesis Methods, 5*(4), 371–385.
Price, D. J. S. (1965). Networks of scientific papers. *Science, 149*(3683), 510–515.
Pudovkin, A. I., & Garfield, E. (2002). Algorithmic procedure for finding semantically related journals. *Journal of the American Society for Information Science and Technology, 53*(3), 1113–1119.
Rainho Bras, O., Cointet, J.-P., Cambrosio, A., David, L., Arriscado Nunes, J., Cardoso, F., et al. (2017). Oncology research in late twentieth century and turn of the century Portugal: A scientometric approach to its institutional and semantic dimensions. *Scientometrics, 113*(2), 867–888.
Rakyan, V. K., Down, T. A., Balding, D. J., & Beck, S. (2011). Epigenome-wide association studies for human common diseases. *Nature Reviews Genetics, 12,* 529–541.
Ramirez-Prado, J. S., Abulfaraj, A. A., Rayapuram, N., Benhamed, M., & Hirt, H. (2018). Plant immunity: From signaling to epigenetic control of defense. *Trends in Plant Science*. https://doi.org/10.1016/j.tplants.2018.06.004.
Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature, 447*(7143), 425–432.
Reik, W., & Walter, J. (2001). Genomic imprinting: Parental influence on the genome. *Nature Review Genetics, 2*(1), 21–32.
Rodriguez-Paredes, M., & Esteller, M. (2011). Cancer epigenetics reaches mainstream oncology. *Nature Medicine, 17*(3), 330–339.
Roos, L., van Dongen, J., Bell, Ch G, Burri, A., Deloukas, P., Boomsma, D. I., et al. (2016). Integrative DNA methylome analysis of pan-cancer biomarkers in cancer discordant monozygotic twin-pairs. *Clinical Epigenetics, 8*(7), 1–16.
Rosenfeld, Ch S. (2010). Animal models to study environmental epigenetics. *Biology of Reproduction, 82*(3), 473–488.
Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20,* 53–65.

Roy, D. M., Walsh, L. A., & Chan, T. A. (2014). Driver mutations of cancer epigenomes. *Protein and Cell, 5*(4), 265–296.
Sandstrom, P. E. (1994). An optimal foraging approach to information seeking and use. *Library Quarterly, 64*(4), 414–449.
Schoofs, T., Berdel, W. E., & Müller-Tidow, C. (2014). Origins of aberrant dna methylation in acute myeloid leukemia. *Leukemia, 28,* 1–14.
Schvaneveldt, R. W., Durso, F. T., & Dearholt, D. W. (1989). Network Structures in Proximity Data. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 24, pp. 249–284). New York, NY: Academic Press.
Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., et al. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell, 119*(7), 941–953.
Shibata, N., Kajikawa, Y., Takeda, Y., & Matsushima, K. (2009). Comparative study on methods of detecting research fronts using different types of citation. *Journal of the American Society for Information Science and Technology, 60*(3), 571–580.
Skvortsova, K., Zotenko, E., Luu, P.-L., Gould, C. M., Nair, S. S., Clark, S. J., et al. (2017). Comprehensive evaluation of genome-wide 5-hydroxymethylcytosine profiling approaches in human DNA. *Epigenetics and Chromatin, 10*(16), 1–20.
Small, H. (2010). Maps of science as interdisciplinary discourse: Cocitation context and the role of analogy. *Scientometrics, 83*(3), 835–849.
Small, H., & Griffith, B. C. (1974). The structure of scientific literatures I: identifying and graphing specialties. *Science Studies, 4*(P.1), 17–40.
Ŝubelj, L., Bajec, M., Boshkoska, B. M., Kastrin, A., & Levnajić, Z. (2015). Quantifying the consistency of scientific databases. *PLoS ONE, 10*(5), e0127390. https://doi.org/10.1371/journal.pone.0127390.
Ŝubelj, L., van Eck, N. J., & Waltman, L. (2016). Clustering scientific publications based on citation relations: A systematic comparison of different methods. *PLoS ONE, 11*(4), e0154404. https://doi.org/10.1371/journal.pone.0154404.
Takai, D., & Jones, P. A. (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proceedings of the National Academy of Sciences of the United States of America, 99*(6), 3740–3745.
Tejedor, J. R., & Fraga, M. F. (2017). Interindividual epigenetic variability: Sound or noise? *BioEssays, 39*(7), 1700055.
Testa, J. (2009). The Thomson Reuters selection process. *Transnational Corporations Review, 1*(4), 59–66.
van Noorden, R., Maher, B., & Nuzzo, R. (2014). The top 100 papers. *Nature, 514*(30), 550–553.
Vargas-Quesada, B., Chinchilla-Rodríguez, Z., & Rodríguez, N. (2017). Identification and visualization of the intelectual structure in graphene research. *Frontiers in Research Metrics and Analytics*. https://doi.org/10.3389/frma.2017.00007.
Velden, T., Boyack, K. W., Gläser, J., Koopman, R., Scharnhorst, A., & Wang, S. (2017). Comparison of topic extraction approaches and their results. *Scientometrics, 111*(2), 1169–1221.
Waddington, C. H. (1942). The epigenotype. *Endeavor*, *1*, 18–20. Reprinted in *International Journal Epidemiology*, *41*(1), 10–13 (2012).
Watanabe, A., Yamada, Y., & Yamanaka, S. (2013). Epigenetic regulation in pluripotent stem cells: A key to breaking the epigenetic barrier. *Philosophical Transactions of the Royal Society, 368*(1609), 20120292.
Waterland, R. A., & Jirtle, R. L. (2003). Transposable elements: Targets for early nutritional effects on epigenetic gene regulation. *Molecular and Cellular Biology, 23*(15), 5293–5300.
Wolff, G. L., Kodell, R. L., Moore, S. R., & Cooney, C. A. (1998). Maternal epigenetics and methyl supplements affect agouti gene expression in $A^{vy}/a$ mice. *Faseb journal, 12,* 949–957.
Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S. H., et al. (2011). Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. *Cancer Cell, 19*(1), 17–30.
Yang, A. Y., Kim, H., Li, W., & Jong, A.-N. T. (2016). Natural compound-derived epigenetic regulators targeting epigenetic readers, writers and erasers. *Current Topics in Medicinal Chemistry, 16*(7), 697–713.
Yao, B., & Jin, P. (2014). Unlocking epigenetic codes in neurogénesis. *Genes and Development, 28*(12), 1253–1271.
Zama, A. P., & Uzumcu, M. (2010). Epigenetics effects of endocrine-disrupting chemicals on female reproduction: An ovarian perspective. *Frontiers in Neuroendocrinology, 31*(4), 420–439.

Zhao, D., & Strotmann, A. (2014). The knowledge base and research front of Information Science 2006–2010: An author co-citation and bibliographic coupling analysis. *Journal of the Association for Information Science and Technology, 65*(5), 996–1006.

Zhu, Y., Kim, M. C., & Chen, C. (2017). An investigation of the intellectual structure of opinion mining research. *Information Research, 22*(1), paper 739. http://InformationR.net/ir/22-1/paper739.html. Accessed March 16, 2018.

Zufferey, F., Williams, F. M., & Spector, T. D. (2014). Epigenetics and methylation in the rheumatic diseases. *Seminars in Arthritis and Rheumatism, 43*(5), 692–700.