Orphan Drugs and Rare Diseases: A Scientometric Review (2000–2014)

Article *in* Expert Opinion on Orphan Drugs · May 2014

DOI: 10.1517/21678707.2014.920251

3 authors, including:

Chaomei Chen
College of Computing and Informatics, Drexe…
322 PUBLICATIONS **5,929** CITATIONS
SEE PROFILE

Meen Chul Kim
Drexel University
18 PUBLICATIONS **44** CITATIONS
SEE PROFILE

EXPERT OPINION

Orphan drugs and rare diseases: a scientometric review (2000 – 2014)

Chaomei Chen†, Rachael Dubin & Meen Chul Kim
Drexel University, College of Computing and Informatics, Philadelphia, PA, USA

Introduction: The literature of orphan drugs and rare diseases between 2000 and 2014 is reviewed. The overall structure of its intellectual landscape is characterized in terms of thematic concentrations of co-cited references and emerging trends of surging keywords and citations to references through a scientometric review – a quantitative study of scientific literature.

Areas covered: The review is based on two sets of bibliographic records retrieved from the Web of Science. The core dataset, consisting of 9461 original research articles and review papers, was constructed from a topic search. The expanded dataset, consisting of 82,765 articles and reviews, was constructed by citation expansion.

Expert opinion: The study has revealed three broad categories of research. The research policy category focuses on the strategic and integral role of the study of orphan drugs and rare diseases in a wide-ranging societal context, including optimizing the allocation of resources and setting appropriate evaluation standards. The basic research category focuses on understanding the complex biological and genetic mechanisms of rare diseases. The disease-specific research category focuses on specific rare diseases with clinical and therapeutic goals. The study has revealed the fundamental role of genetic variation in linking a diverse range of rare and complex phenotypic patterns. The field would benefit from more comprehensive reviews of all three categories as a whole and disseminating findings and technical advances across various specialties. Finally, building stronger connections between the study of rare diseases and the study of more common diseases is recommended for all three categories.

Keywords: CiteSpace, orphan drugs, rare diseases, scientometrics

Expert Opinion on Orphan Drugs [Early Online]

1. Introduction

Rare diseases are defined as a disease that affects fewer than 200,000 people in the US or a disease for which it is unlikely that the cost of drug development would be able to be recuperated through drug sales (either due to high cost of development or due to limited market) [1]. Economic, regulatory and societal implications of rare diseases have been recognized. The 1983 US Orphan Drug Act provides incentives such as public funding, tax credits and market exclusivity for the developers of drugs for rare diseases. In Europe, a rare disease is defined to have a prevalence of no more than 5 per 10,000 of the population. The European Union's Orphan Drug Regulation provides market exclusivity and assistance with research protocols and associated fees but does not include specific grants or tax credits [2]. Although only a very small portion of the population is impacted by each rare disease, it is estimated that 30 million Europeans and 25 million Americans are potentially affected by rare diseases.

The goal of this article is to reveal the intellectual landscape of the study of orphan drugs and identify thematic patterns, landmark articles and emerging trends.

informa
healthcare

Article highlights.

- The literature of orphan drugs and rare diseases between 2000 and 2014 is visualized and reviewed.
- Three major types of research are prominent, namely, research policy, basic research and disease-specific research.
- Research in genetic variation plays a fundamental role in linking a diverse range of rare and complex diseases.
- Holistic and comprehensive reviews of the field as a whole are recommended.
- Increasing the breadth and depth of research in established and potential connections between rare diseases and more common diseases is recommended.

This box summarizes key points contained in the article.

Unlike a conventional literature survey, which is typically prepared by a domain expert, our review is guided by a computational approach implemented in CiteSpace, a visual analytic system for visualizing emerging trends and critical changes in scientific literature [3,4].

The input of the process is a collection of scientific publications representative for the study of orphan drugs. Thematic patterns and emerging trends are computationally identified and detected. The computational and visual analytic approach that guides our review is drawn from the field of scientometrics and information visualization. Scientometrics is the quantitative study of science. This kind of approach has a few distinct advantages over a conventional expert-compiled review. First, a much broader and more diverse range of relevant topics can be reached than the conventional approach. Second, such reviews can be generated as frequently as needed, although one does not have much control over whether and when a conventional review of their field will be available. Computationally assisted literature reviews are not designed to replace expert-made reviews; rather, they are intended to provide an additional point of reference with a certain set of benefits. The depth of reviews and insights produced by domain experts are indispensable for the development of an understanding of complex subject matter.

2. Methods

Two datasets of bibliographic records on orphan drugs were retrieved from the Web of Science using a topic search and a subsequent expansion through citation links. The topic-search dataset is referred as the core dataset. The expanded dataset represents a broader context of the core. Key findings are highlighted based on the core dataset and then move onto a more detailed study of the expanded dataset using various visual analytic functions implemented in CiteSpace [3,5].

2.1 Bibliographic records

Each bibliographic record contains the metadata of a published article, including a list of authors, the title, the abstract, a set of keywords and a set of references cited by the article. Each reference contains the first author's name, the year of publication, the source (i.e., the journal, the conference or other form of publication), the volume number, the number of the first page and the DOI reference. Given a DOI reference, one can access the full text of the corresponding article.

2.1.1 *The core dataset*

The core dataset is retrieved by a topic search in the Web of Science. The search query consists of four phrases about orphan drugs:

'*orphan drug**' OR '*orphan disease**' OR '*rare disease**' OR '*orphan medicinal product**'

The wildcard * was used to capture relevant variations of a word, such as orphan drug and orphan drugs. A record is considered relevant if any of terms is found in the title, abstract or keyword fields of the record. The query resulted in 9461 records of original research articles and review papers. Among them, 7753 records were published during the period of 2000 – 2014.

2.1.2 *The expanded dataset*

The expanded dataset is a superset of the core dataset with extra records obtained by association through citation links. Even if an article does not contain any of the query terms in the topic search, if it cites at least one article in the core set, it becomes reasonable to assume that it *may* be thematically relevant to the subject matter underlying the core dataset. In subsequent analyses, any irrelevant articles introduced through this citation expansion will eventually be filtered out. Thus, it is justifiable for us to cast a wider net in the expansion. The citation expansion method is originated in the principle of citation index by Eugene Garfield in the 1960s [6].

In the Web of Science, the core dataset as a whole was cited by 76,897 articles. These records were merged into the core dataset. The resultant expanded dataset consists of 82,765 unique records (**Table 1**). Our scientometric review is based on both datasets.

2.2 Revealing the intellectual landscape and detecting emerging topics

The intellectual landscape of a scientific field can be represented by a network of a variety of entities such as cited references, collaborating authors and co-occurring keywords. CiteSpace supports the construction of several types of networks from bibliographic sources [7]. This study focuses on document co-citation networks and networks of co-occurring keywords.

2.2.1 *Bibliographic landscape*

A link in a document co-citation network represents how frequently two articles are cited together by other articles in a dataset such as the core and expanded datasets. Individual nodes in the network can be aggregated into groups, or

Table 1. The core and expanded datasets used in the analysis.

Dataset	Duration	Results	Articles	Reviews	Authors	References	Keywords	Institutions
Core	1946 – 2014	9461	6765	1154	46,587	247,562	82,861	23,889
Expanded	1946 – 2014	82,765	56,822	14,923	442,709	3,946,944	938,718	243,686

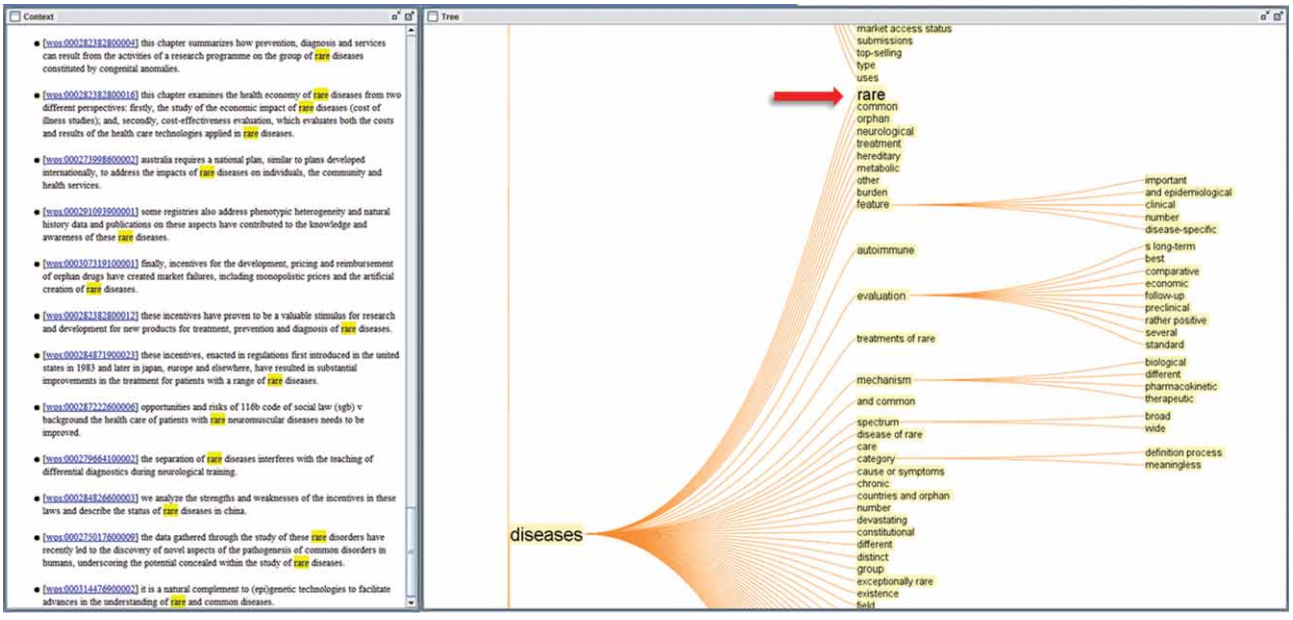

Figure 1. Right: Representative concepts of the largest cluster derived from the core dataset. Left: Sentences in which the selected concept – *rare diseases* – was found in various records.

clusters, based on their interconnectivity. Each cluster represents a distinct specialty or a thematic concentration. Other points of interest include highly cited landmark articles, articles with strong citation bursts and keywords with a strong surge of frequency. Emerging topics are identified based on these properties.

CiteSpace is designed to synthesize and visualize a time series of individual networks extracted from each year's publications. The resultant network can be divided into clusters, that is, groups of entities, such that entities within the same group are more similar to each other than entities from different groups. The homogeneity of each group is measured by a silhouette score from the lowest -1 to the highest 1. The quality of the overall division is measured by the modularity measure.

2.2.2 *Emerging topics*

CiteSpace identifies emerging topics in terms of highly cited landmark articles, articles with strong citation bursts and keywords with a strong surge of frequency. The goal of burst detection is to determine whether the appearance of an entity increases sharply with reference to its peers [8]. For example, if the number of articles with the term *genome-wide association studies* (GWAS) in their titles or abstracts sharply increased in the last 2 years at a much faster rate than other terms, then the term is defined as a burst term. Similarly, if an article is found to have a steep increase of its citation counts, then the article is regarded as having a citation burst. A citation burst indicates an increased attention to the underlying work, thus providing a key indicator of an emerging topic.

2.2.3 *Exploration*

In CiteSpace, users may explore the intellectual landscape of a field through interactive visualizations. For example, the nature of a cluster of co-cited references can be identified by algorithmically generated labels of the cluster, representative concepts in the cluster (**Figure 1**) and the full text of an article of interest [9].

3. Bibliographic landscapes

3.1 The landscape based on the core dataset

Figure 2 shows the overview of a document co-citation network generated from the core dataset. The network contains 1191 references cited between 2000 and 2014. The largest cluster is in the center of the visualization, that is,

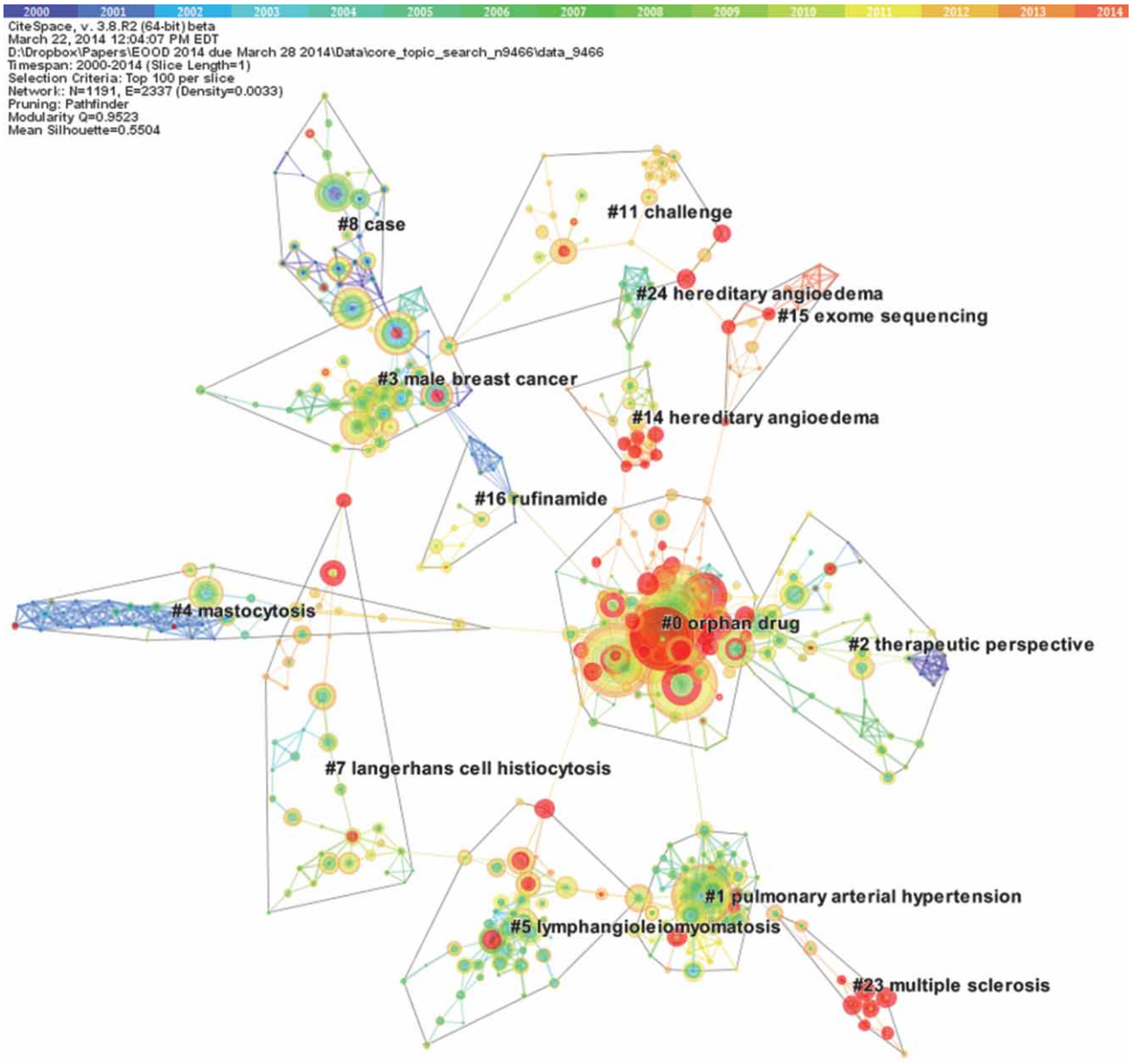

Figure 2. A document co-citation network of the core dataset.

#0 orphan drug. Each node depicts a cited reference. The citation history is visualized in terms of 'tree rings' of different colors and thickness. Nodes with citation bursts are visualized with rings in red.

3.1.1 *Orphan drug*

The largest cluster, #0 orphan drug, contains 103 member references. The average year of publication is 2005, making this the most recent among the largest 10 clusters. The homogeneity of the cluster, measured by the silhouette score, is 0.954, which is very close to the highest value of 1.00, suggesting a reliable quality. As shown in **Figure 2**, this cluster has a high concentration of articles with citation bursts. Thus, the orphan drug cluster is essential to the literature represented by the core dataset.

The effectiveness of the Orphan Drug Act in promoting the development of treatments for rare diseases is undeniable; however, it is not without its limitations. Companies may collect multiple orphan drug designations to treat extremely similar diseases; for example, the drug azidothymidine received orphan status for AIDS as well as AIDS-related complex [1]. Exclusive marketing rights (EMRs) granted to orphan drugs are intended to provide incentives for research in relevant areas. On the other hand, as EMRs are granted only to the first drug approved by the FDA unless a newly introduced drug is more effective, capturing multiple orphan drug

designations for a group of related diseases can greatly limit competition. The determination of whether a new drug is in fact different enough from an existing orphan drug to be granted an additional orphan drug designation has been a source of many legal battles [10].

In addition to the actions of pharmaceutical companies, the existing orphan drug legislation itself has also been questioned. A disease classified as an 'orphan' may become more prevalent over time; in the case of a disease such as AIDS, the financial burden of developing a drug would have been recouped easily without the need of additional incentives [1]. Even if the original disease remains rare, an orphan drug may also be found to be effective for other, less rare diseases [11]. Others have questioned whether it is in fact justifiable for extensive public resources to be provided to private industry in the form of orphan drug incentives, particularly when other, more widespread diseases could potentially be treated at lower cost [12].

3.1.2 *Male breast cancer*

Other large clusters in the core dataset focus on specific rare diseases and their treatment. The fourth cluster (#3), labeled as male breast cancer, has 49 members and an average year of 1993. Cancers make up a large proportion (31%) of orphan diseases [13]; as such, it is unsurprising that a form of cancer is the focus of one of the large clusters in the core orphan drug dataset. Although breast cancer is not a rare disease for women, males are affected by < 1% of all breast cancer cases [14]. Although survival rates between men and women with comparable levels of disease severity seem to be relatively similar, men typically recognize breast cancer at a later stage of the disease and more frequently present with metastasis to the lymph nodes, leading to a higher death rate for diagnosed males [15].

3.2 Articles with citation burst

The dynamics of a field can be characterized in part by articles that have received the steepest increase of citations, that is, citation bursts. **Figure 3** shows the top 30 references with the strongest citation bursts. A citation burst indicates the likelihood that the scientific community has paid or is paying special attention towards the underlying contribution. Instead of discussing all the 30 references, the following discussions will be limited to the ones with the strongest burst in the group of articles that started to burst at the same time. If several references started to burst in year 2000, only the one with the greatest magnitude of burst will be concentrated on. References selected in this way are indicated by red arrows in **Figure 3**.

As shown in **Table 2**, among citation bursts starting in 2000, the strongest burst is associated with a 1958 paper by Kaplan and Meier [16]. This episode of burst ended in 2002. The paper is about nonparametric estimation from incomplete observations. This is an extremely highly cited paper with 38,905 citations on the Web of Science at the time of writing. In particular, this analytic methodology paper is mostly cited by articles on oncology (18,048), surgery (5215) and hematology (4360).

The strongest burst starting from 2006 is associated with a 2004 paper by Simonneau *et al.* on a clinical classification of pulmonary hypertension [17]. According to the authors, because the absolute risk of known risk factor for pulmonary arterial hypertension (PAH) is generally low, individual susceptibility or genetic predisposition is likely to play an important role. The article particularly addressed the possible connections between PAH and a few rare diseases such as autosomal recessive disorder, Gaucher's disease and Osler-Weber-Rendu disease. The clinical classification was subsequently updated in 2009. The updated version is prominent in the expanded dataset.

The strongest burst from 2009 is due to a 2008 paper by Buckley on clinical trials of orphan medicines. Buckley's 2008 paper [18] belongs to cluster #0 orphan drug. Because orphan diseases are so rare, gaining access to a large enough sample of affected individuals poses a significant challenge to clinical trials. The article discusses these challenges and some of the ways in which they have and have not been addressed. For example, alternative study designs (such as Bayesian methods) may reduce the number of study participants needed. If a drug already in use is being repurposed for an orphan disease, existing literature may provide insights into likely levels of effectiveness and risk. However, Buckley notes that the standards used to determine whether sufficient evidence is provided for an orphan drug appear to vary wildly based on his analysis of drugs authorized for use in the EU.

Citation bursts starting in 2010 are led by Drummond *et al.*'s article published in 2007, written by five authors from the UK, the US and Spain [19]. Its key argument is that the standard health technology assessment (HTA) method is not suitable to assess orphan drugs because orphan drugs would be deemed to be not cost-effective by the economic evaluation standard in HTA. The authors argue that there was a policy gap in accessing to drugs for rare diseases and there was a need to address the shortcomings in the evaluation of orphan drugs. The authors proposed a research agenda of two key themes: i) assessing the societal value of orphan drugs; and ii) funding the development and use of orphan drugs. Drummond's paper also belongs to cluster #0 orphan drug.

The next article is by Griggs *et al.* in 2009 [20], which is again from cluster #0 orphan drug. It has the strongest citation burst starting in 2011, and it is still going strong. Their article reviewed important issues facing clinical investigators who are interested in the study of rare diseases. For example, the most frequent problem is the recruitment of study subjects for an observational cohort or a clinical trial. Studying rare diseases requires extra care to protecting the privacy of study subjects because of the small sample and population size. The paper introduced a series of valuable resources for

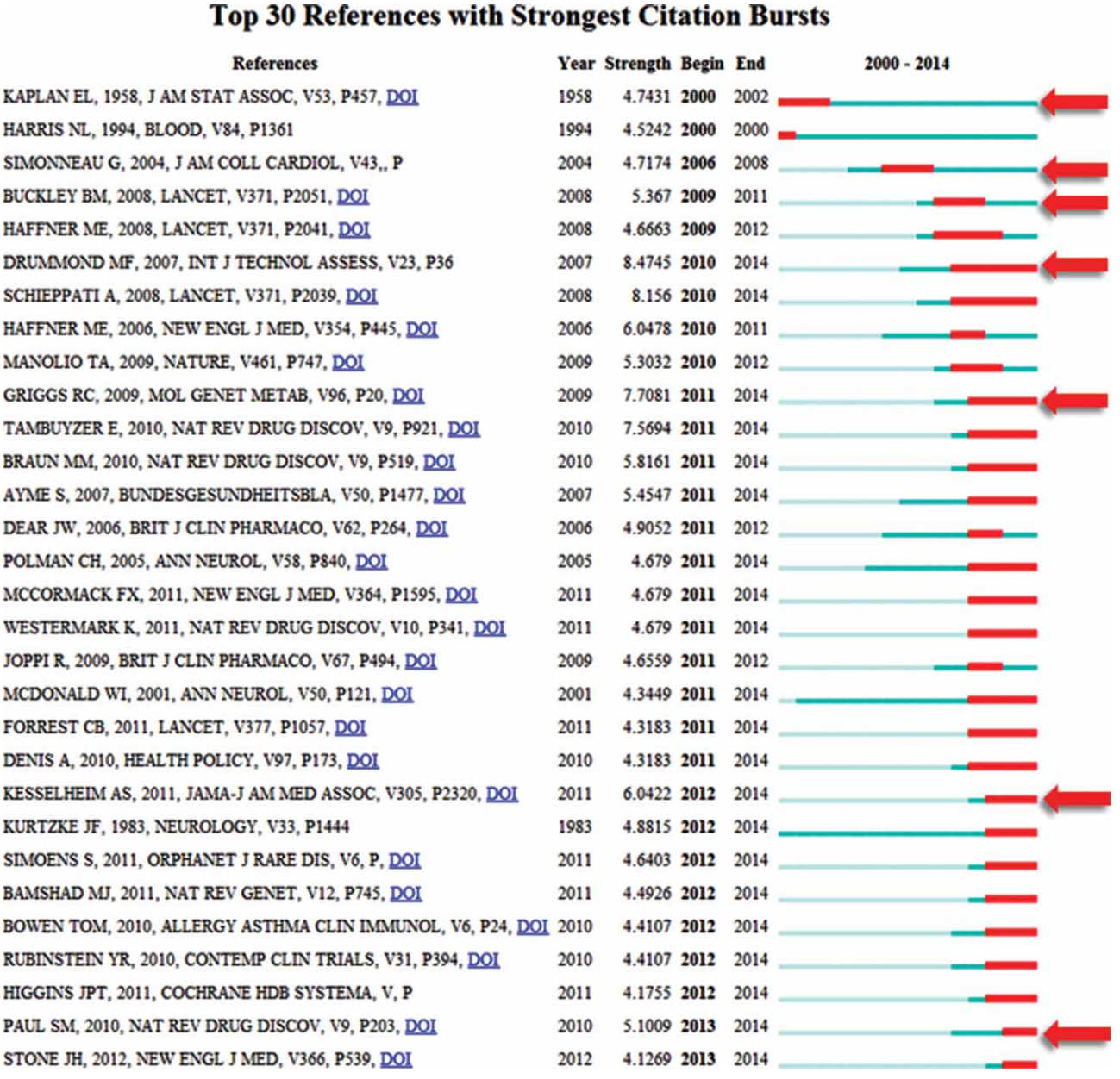

Figure 3. Top 30 references with strong citation bursts. The ones pointed by red arrows are reviewed in the article.

Table 2. References with the strongest citation bursts starting from the same years.

Study	Citation burst		
	Strength	Begin	End
Kaplan *et al.* 1958 [16]	4.74	2000	2002
Simonneau *et al.* 2004 [17]	4.72	2006	2008
Buckley *et al.* 2008 [18]	5.37	2009	2011
Drummond *et al.* 2007 [19]	8.47	2010	2014
Griggs *et al.* 2009 [20]	7.71	2011	2014
Kesselheim *et al.* 2011 [21]	6.04	2012	2014
Paul *et al.* 2010 [22]	5.10	2013	2014

research and training in the study of rare diseases. Its citation burst started in 2011, and it is still elevated.

With the strongest burst starting in 2012, Kesselheim *et al.* in 2011 characterized orphan drugs for cancer and corresponding pivotal clinical trials in comparison with non-orphan drugs [21]. Based on 15 orphan drugs and 12 non-orphan drugs approved between 2004 and 2010, the study found that pivotal trials of orphan drugs had smaller participant numbers, less likely to be randomized and less likely to be double-blind.

The most recent strongest burst started in 2013 is due to Paul *et al.* in 2010 [22]. The paper addresses how to improve R&D productivity in the pharmaceutical industry as its grand challenge, although the paper does not seem to address orphan drugs or rare diseases specifically.

Examining the group of most highly impactful articles on orphan drugs, identified based on citation bursts, shows that with the exception of Simonneau *et al.* these articles primarily focused on logistical issues concerning assessment standards, challenges faced by clinical investigators in the study of rare diseases and characteristics of clinical trials associated with orphan drugs. The influence of this particular set of articles within the core dataset suggests that questions of scientific rigor and the role of regulations in designing and executing clinical trials for orphan drugs are of significant and wide-reaching concern within this community.

4. Emerging trends and new developments in a broader context

The expanded dataset, consisting of 82,765 records (or about nine times as many as the core dataset), sets the core dataset in a broader context with contributions from a total of 442,709 distinct authors of 243,686 institutions. The additional records are included because they cited one or more articles in the core dataset. The expanded dataset contains over 3.9 million references and 938,718 keywords.

4.1 Keywords as indicators of emerging trends (2000 – 2014)

Just as citation bursts may indicate the degree of attention from the scientific community to a published article, burst detection can also identify bursts keywords as indicators of emerging trends. **Figure 4** shows the top 30 keywords with the strongest bursts in their appearances during 2000 and 2014. A large number of keywords started to burst since year 2000. The strongest ones include *ataxia-telangiectasia* (AT), *prognostic factors* and *linkage disequilibrium* (LD). Subsequently surged keywords include *s-phase checkpoint*, *Crohn's disease* and *single-nucleotide polymorphisms* (SNPs). The most recent burst of keyword is *genome-wide association*. The role of these key concepts and their interrelationships will become clear when the structure and dynamics of the underlying intellectual landscape is discussed.

4.2 References with strong citation bursts (2000 – 2014)

Figure 5 shows the top 30 references with the strongest citation bursts in the expanded dataset. Furthermore, references that started to burst at the same time are considered as a group. For each group, the reference with the strongest burst within the group will be focused on. For example, as **Figure 5** shows, nine references started to burst in year 2000. In this group, the reference with the strongest citation burst is regarded as the most representative, namely, Lim *et al.*'s 2000 article [23].

Table 3 shows the representative references for each of five groups of references by the beginning time of burst. Lim *et al.* in 2000 investigated the phenotypic similarities in AT and Nijmegen breakage syndrome (NBS) in terms of the functional interactions between the two genes whose mutations are responsible for the two diseases [23]. The article has the second strongest citation burst in the entire expanded dataset. The burst lasted for 5 years from 2000 till 2004.

Rioux *et al.*'s article published in 2001 [24] has the strongest citation burst in the entire expanded dataset. They applied LD mapping to identify the linkage between a gene (IBD5) and Crohn's disease. This was the first comprehensive application of hierarchical LD mapping involving a systematic search for LD across a linkage peak, rigorous bounding of the critical region and exhaustive ascertainment of SNPs in the critical regions.

Gadodiamide is used in MRI to enhance the visualization of blood vessels. In 2006, Marckmann *et al.* investigated the possible connection between gadodiamide and nephrogenic systemic fibrosis, which is a rare disease [25]. The study indicates that gadodiamide plays a causative role in nephrogenic systemic fibrosis. This article is ranked the third by the strength of its citation burst.

In 2009, Simonneau *et al.* published an update of the previous clinical classification of pulmonary hypertension [26]. The previous article about the clinical classification is identified in the analysis of the core dataset.

In 2010, Bertram *et al.* reviewed the genetic research in Alzheimer's disease [27]. Having reviewed the advances made by GWAS in identifying Alzheimer's disease susceptibility genes, the review concluded that GWAS have substantially reshaped the landscape of late-onset Alzheimer's disease genetics during the courses of only 3 years. Furthermore, the review also looked beyond GWAS for discovering rare causal genetic variants of larger effects.

4.3 The landscape based on the expanded dataset

Figure 6 shows a document co-citation network derived from the expanded dataset. The network represents the collective pattern of the citation decisions of 500 articles that are most cited each year themselves for each year in the period of 2000 – 2014. The resultant network represents the intellectual landscape of orphan drugs and rare diseases in a broader context.

Top 30 Keywords with Strongest Citation Bursts

Keywords	Year	Strength	Begin	End	2000 - 2014
ataxia-telangiectasia	2000	70.3247	**2000**	2006	
prognostic factors	2000	49.8535	**2000**	2005	
linkage disequilibrium	2000	41.9375	**2000**	2006	
ionizing-radiation	2000	39.2227	**2000**	2005	
bone-marrow transplantation	2000	37.6099	**2000**	2003	
double-strand breaks	2000	37.5917	**2000**	2006	
linkage	2000	36.4447	**2000**	2005	
saccharomyces-cerevisiae	2000	31.0545	**2000**	2004	
gene	2000	29.8691	**2000**	2002	
nijmegen breakage syndrome	2000	27.5811	**2000**	2006	
dna-damage response	2000	26.4568	**2000**	2004	
smooth-muscle cells	2000	26.1191	**2000**	2003	
s-phase	2000	23.5414	**2000**	2004	
cell-cycle	2000	22.0682	**2000**	2005	
malignant-lymphoma	2000	21.0953	**2000**	2004	
cloning	2000	20.606	**2000**	2004	
localization	2000	20.2864	**2000**	2003	
protein	2000	20.0571	**2000**	2003	
aids	2000	18.7765	**2000**	2005	
potassium channel	2000	17.8574	**2000**	2006	
coronary heart-disease	2000	17.7765	**2000**	2004	
molecular-cloning	2000	17.0402	**2000**	2003	
potassium channel gene	2000	16.9104	**2000**	2002	
s-phase checkpoint	2000	38.0409	**2002**	2007	
crohns-disease	2000	28.8567	**2002**	2006	
single-nucleotide polymorphisms	2000	20.5066	**2002**	2006	
dermopathy	2000	49.8578	**2007**	2009	
gadolinium	2000	39.1106	**2007**	2009	
nephrogenic systemic fibrosis	2000	40.1299	**2008**	2009	
genome-wide association	2000	38.7906	**2011**	2014	

Figure 4. Keywords with the strongest frequency bursts in the expanded dataset.

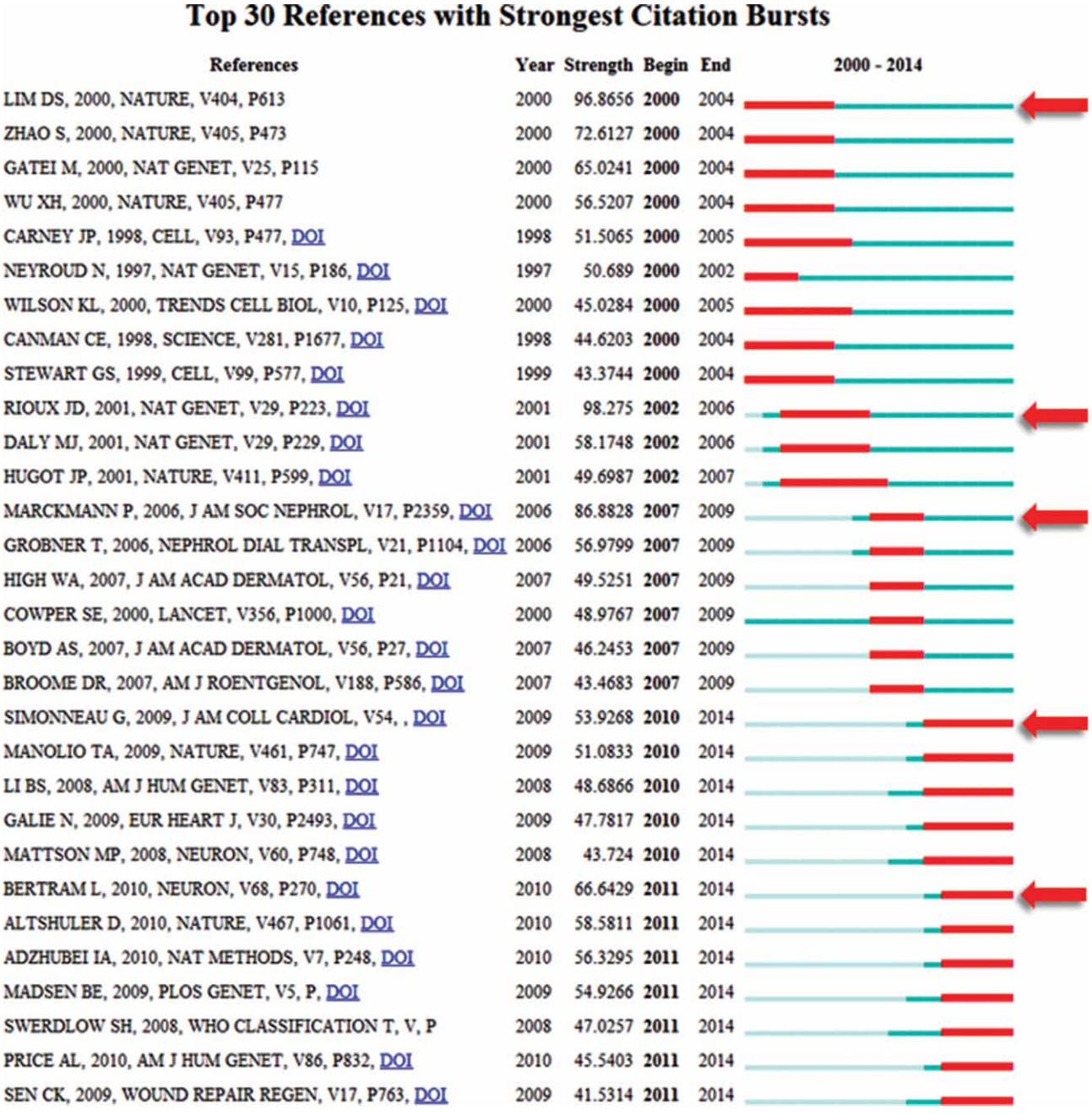

Figure 5. Top 30 cited references with the strongest citation bursts in the expanded dataset.

Nodes in the network represent cited references. The size of a node in the visualized network is proportional to the number of citations received by the cited reference. Red rings of a node depict that citation bursts are detected for the reference.

Lines between nodes represent co-citation links. The colors of links denote when a particular connection was made for the first time. Blue colors indicate the earliest connections, whereas orange colors indicate the most recently made connections. The labels starting with the # and a number characterize the impact of the contributions in the cluster on subsequent research. For example, #0 ATM suggests that the first cluster's impact is primarily on ATM-related research, that is, ataxia-telangiectasia mutated (ATM).

Citation counts restricted to a particular subset of the Web of Science are known as local citations, as opposed to the global citations across the entire Web of Science. Local and global citations essentially agree with each other on the central topics but may differ on peripheral topics. **Figure 7** shows the same underlying network as **Figure 6** but labels articles with

Table 3. Representative references with the strongest citation bursts in the same-year groups.

Study	Citation burst		
	Strength	Begin	End
Lim *et al.* 2000 [23]	96.87	2000	2004
Rioux *et al.* 2001 [24]	98.28	2002	2006
Marckmann *et al.* 2006 [25]	86.88	2007	2009
Simonneau *et al.* 2009 [26]	53.93	2010	2014
Bertram *et al.* 2010 [27]	66.64	2011	2014

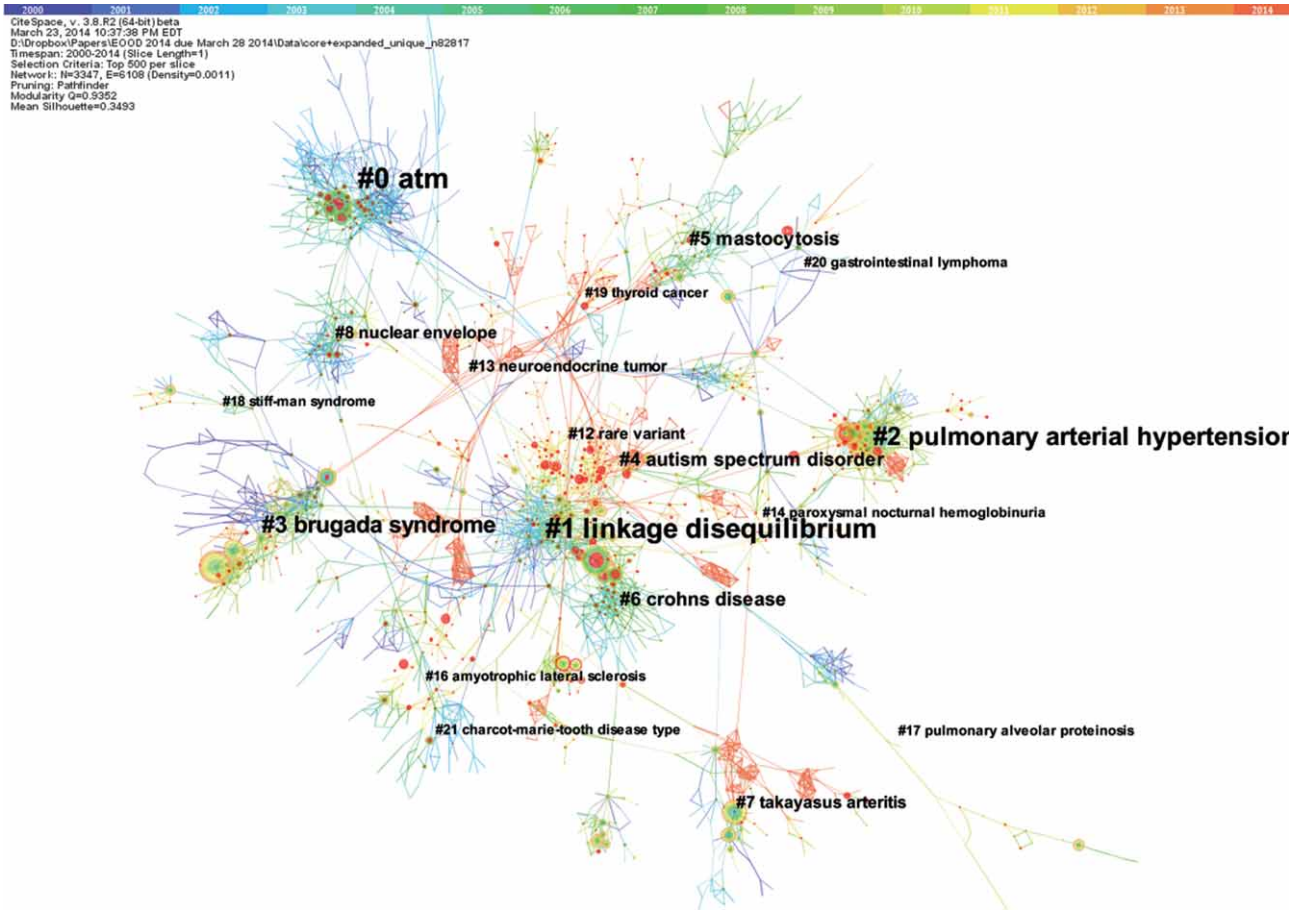

Figure 6. A document co-citation network derived from the expanded dataset.

the highest local citations. These highly cited articles are often considered to be landmark articles in the bibliographic landscape of a subject matter. For example, the reference with the highest citations is an article published in 2005 by Antzelevitch *et al.* [28]. It has a local citation count of 600 within the expanded dataset. It belongs to cluster #3 Brugada syndrome. The work tends to be cited by research relevant to Brugada syndrome. The second landmark reference is the 2001 article by Rioux *et al.* [24] in cluster #6 on Crohn's disease. It has 574 local citations. The third one, Lim 2000, has a local citation count of 548. It is a member of the largest cluster #0 on ATM. Humbert 2006 is a landmark paper for PAH research. Table 4 shows the 10 most prominent citation landmark references.

4.4 Thematic concentrations in the landscape

Each cluster represents a thematic concentration in the bibliographic landscape. The nature of each cluster can be understood from two aspects. First, each cluster is a group of tightly coupled references. The group itself is known as the intellectual base of a research specialty. Second, articles that are responsible for the citation patterns are known as

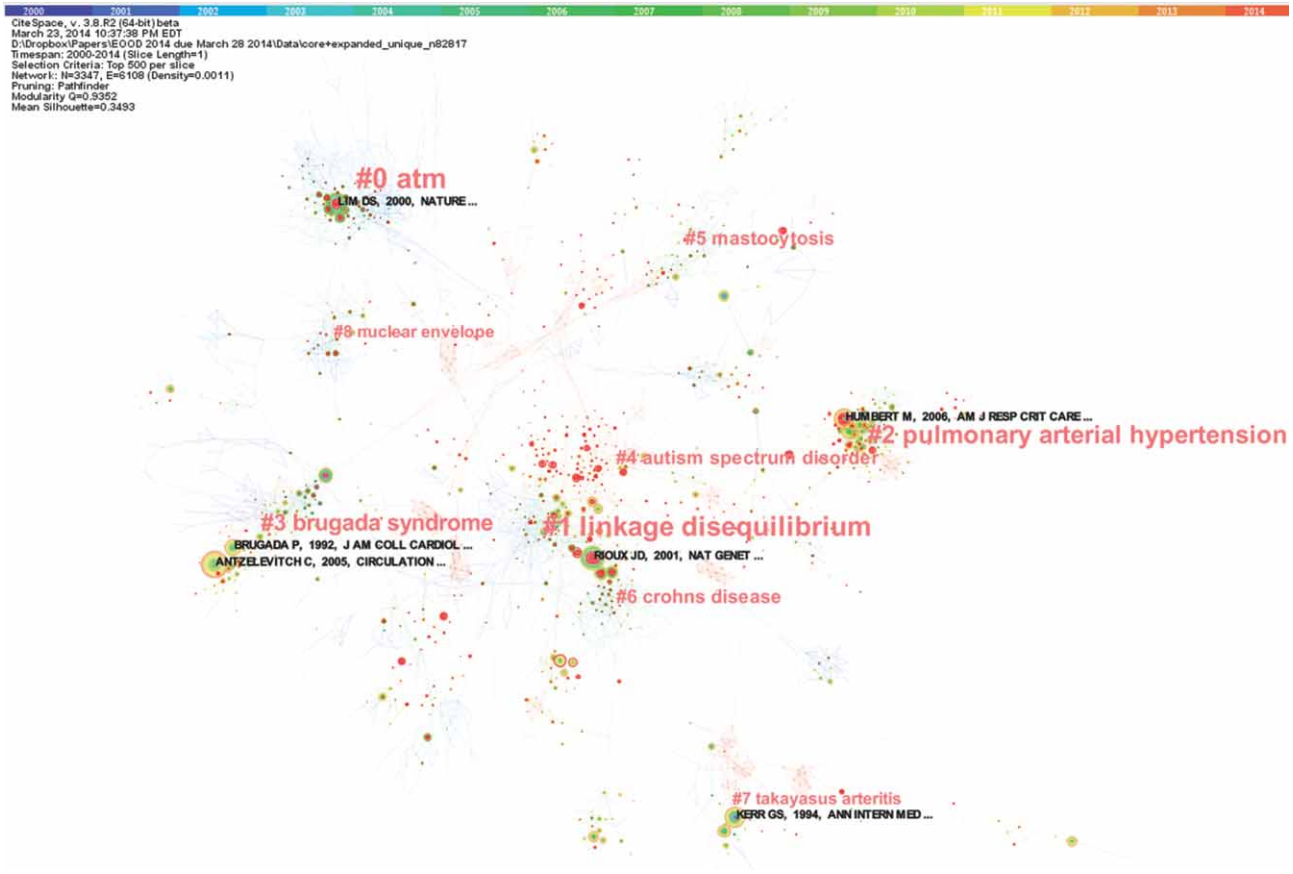

Figure 7. An overview of a document co-citation network derived from the expanded dataset.

Table 4. Landmark references by local citation counts.

Citation Count	Study	Cluster
600	Antzelevitch *et al.* 2005 [28]	#3 Brugada syndrome
574	Rioux *et al.* 2001 [24]	#6 Crohn's disease
548	Lim *et al.* 2000 [23]	#0 ATM
497	Kerr *et al.* 1994 [30]	#7 Takayasu's arteritis
486	Humbert *et al.* 2006 [31]	#2 Pulmonary arterial hypertension
476	Marckmann *et al.* 2006 [25]	#11 Nephrogenic systemic fibrosis
428	Brugada *et al.* 1992 [32]	#3 Brugada syndrome
393	D'Alonzo *et al.* 1991 [33]	#2 Pulmonary arterial hypertension
381	Neyroud *et al.* 1997 [34]	#3 Brugada syndrome
352	Hugot *et al.* 2001 [35]	#6 Crohn's disease

ATM: Ataxia-telangiectasia, mutated.

the research front of the research specialty. A few prominent clusters emerged from the expanded dataset will be inspected and the ways in which they are related to orphan drugs and rare diseases research will be identified.

The expanded dataset yielded 375 clusters of co-cited references. Table 5 summaries the largest 10 clusters. The silhouette value of a cluster is a measure of the cluster's homogeneity. The average age of its members is the average year of member references' publication date. The label of a cluster summarizes the impact of the cluster on more recent research. The most recently emerged topic is captured by cluster #4 – *autism spectrum disorder* with an average year of 2007. *Pulmonary arterial hypertension* (#2) and *Crohn's disease* (#6) are two relatively young areas. Topics of *ATM* and *linkage disequilibrium*, the largest and the second largest clusters, are both dated 1999 on average.

4.4.1 *Ataxia-telangiectasia, mutated*

The most prominent thematic concentration is research that is concerned with rare disorder ataxia-telangiectasia (ATM) and NBS, which is also a rare disorder. The theme is supported by the largest cluster (#0) of 248 cited references with a high level of homogeneity (a silhouette value of 0.98).

The ATM refers to the gene that is responsible for the rare autosomal-recessive inherited disorder. It has a complex clinical phenotype. NBS is a rare disorder that causes chromosomal instability. The most cited references in this cluster are related either to ATM or to NBS, or both.

Table 5. A summary of thematic concentrations emerged from the expanded dataset.

Cluster #	Size	Silhouette	Label (log-likelihood ratio)	Average age of members
0	248	0.98	ATM	1999
1	221	0.88	Linkage disequilibrium	1999
2	171	0.99	Pulmonary arterial hypertension	2002
3	167	0.96	Brugada syndrome	2001
4	115	0.94	Autism spectrum disorder	2007
5	112	0.95	Mastocytosis	2001
6	111	0.98	Crohn's disease	2002
7	97	0.95	Takayasu's arteritis	2000
8	96	0.97	Nuclear envelope	2000
9	87	1.00	Fabry disease	2001

ATM: Ataxia-telangiectasia, mutated.

Table 6. Most cited references in the ATM conconcentration.

Citations	Burst	Study
548	99.52	Lim *et al.* 2000 [23]
285	74.01	Zhao *et al.* 2000 [36]
266	66.33	Gatei *et al.* 2000 [37]
248	57.75	Wu *et al.* 2000 [38]
202	53.06	Carney *et al.* 1998 [39]

ATM: Ataxia-telangiectasia, mutated.

Table 7. Articles that shape the cluster.

Coverage %	Study
23	Kastan *et al.* 2000 [40]
23	Kerr *et al.* 1994 [30]; Zhou and Elledge 2000 [60]
17	Zheng *et al.* 2000 [41]
15	Shiloh 2000 [42]
15	Shiloh 2000 [42]; Schultz 2000 [61]

Articles that are essentially built on the intellectual base formed by the cluster provide rich information of major research interests associated with the concentration of co-cited references. Central concepts extracted from these articles include *ATM*, *mutation*, *DNA damage* and *DNA double-strand break* (DSB). These terms correspond to major branches in concept trees constructed by CiteSpace from the unstructured text of titles and abstracts of the citing articles, that is, articles that cited members of the cluster. More specific concepts include *dysfunction of ATM*, *damaged DNA*, such as *UV-damaged* and *radiation-damaged DNA*, *cellular response to DNA DSBs*.

CiteSpace allows the user to interact with a visualized concept tree so that concrete sentences relevant to a specific concept can be presented as the user moves the mouse over various concepts. The following two sentences from the abstracts of two distinct records are obtained by moving the mouse over the concept of ATM on a visualized concept tree in CiteSpace. The wos numbers are unique identifiers of the records in the Web of Science.

- [wos:000090011900003] AT results from mutations of ATM and is characterized by severe neurodegeneration and defective responses to DNA damage.
- [wos:000281161100002] multiple roles of ATM in monitoring and maintaining DNA integrity.

Table 6 shows five most cited references in the ATM concentration. As their titles indicate, these articles address various issues concerning ATM and NBS.

Table 7 lists articles that have shaped the cluster in that these articles' citation behavior determined the groupings of the cited references. The percentage of coverage means the percentage of member references cited by a citing article. The titles of these citing articles confirm the central topic of ATM, especially in terms of cellular responses to DNA damages.

4.4.2 *Linkage disequilibrium*

LD refers to the nonrandom associations of alleles at different loci in genetic variation. This is the second largest topic revealed by the expanded dataset. Its intellectual base is represented by 221 members of cluster #1 with a slightly lower homogeneity of 0.88.

Major concepts in the concept tree of this cluster include *genes*, *chromosome* and *patients*. In particular, phrases such as *several genes*, *a number of genes*, *many genes* and *multiple genes* most frequently appeared in this cluster, suggesting this thread may have a different focus from genetic variation of a single gene. Other popular terms of this cluster include *GWAS*, *genetic associations* and *evidence of association*. The following excerpt illustrates a typical context of these terms.

- [wos:000280922800004] Although they have demonstrated success in searching for common variants for complex diseases, genome-wide association (GWA) studies are less successful in detecting rare genetic variants because of the poor statistical power of most of current methods.

Table 8. Most cited references in the linkage disequilibrium's intellectual base.

Citation	Burst	Study
247	38.07	Risch and Merikangas 1996 [43]
237	40.87	Purcell *et al.* 2007 [44]
230	8.14	Barrett *et al.* 2005 [45]
219	10.22	Lander *et al.* 2001 [46]
218	11.22	Peltonen *et al.* 2000 [29]

Table 9. Citing articles to the linkage disequilibrium cluster.

Coverage %	Study
18	Jorde 2000 [47]
12	Peltonen *et al.* 2000 [29]
11	Jorde *et al.* 2000 [48]
10	Risch 2000 [49]
9	Terwilliger and Goring 2000 [50]

Table 10. Most cited references in the PAH cluster.

Citation	Burst	Study
486	35.65	Humbert *et al.* 2006 [31]
393		D'alonzo *et al.* 1991 [33]
306	4.15	Rubin *et al.* 2002 [51]
270		Barst *et al.* 1996 [52]
243		Rich *et al.* 1987 [53]
220	3.34	Channick *et al.* 2001 [54]
217	52.87	Simonneau *et al.* 2009 [26]

PAH: Pulmonary arterial hypertension.

Table 11. Citing articles to the PAH cluster.

Coverage %	Study
23	Anderson and Nawarskas 2010 [55]
21	Girgis 2010 [56]
18	Galie *et al.* 2010 [57]
18	Burt *et al.* 2010 [58]
17	Reis *et al.* 2010 [59]

PAH: Pulmonary arterial hypertension.

Table 8 lists five most cited references in the LD thread. The relevance of these individual references is reflected in their titles, for example, through terms such as *genetic studies*, *whole-genome association*, *linkage analysis* and *complex traits*.

Table 9 shows five representative articles from the research front of LD. In other words, these articles are built on the LD cluster. Note Peltonen's 2000 article [29] appears in both **Tables 8** and **9**. Thematic connections are also evident from the titles, for example, *search for complex disease genes*, and *mapping complex traits.*

4.4.3 *Pulmonary arterial hypertension*

PAH is an orphan disease with an estimated prevalence of up to 15 in a million. It is characterized by high blood pressure of the main artery of the lungs for no apparent reason. PAH is the third largest topic. It is defined by cluster #2's 171 members.

In 2009, Simonneau *et al.* published an updated clinical classification of pulmonary hypertension [26]. The article has a very strong citation burst of 52.87 (**Table 10**). In fact, its citation burst ranked among the top 30 strongest citation bursts of the entire expanded dataset. In contrast, the most cited reference in this cluster has the citation burst of 35.65. The initial version of the clinical classification was featured in the analysis of the core dataset. The themes of this cluster are highly consistent with the term *pulmonary arterial hypertension* or *pulmonary hypertension* in the titles of the five most active citers to the cluster.

Table 11 lists five articles that cited the most references in the PAH cluster. The titles indicate a strong concentration on the topic of PAH.

Other clusters identified in the bibliographic landscape include a Brugada syndrome cluster (#3), autism spectrum disorder (#4), mastocytosis (#5) and Crohn's disease (#6).

The Brugada syndrome is a genetic disease with an increased risk of sudden cardiac death. The Brugada syndrome cluster has been mostly cited by articles focusing on Brugada, long QT and arrhythmic syndromes. The *autism spectrum disorder* cluster is associated with terms such as *cognitive impairment* and *intellectual disability*. Mastocytosis is a rare disease characterized by a primary pathological increase in mast cells in different tissues, which may present in a variety of clinical patterns. Crohn's disease is a type of chronic inflammatory bowel disease (IBD), commonly found in European and East Asian countries.

5. Expert opinions

The scientometric study has identified three broad categories of research in the study of orphan drugs and orphan diseases. The first category focuses on research policy issues and the strategic role of research on orphan drugs and rare diseases in a wide-ranging societal context, including optimizing the allocation of resources and improving evaluation standards that are particularly suitable for orphan drugs and rare diseases. The second category is basic research in nature. It focuses on understanding the complex biological and genetic mechanisms of rare diseases. The third category is primarily disease-specific research with clinical and therapeutic focuses.

The research policy category studies how scientific, clinical, educational and other aspects of the study of orphan drugs and rare diseases should be developed, evaluated and optimized. The unique characteristics of rare diseases have a

wide variety of implications, from the design of clinical trials, methodologies of data analysis, to evaluation standards, and many other dimensions. Orphan drugs and rare diseases present unique challenges and opportunities to translational medicine.

The basic research category investigates fundamental questions about the complex mechanisms of a rare disorder. For example, what are the types of genetic variation and environmental risk factors that can cause a rare disease? What would it take to be able to predict so far unseen rare diseases? The development of increasingly advanced enabling techniques such as methods of reliably analyzing small-sized samples and populations or methods of studying genome-wide associations also belongs to the category.

The clinical and therapeutic category is a critical part of the long-value chain of healthcare. So far, the scientometric study has not identified any strong presence of translational research in the literature of organ drugs and rare diseases.

The research policy category is prominent in the bibliographic landscape generated from the core dataset. In hindsight, this is not a surprise considering the way the core dataset was constructed. Given that the core dataset resulted from a topic search, all publications in the core dataset explicitly contain terms such as orphan drugs and rare diseases. These conceptually broader terms suggest that research activities are collectively and categorically at the center of these publications' primary focus. In contrast, conceptually narrower terms are more commonly seen in the expanded dataset, for example, ultraviolet (UV)-damaged DNA, GWA and LD.

In the basic research category, the appearances of conceptually broad terms such as orphan drugs and rare diseases are less prominent than they are in the research policy category. Instead, prominent terms are conceptually narrower, that is, corresponding to specific mechanisms of a rare disease. For example, the largest three thematic concentrations identified in the expanded dataset are ATM, LD and PAH, whereas the largest one in the core dataset is Orphan Drug, which is at a conceptually higher level of abstraction. Furthermore, the more focused themes are associated with concepts at even finer grained levels of granularity such as cellular responses to UV-damaged DNA and methodological terms such as GWAS.

The clinical and therapeutic category is also largely revealed by the study of the expanded dataset. While the footprint of the basic research category is essentially at the center of the intellectual landscape, research in the clinical and therapeutic category tends to be distributed either in citation valleys between those more prominent clusters of the basic research category or in peripheral areas of the landscape. For example, the Crohn's disease cluster is like a small satellite town in the suburbs of the big city of the LD cluster. Ontologically speaking, research on LD is conceptually more fundamental than the study of Crohn's disease because findings in the former would influence the latter more than the other way round.

Our survey has revealed the intellectual structure of the research landscape relevant to orphan drugs and rare diseases. It also raises some questions that our methodology is unlikely to be able to answer. For example, what are historical, scientific and pragmatic reasons behind the current size of each cluster? Are they the result of policy intervention, the consequence of the pursuit of a curiosity-driven research agenda or motivated by the intent to find a cure of a specific rare disease? Why does research in ATM have the largest footprint, about twice as big as the next largest cluster? Why some rare diseases have apparently attracted more research attention than others? Addressing these questions is relevant to all three current categories of research because individual topics are no longer considered in isolation to other topic areas. One area of expertise might hold the key to solve a technical solution in another.

Our study of the literature has confirmed the role of genetic variation as a fundamental common ground that sustains investigations of the diverse range of rare but complex phenotypic patterns. Furthermore, our analysis has revealed the diversity of orphan drugs and rare diseases and the potential barrier for disseminating findings and technical advances across various specialties. Finally, the study has analyzed both original research articles and review papers, but more comprehensive review papers that cover all three categories would be welcomed. Reviews that consider fruitful connections between rare diseases and common diseases would be welcomed too. Our recommendation is, therefore, that more comprehensive review papers would be beneficial to identify the achievements of orphan drugs and rare diseases research, the common ground of numerous research threads within the field, and fundamental challenges that should be addressed and coordinated in an even broader context, for example, including the study of common diseases.

Declaration of interest

The author has no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Bibliography

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Arno PS, Bonuck K, Davis M. Rare diseases, drug development, and AIDS: the impact of the Orphan Drug Act. Milbank Q 1995;73(2):231-51
2. Dear JW, Lilitkarntakul P, Webb DJ. Are rare diseases still orphans or happily adopted? The challenges of developing and using orphan medicinal products. Br J Clin Pharmacol 2006;62(3):264-71
3. Chen CM, CiteSpace II. Detecting and visualizing emerging trends and transient patterns in scientific literature. J Am Soc Inf Sci Technol 2006;57(3):359-77
4. Chen C, Hu Z, Liu S, Tseng H. Emerging trends in regenerative medicine: a scientometric analysis in CiteSpace. Expert Opin Biol Ther 2012;12(5):593-608
5. Chen CM, Ibekwe-SanJuan F, Hou JH. The structure and dynamics of cocitation clusters: a multiple-perspective cocitation analysis. J Am Soc Inf Sci Technol 2010;61(7):1386-409
6. Garfield E. Citation indexes for science: a new dimension in documentation through association of ideas. Science 1955;122(3159):108-11
7. Small HG. A co-citation model of a scientific specialty: a longitudinal study of collagen research. Soc Stud Sci 1977;7:139-66
8. Kleinberg J. Bursty and hierarchical structure in streams. Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 23 – 26 July 2002; ACM Press, Edmonton, Alberta, Canada; 2002. p. 91-101, DOI:10.1145/775047.775061
9. Chen C. Hindsight, insight, and foresight: a multi-level structural variation approach to the study of a scientific field. Technol Anal Strateg Manage 2013;25(6):619-40
10. Cheung RY, Cohen JC, Illingworth P. Orphan drug policies: implications for the United States, Canada, and developing countries. Health Law J 2004;12:183-200
11. Rinaldi A. Adopting an orphan. EMBO Rep 2005;6(6):507-10
12. McCabe C, Tsuchiya A, Claxton K, Raftery J. Orphan drugs revisited. QJM 2006;99(5):341-5
13. Haffner ME, Whitley J, Moses M. Two decades of orphan product development. Nat Rev Drug Discov 2002;1(10):821-5
14. Fentiman IS, Fourquet A, Hortobagyi GN. Male breast cancer. Lancet 2006;367(9510):595-604
15. Salvadori B, Saccozzi R, Manzari A, et al. Prognosis of breast cancer in males: an analysis of 170 cases. Eur J Cancer 1994;30A:930-5
16. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958;53(282):457-81

•• **This is an extremely highly cited methodology paper on nonparametric estimation from incomplete observations.**

17. Simonneau G, Galiè N, Rubin LJ, et al. Clinical classification of pulmonary hypertension. J Am Coll Cardiol 2004;43(12 Suppl S):5S-12S
18. Buckley BM. Clinical trials of orphan medicines. Lancet 2008;371(9629):2051-5

•• **This paper on clinical trials of orphan medicines addresses unique challenges to the field of study. It has a strong citation burst from 2009.**

19. Drummond MF, Wilson DA, Kanavos P, et al. Assessing the economic challenges posed by orphan drugs. Int J Technol Assess Health Care 2007;23(1):36-42

• **This is a representative research policy paper.**

20. Griggs RC, Batshaw M, Dunkle M, et al. Clinical research for rare disease: opportunities, challenges, and solutions. Mol Genet Metab 2009;96(1):20-6

• **This paper is suitable for researchers considering their career choices in connection to the study of rare diseases. It has a strong citation burst since 2011.**

21. Kesselheim AS, Myers JA, Avorn J. Characteristics of clinical trials to support approval of orphan vs nonorphan drugs for cancer. J Am Med Assoc 2011;305(22):2320-6
22. Paul SM, Mytelka DS, Dunwiddie CT, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nat Rev Drug Discov 2010;9:203-14
23. Lim DS, Kim ST, Xu B, et al. ATM Phosphorylates p95/nbs1 in an S-phase checkpoint pathway. Nature 2000;404(6778):613-17

•• **This is a representative basic research paper on two rare diseases in terms of genetic mechanisms.**

24. Rioux JD, Daly MJ, Silverberg MS, et al. Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. Nat Genet 2001;29(2):223-8

•• **This is a representative basic research paper on identifying the genetic linkage of Crohn's disease using linkage disequilibrium mapping.**

25. Marckmann P, Skov L, Rossen K, et al. Nephrogenic systemic fibrosis: suspected causative role of gadodiamide used for contrast-enhanced magnetic resonance imaging. J Am Soc Nephrol 2006;17(9):2359-62
26. Simonneau G, Robbins IM, Beghetti M, et al. Updated clinical classification of pulmonary hypertension. J Am Coll Cardiol 2009;54(1 Suppl):S43-54

• **This is a representative disease-specific paper on pulmonary hypertension, providing an updated clinical classification of pulmonary hypertension.**

27. Bertram L, Lill CM, Tanzi RE. The genetics of Alzheimer disease: back to the future. Neuron 2010;68(2):270-81

•• **This is a representative basic research paper on the role of genome-wide association studies in genetic research of Alzheimer's disease.**

28. Antzelevitch C, Brugada P, Borggrefe M, et al. Brugada syndrome: report of the second consensus conference: endorsed by the Heart Rhythm Society and the European Heart Rhythm Association. Circulation 2005;111(5):659-70

• **This is a representative disease-specific paper on Brugada syndrome. It has the highest citations in the expanded dataset of over 80,000 papers.**

29. Peltonen L, Palotie A, Lange K. Use of population isolates for mapping complex traits. Nat Rev Genet 2000;1(3):182-90
30. Kerr GS, Hallahan CW, Giordano J, et al. Takayasu arteritis. Ann Intern Med 1994;120(11):919-29

31. Humbert M, Sitbon O, Chaouat A, et al. Pulmonary arterial hypertension in France: results from a national registry. Am J Respir Crit Care Med 2006;173(9):1023-30

32. Brugada P, Brugada J. Right bundle branch block, persistent ST segment elevation and sudden cardiac death: a distinct clinical and electrocardiographic syndrome. A multicenter report. J Am Coll Cardiol 1992;20(6):1391-6

33. D'Alonzo GE, Barst RJ, Ayres SM, et al. Survival in patients with primary pulmonary hypertension. Results from a national prospective registry. Ann Intern Med 1991;115(5):343-9

34. Neyroud N, Tesson F, Denjoy I, et al. A novel mutation in the potassium channel gene KVLQT1 causes the Jervell and Lange-Nielsen cardioauditory syndrome. Nat Genet 1997;15(2):186-9

35. Hugot JP, Chamaillard M, Zouali H, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 2001;411(6837):599-603

36. Zhao S, Weng YC, Yuan SS, et al. Functional link between ataxia-telangiectasia and nijmegen breakage syndrome gene products. Nature 2000;405(6785):473-7

37. Gatei M, Young D, Cerosaletti KM, et al. ATM-dependent phosphorylation of nibrin in response to radiation exposure. Nat Genet 2000;25(1):115-19

38. Wu X, Ranganathan V, Weisman DS, et al. ATM phosphorylation of Nijmegen breakage syndrome protein is required in a DNA damage response. Nature 2000;405(6785):477-82

39. Carney JP, Maser RS, Olivares H, et al. The hMre11/hRad50 protein complex and Nijmegen breakage syndrome: linkage of double-strand break repair to the cellular DNA damage response. Cell 1998;93(3):477-86

40. Kastan MB, Lim DS. The many substrates and functions of ATM. Nat Rev Mol Cell Biol 2000;1(3):179-86

41. Zheng L, Li S, Boyer TG, Lee WH. Lessons learned from BRCA1 and BRCA2. Oncogene 2000;19(53):6159-75

42. Shiloh Y. ATM: sounding the double-strand break alarm. Cold Spring Harb Symp Quant Biol 2000;65:527-33

43. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996;273(5281):1516-17

44. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81(3):559-75

45. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21(2):263-5

46. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. Nature 2001;409(6822):860-921

47. Jorde LB. Linkage disequilibrium and the search for complex disease genes. Genome Res 2000;10(10):1435-44

48. Jorde LB, Watkins WS, Kere J, et al. Gene mapping in isolated populations: new roles for old friends? Hum Hered 2000;50(1):57-65

49. Risch NJ. Searching for genetic determinants in the new millennium. Nature 2000;405(6788):847-6

50. Terwilliger JD, Göring HH. Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. Hum Biol 2000;72(1):63

51. Rubin LJ, Badesch DB, Barst RJ, et al. Bosentan therapy for pulmonary arterial hypertension. N Engl J Med 2002;346(12):896-903

52. Barst RJ, Rubin LJ, Long WA, et al. A comparison of continuous intravenous epoprostenol (prostacyclin) with conventional therapy for primary pulmonary hypertension. N Engl J Med 1996;334(5):296-301

53. Rich S, Dantzker DR, Ayres SM, et al. Primary pulmonary hypertension. A national prospective study. Ann Intern Med 1987;107(2):216-23

54. Channick RN, Simonneau G, Sitbon O, et al. Effects of the dual endothelin-receptor antagonist bosentan in patients with pulmonary hypertension: a randomised placebo-controlled study. Lancet 2001;358(9288):1119-23

55. Anderson JR, Nawarskas JJ. Pharmacotherapeutic management of pulmonary arterial hypertension. Cardiol Rev 2010;18(3):148-62

56. Girgis RE. Emerging drugs for pulmonary hypertension. Expert Opin Emerg Drugs 2010;15(1):71-85

57. Galiè N, Palazzini M, Leci E, Manes A. Current therapeutic approaches to pulmonary arterial hypertension. Rev Esp Cardiol 2010;63(6):708-24

58. Burt C, Pepke-Zaba J, Falter F. Pulmonary arterial hypertension. Curr Vasc Pharmacol 2010;8(3):412-20

59. Reis A, Rocha N, Barros R, et al. Guidelines for the management of pulmonary hypertension patients. Rev Port Cardiol 2010;29(2):253-89

60. Zhou BB, Elledge SJ. The DNA damage response: putting checkpoints in perspective. Nature 2000;408:433-9, DOI:10.1038/35044005

61. Schultz LB. p53 binding protein 1 (53bp1) is an early participant in the cellular response to DNA double-strand breaks. J Cell Biol 2000;151:1381-90, DOI:10.1083/jcb.151.7.1381

Affiliation

Chaomei Chen[†], Rachael Dubin & Meen Chul Kim
[†]Author for correspondence
Drexel University, College of Computing and Informatics, 3141 Chestnut Street, Philadelphia, PA 19104-2875, USA
E-mail: chaomei.chen@drexel.edu