Current status and thinking of gene editing technology

Junfeng Yu¹,²,³, Yanqing Zhao¹,², Haizhou Feng¹,², Jiazeng Yu⁴,* and Tie Li⁵

¹ Qilu University of Technology (Shandong Academic of Sciences), Jinan, China
² Information Research Institute of Shandong Academy of Sciences, Jinan, China
³ Institute of Telecommunication Systems of National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine
⁴ Ruiha Affiliated Hospital of Soochow University, Suzhou, China
⁵ Shandong Normal University, Jinan, China

*Corresponding author. Email: 120652688@qq.com

Abstract. Gene editing refers to operations such as deleting, replacing, and inserting target genes to obtain new functions or phenotypes, and even create new species. As an important research field with the rapid development of life sciences, the development and application of gene editing technology has made the genetic modification of organisms into an unprecedented depth and breadth. Through the analysis of patent documents related to gene editing, the article reveals the development status of gene editing technology in terms of global output trends, patent application status of major institutions, life cycle, technical composition, and geographical distribution. It also puts forward countermeasures and suggestions in basic research, intellectual property protection, international exchanges and cooperation, and policy supervision to provide references for researchers and decision makers in related fields.

Keywords: gene editing, ZNF, TALENs, CRISPR, bibliometric.

1. Introduction
Gene editing, also known as genome editing or genome engineering, is an emerging and more precise genetic engineering technology or process that can modify specific target genes in the genome of an organism. In short, gene editing technology is a technology for inserting, deleting, and replacing gene sequences containing genetic information. Gene editing technology uses nuclease, an enzyme that cuts a specific DNA sequence in cells, to delete, repair or replace specific disease-related genes. These technologies include homing endonucleases (or meganucleases), zinc finger nucleases (ZNF), transcription activator-like effector nucleases (TALEN) and cluster regularly interspaced short palindromic repeats (CRISPR). The principle lies in: using artificial endonucleases to precisely target and specifically cut DNA double-strands to induce site-specific DNA double-strand breaks (DSB). After DSB is produced, activated cells initiate two main natural repair mechanisms: non-homologous end joining (NHEJ) and homologous recombination repair (HDR).

Normally, cells are mainly repaired through NHEJ. In the process of repairing and reconnecting broken DNA, NHEJ can generate random insertions or indels of bases at the DSB site, often causing frameshift mutations to inactivate genes, thereby achieving the knockout of target gene. If an exogenous donor gene sequence exists, the NHEJ mechanism will link it to the DSB site of the double-strand break,
thereby realizing targeted exogenous gene knock-in. When a recombination donor with a homologous arm exists, the cell will also use HDR to repair the DSB. The foreign target gene in the donor will be completely integrated into the target site through the homologous recombination process to achieve accurate insertion, deletion or base substitution at a specific site, without random base insertion or loss [1-2].

Tracing back to its history, the chronicle of gene editing of the American Society for Gene and Cell Therapy (ASGCT) began more than 30 years ago. The first gene editing was done in yeast cell experiments. The development of artificial nuclease technology makes it a reality to induce DSBs artificially, and zinc finger nucleases (ZFNs) is a milestone breakthrough, also known as the first-generation gene editing technology [3]. ZFN technology has led gene editing into an era that no longer relies solely on naturally occurring DSBs, but it has great limitations such as high cost and difficulty in achieving multi-target editing [4]. The discovery of TALE (transcription activator-like effector) motifs gave birth to the second generation of gene editing technology-TALENs (TALE nucleases). In 2012, Doudna Jennifer and Emmanuelle Charpentier transformed CRISPR/Cas9 into a low-cost tool that can efficiently modify specific DNA sequences of target genes. The advent of CRISPR/Cas9 marks the advent of a new generation of biological core technology and makes it possible to precisely control the life process. The upsurge of genome editing has swept the entire field of life science research, and life science research has opened the era of genome editing [5].

Through the analysis of patent documents related to gene editing, the article reveals the development status of gene editing technology in terms of global output trends, patent application status of major institutions, life cycle, technical composition, and geographical distribution. It also puts forward countermeasures and suggestions in basic research, intellectual property protection, international exchanges and cooperation, and policy supervision to provide references for researchers and decision makers in related fields.

2. Data selection and method
The patent search database related to gene editing is the incoPat patent database. incoPat deeply integrates the global invention wisdom, complete collection of more than 100 million basic patent data from 120 countries and regions around the world, and collects and processes patent data from 22 major countries with diverse fields and complete data. The earliest patent on gene editing appeared in 1989. The 199 patent (US781559), that is, "cultured cells modified with target DNA sequences in the genome are used to produce transgenic animals or to regulate plant cells, including the transformation of cells with positive and negative selection vectors." The earliest priority can be traced back to 1989 (US397707) [6]. In the following 10 years, the number of relevant patent applications per year was less than 10; since 2003, the number of patent applications has increased year by year[4]. Therefore, the search time range chosen in this article is from January 1, 2001 to December 31, 2020. The subject search was carried out in the patent database with terms such as gene editing, genome editing, genome engineering, ZFN, TALENs, CRISPR, transcription activator-like effector nucleases, homing endonucleases, clustered regularly interspaced short palindromic repeats, etc. The search time was February 9, 2021. The number of retrieved records is 14137, and 11665 after the combination of application numbers.

3. Analysis of patent documents of gene editing technology

3.1. Trend analysis

3.1.1. Application trend. The chart shows the development trend for patents applications, through which you can get an idea of application changes in different periods of the analysis object from the macro level. Normally, invention patents will be published 3-18 months later than application filed and utility models and design patents will be published 6 months later than application filed. It can be seen from Figure 1 that before 2010, the number of patents each year was less than 100. After 2012, the number of applications began to grow rapidly, reaching 2,327 in 2018. Since patent applications are
granted after an 18-month publication period; therefore, the data for 2019 and 2020 cannot represent all the data for that year.

Figure 1. Application trend.

3.1.2. Life cycle. Life cycle analysis is normally the most commonly used method in patents quantity analysis, through which you could predict the future technology developing direction by analysing patents current technology developing period. The object could be the life cycle of a technology represented by a specific patent or the whole technology life cycle of a technical area. This article describes its life cycle in the field of gene editing technology. From Figure 2, it can be seen that its slow development before 2012 is in its infancy, and it has reached a rapid development stage from 2012 to 2018. Since patent applications are granted after an 18-month publication period; therefore, the data for 2019 and 2020 cannot represent all the data for that year.

Figure 2. Life cycle.
3.2. Technology analysis

3.2.1. Technology construction (IPC sub-category). The chart shows the quantity distribution of various technology directions of the analysis object through which you can learn the covered technology classifications of the analysis object and the innovation intensity of various technology branches. This article selects the top 10 of the IPC sub-category for analysis. From Figure 3, it can be seen that C12N involves the largest number of patents, 9351; followed by A61K, involving 2599 patents; and ranking third is C07K involving 1358 patents. The meaning of these three subcategories is explained as follows:

C12N: MICROORGANISMS OR ENZYMES; COMPOSITIONS THEREOF (biocides, pest repellents or attractants, or plant growth regulators containing microorganisms, viruses, microbial fungi, enzymes, fermentates, or substances produced by, or extracted from, microorganisms or animal material A01N63/00; medicinal preparations A61K; fertilisers C05F); PROPAGATING, PRESERVING, OR MAINTAINING MICROORGANISMS; MUTATION OR GENETIC ENGINEERING; CULTURE MEDIA (microbiological testing media C12Q1/00)

A61K: PREPARATIONS FOR MEDICAL, DENTAL, OR TOILET PURPOSES (devices or methods specially adapted for bringing pharmaceutical products into particular physical or administering forms A61J3/00; chemical aspects of, or use of materials for deodorisation of air, for disinfection or sterilisation, or for bandages, dressings, absorbent pads or surgical articles A61L; soap compositions C11D)

C07K: PEPTIDES (peptides containing β-lactam rings C07D; cyclic dipeptides not having in their molecule any other peptide link than those which form their ring, e.g. piperazine-2,5-diones, C07D; ergot alkaloids of the cyclic peptide type C07D519/02; single cell proteins, enzymes C12N; genetic engineering processes for obtaining peptides C12N15/00)

3.2.2. Technology application trend. The chart shows the distribution and development trend for patents applications of various technology directions, through which you can have an idea of technical distribution of all stages or key technologies layout of specific period, mine the development direction of hot technologies and get a general picture of the industry. It will also help the R&D team to optimize its R&D focus and roadmap. By comparing the development trend of various technologies directions will help users to learn what kinds of technical developed earlier, quicker and stronger. It can be seen from Figure 4 that the C12N layout is the earliest and the largest, and the number of patents in other fields is far less than that of C12N. Before 2013, except for C12N, the number of patent applications in other sub-categories was less than 100. From 2014 to 2019, the number of patent applications in seven sub-categories remained above 100.

Figure 3. Technology construction.
3.3. Applicant analysis

3.3.1. Applicant ranking. The chart shows the applicants (patentee) ranking according to patents quantities, through which you could find out the patent applicants with more innovation achievements so as to do further study on their patents competitiveness. It can be seen from Figure 5 that the top three institutions are massachusetts institute of technology, the broad institute inc, and Harvard college, and the number of patents is far more than twice the number of ranked fifth and sixth, and these three institutions are in the United States.

3.3.2. Applicant application trend. The chart shows applications developing trend of various applicants through which you could identify the starting, growing, leading and descending time for applicants' innovation and grasp the application strategy and innovation strength of all applicants. This
will also help you learn the innovation research input and technology activity levels of specific period and predict the future top players. It can be seen from Figure 6 that the top three institutions (Massachusetts Institute of Technology, the Broad Institute Inc., Harvard College) patent applications are concentrated after 2012, and the numbers are basically close. This has a certain relationship with the three organizations often cooperating to apply for patents. Before 2012, the patent applications of Sangamo, Dow AgroSciences LLC, and Cellectis were more active than other organizations.

**Figure 6.** Applicant application trend.

### 3.4. Region analysis

**3.4.1. Global region ranking (TOP 8).** The chart shows the analysis object's patents layout in various countries and areas, through which you can learn the technology innovation condition in different countries so as to find out main innovation sources and important target markets. It can be seen from Figure 7 that from a geographical point of view, the number of applications from China and the United States is far ahead, and the number is more than five times that of Canada, Japan, and Australia.

**Figure 7.** Global region ranking (TOP 8).
3.4.2. Global application trend. The chart shows the analysis object's patent applications trend in different countries and areas, through which you will learn the patent technology origins and developing condition in different countries. By comparing the technology activity of various areas, you could analyse the patents global layout and predict future trend and provide intelligence for global risks prevention strategy and competition strategy. It can be seen from Figure 8 that the main application areas have applied for more patents after 2013, especially China, the United States, the World Intellectual Property Organization and the European Patent Office. In addition, the World Intellectual Property Organization (WO) and the European Patent Office (EP) also received more patents. Since the World Intellectual Property Organization has nearly 150 member states, a technology can enter all member states in the national phase after an application is submitted to it in the international phase, while the European Patent Office has nearly 40 European member states, and patents that pass its examination can be protected in all its member states. Therefore, both the World Intellectual Property Organization and the European Patent Office are effective ways for patent applicants to distribute their patents globally. Through the acceptance volume of these two organizations, the trend of the global patent layout of gene editing technology can be inferred.

![Figure 8. Global application trend.](image)

3.5. Topic clustering
This article uses the 3D patent sand table in incoPat to cluster themes. The 3D patent sand table uses a three-dimensional topographic map to visually display the competitive situation of technology. The peaks represent the technology-intensive areas and the valleys represent the technical gaps. Each dot represents a patent, and the closer the dots are, the higher the technical relevance. The 3D patent sandbox also has a clustering function, which automatically groups patents according to their relevance through intelligent semantic analysis. Figure 9 shows the research themes from 2001 to 2011. Due to the small number of patents before 2011, fewer peaks were formed. The main topics are (1) zinc finger, nuclease, dna modification (2) phage resistance, cultures, streptococcus thermophilus (3) data analysis, endonucleases, homing (4) crispr, lactococcus, subtyping. Figure 10 shows the research themes from 2012 to 2020. After 2012, there are more patents filed and many peaks formed. The main themes are (1) knocking out (2) hybrid dna, cas, rnas (3) gene knockout, knocking out, nuclease (4) male sterile line, haploid, rice (5) edited (6) hemoglobinopathies, effector cells, endonuclease.
Figure 9. Topic clustering from 2001 to 2011.

Figure 10. Topic clustering from 2012 to 2020.
4. Suggestions

4.1. Emphasize original innovation
Using gene editing technology, precise modification of genomic DNA can be achieved, thereby strengthening the analysis of gene function, understanding the growth and development mechanism, and revealing the essence of life. In addition to the relatively mature CRISPR system, there are still new gene editing systems to be developed; the gene editing technology itself still has much room for improvement and perfection in terms of the precision and efficiency of targeted modification and the reduction of off-target effects. Therefore, strengthen the exploration of source technology and create original gene editing technology with independent intellectual property rights.

4.2. Focus on international cooperation
The partnership demonstrates the importance and effectiveness of high-level scientific research cooperation. Massachusetts Institute of Technology, Bode Research Institute, and Harvard University are the core institutions of gene editing research, and have related research cooperation with many universities, research institutes or companies around the world. It is recommended that special attention should be paid to cooperation in research and development; in addition to cooperation with leading international laboratories, we should also focus on the strong cooperation of domestic superior units. At the same time, it is necessary to strengthen the integration of production, education and research, and accelerate the transfer and transformation of research results through the cooperation of research institutions, universities and enterprises.

4.3. Strengthen intellectual property protection
To a large extent, the realization of technological competition has evolved into a dispute over intellectual property rights. The intellectual property system is the key system guarantee that plays the decisive role of the market mechanism, encourages innovation, stimulates creation, and promotes application. Focus on the patent layout of major foreign applicants, and combine its own technical planning to build a containment-type patent layout for our competitors to effectively avoid future infringement risks.

4.4. Introduce relevant regulatory and ethical norms
The biggest impact of new gene editing technology is its immeasurable application value. It has been in a leading position in many fields of gene editing applications, such as the construction of humanized animal models, the breeding and improvement of important crops and livestock and poultry, and the development of clinical trials [7]. Ethical norms and related regulatory policies that are conducive to the transformation of gene editing research results, and accelerate the advancement of gene editing technology for crop improvement and research and transformation of major diseases.

Authors’ contributions
Junfeng Yu: Conceptualization, Methodology, Formal analysis; Yanqing Zhao: Project administration; Haizhou Feng: Visualization, Investigation; Jiazeng Yu: Supervision; Tie Li: Data Curation.

Acknowledgments
This work was supported by Youth Fund Project of Shandong Academy of Sciences under Grant No. 2020QN0011.

References
[1] Xie Fangyan. Advances in the Application of Gene Editing Technology in Gene Therapy [J]. Biological Chemical Engineering, 2019(2):162-165.
[2] Wang Cuiping. Clinical Ethics of Human Embryo Gene Editing Mediated by CRISPR/Cas9 Technology [J]. Journal of Dialectics of Nature, 2018,40(11):105-112. DOI:10.15994/j.1000-0763.2018.11.014.
[3] Lu junnan, Chu Xin, pan Yanping, Chen Yingxi, Wen Luan, Dai Junbiao. Advances and Challenges in Gene Editing Technologies [J]. Bulletin of Chinese Academy of Sciences, 2018, 33 (11): 1184-1192. DOI:10.16418/j.issn.1000-3045.2018.11.006.

[4] Chandrasegaran S, Carroll D. Origins of programmable nucleases for genome engineering[J]. Journal of Molecular Biology, 2016, 428(5): 963-989. DOI: 10.1016/j.jmb.2015.10.014.

[5] Song Xiufang, Wei Xuemei, Zheng Lili, Zhao Yajuan, Zhang Kexin, Liu Chunguang, Xu Weihua. Analysis and Review on Gene Editing [J]. Bulletin of Chinese Academy of Sciences, 2020,35 (12): 1510-1524. DOI:10.16418/j.issn.1000-3045.20200601001.

[6] Capecchi M R. Culturing cells having a modification of a target DNA sequence in its genome, useful for producing transgenic animals or for manipulating plant cells, comprises transforming cells with a positive-negative selector vector: US,781559. 1997-01-09.

[7] Xu Li, Wang Yue, Yao Chiyuan, Xu Ping. Trends and Development Bottleneck Analysis of Gene Editing Technology [J]. China Biotechnology, 2018,38 (12): 113-122. DOI:10.13523/j.cb.20181214.