



**HAL**  
open science

## Innovative therapy in European Parliament's positions: a numerical science-based vocabulary analysis

Thomas Allouche, Frédéric Bechet, Valentin Brunel, Marie Glinel, Aurélie  
Mahalatchimy

### ► To cite this version:

Thomas Allouche, Frédéric Bechet, Valentin Brunel, Marie Glinel, Aurélie Mahalatchimy. Innovative therapy in European Parliament's positions: a numerical science-based vocabulary analysis. European Society for Gene & Cell Therapy Congress 2023, Oct 2023, Bruxelles (BEL), Belgium. halshs-04290599

**HAL Id: halshs-04290599**

**<https://shs.hal.science/halshs-04290599v1>**

Submitted on 16 Nov 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Innovative therapy in European Parliament's positions: a numerical science-based vocabulary analysis

T. Allouche (1), V. Brunel (2), F. Béchet (3), M. Glinel (4), A. Mahalatchimy (5)

1 Masters student in Computer Science, Aix Marseille University

2 Research Engineer, CNRS, UMR 7318 International, Comparative and European laws (DICE) CERIC, CNRS, Aix Marseille Univ., Toulon Univ., Pau & Pays de l'Adour Univ., Aix-en-Provence, France

3 Computer Science Professor, Director of LIS, UMR 7020 - CNRS / AMU / UTLN

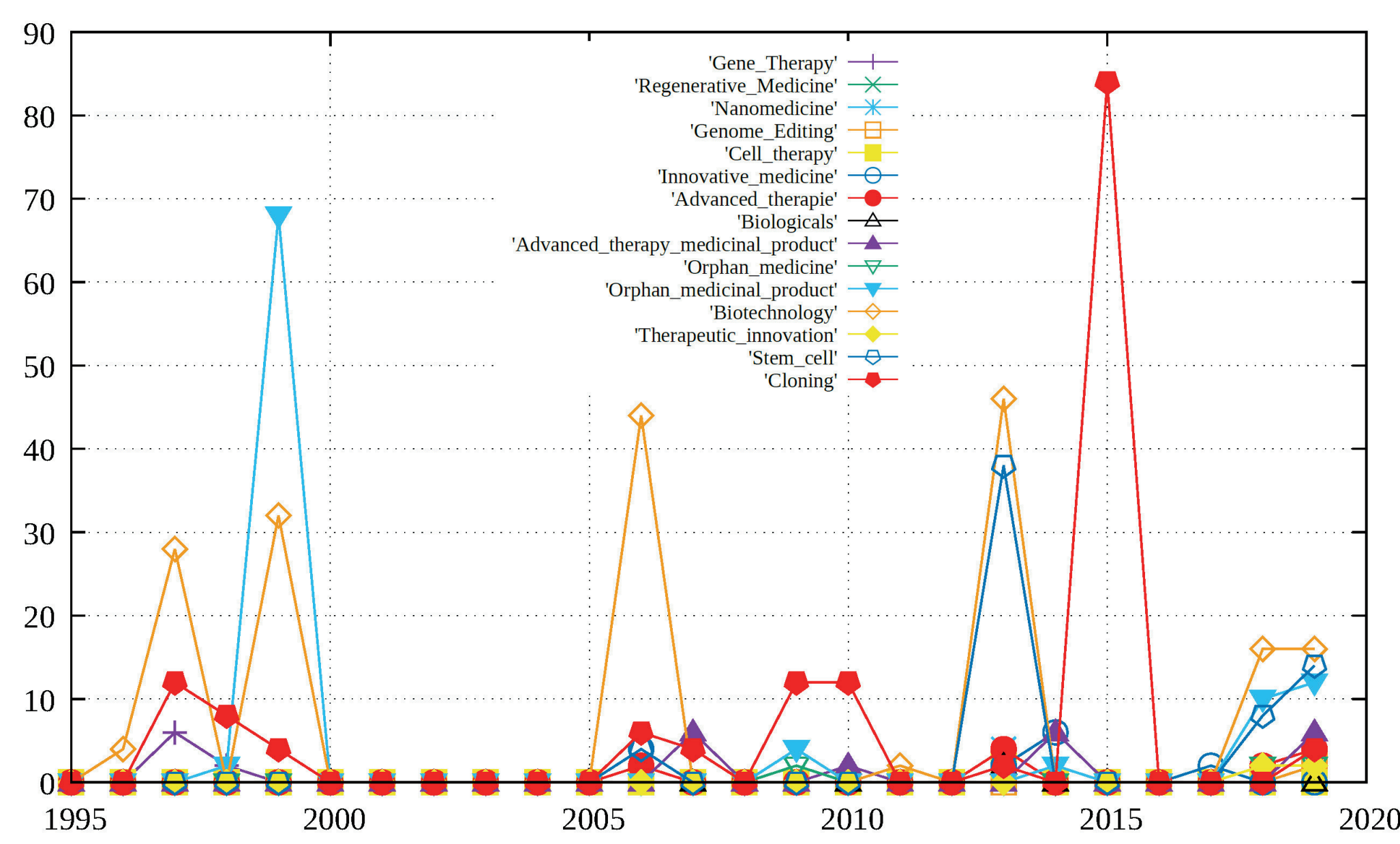
4 Post-doctorate fellow in the I-Biolex project, UMR 7318 DICE CERIC, CNRS, Aix Marseille Univ., Toulon Univ., Pau & Pays de l'Adour Univ., Aix-en-Provence, France.

5 EuroGCT Deputy coordinator and WP4 leader Permanent, Researcher in law at the French National Centre for Scientific Research (CNRS, chargée de recherche), Law Faculty, UMR 7318 DICE CERIC, CNRS, Aix Marseille Univ., Toulon Univ., Pau & Pays de l'Adour Univ., Aix-en-Provence, France. Contact: valentin.BRUNEL@univ-amu.fr / This work has been supported by ANR-funded I-BioLex project (ANR-20-CE26-0007-01), by the European Union's Horizon 2020 EuroGCT project (GA: No 965241) and the Institut Laënnec (Aix Marseille Université)

**INTRODUCTION.** Through the evolution of European Union (EU) law, the European Parliament has been more and more involved in the adoption of EU binding law. Indeed, the co-decision procedure (i.e. the joint adoption of legislative acts both by the European Parliament and the Council of the European Union) has become the ordinary legislative procedure from the adoption of the Lisbon Treaty in 2007. But the Parliament has also been involved in other types of procedures such as the consultation or cooperation procedures. Such participation takes the form of documents that are called European Parliament's positions. The latter constitute an interesting corpus to know more about the fields of activities of the European Parliament, as the European institution representing EU citizens. Focusing on innovative therapy in a wide meaning, this poster aims at conducting an exhaustive study of the mention of terms related to this topic in European Parliament's positions.

**METHODOLOGY AND CORPUS.** From the official website of EU law (EUR-Lex), we constructed a database comprising all positions of the European Parliament, with available metadata. This metadata mainly includes the date and type of documents, as well as the related legal basis and all instruments cited. The database was then searched in order to find every occurrence of a selected vocabulary on innovative therapy, including general terms such as "biomedical innovation" or more specific terms such as "gene therapy", "cell therapy" or "advanced therapy medicinal products". We thus were able to see what terms were used and when they were first mentioned, and in what kind of texts they were most present. In our corpus of texts including our selected terms, another point of interest was the way this vocabulary was interconnected, allowing us to group words according to their co-occurrence in documents. From a legislative mapping perspective, we were also able to see which legal basis were more often used, as well as which documents were citing each other. This work is presented under the form of a network graph comprising all documents and their interrelations.

## Findings 1.1: Frequency over time and by category of keywords



Using the database from Eur-Lex, our first task was to look for occurrences of our selected vocabulary. The first graph shows through time the frequency of each occurrence.

Graph 1: Frequency of keyword occurrences in our corpus by date

- All our keywords relating to organoids have never appeared among the EP resolutions: Organoid, Organ on a chip, Body on a chip, Cerebroid, Cerebral organoid, Brain organoid, neural organoid, Gastruloid, Blastoid, Iblastoid, Tumoroid.
- 4 keywords have never appeared: Biomedical innovation, Tissue engineering, Advanced medicine, Biological medicine
- 4 keywords have appeared the most: Biotechnology, Cloning, Orphan medicinal product, Stem cell
- Biotechnology has appeared the most and in 20 different documents, by descending order, especially:
  - Around 1998: Research funding (4th framework programme), directive on legal protection of biotechnological inventions, directive on genetically modified organisms (GMOs)
  - Around 2006: Research funding (7th framework programme)
  - Around 2014: Research funding (Horizon 2020) and Regulation related to the Nagoya Protocol on Access to Genetic Resources
  - Around 2020: research funding (InvestEU & Horizon Europe)
  - In 2011: directive on GMOs
- Cloning has also appeared a lot and in 15 different documents, by descending order, especially:
  - In 2015 on a finally withdrawn proposal of a directive on the cloning of animals
  - Around 1998: research funding (fifth framework programme) and directive on the legal protection of biotechnological
  - Between 2007 & 2010: novel foods and on organic production
  - In 2006: research funding (7th framework programme)
  - In 2013: research funding (Horizon 2020)
- Stem cell has also appeared a lot and in 11 different documents, by descending order, especially:
  - In 2013: research funding (Horizon 2020)
  - Around 2020: research funding (Horizon Europe, InvestEU)
  - In 2019: research funding (Horizon Europe, InvestEU)
  - In 2006: research funding (7th framework programme)
- Orphan medicinal product has also appeared a lot and in 7 different documents, by descending order of occurrence, especially:
  - In 1999: Regulation on Orphan medicinal products (most occurrence)
  - In 2019: Regulation on Health Technology Assessment (HTA)
  - In 1998, 2009 on actions on rare diseases; and in 2014 on a proposal related to the use of the regulatory procedure with scrutiny
- Advanced therapy medicinal product (ATMP) has not appeared many times, but in 9 different documents, especially:
  - In 2007: Regulation on ATMP
  - In 2010: Regulation on pharmacovigilance
  - In 2014: Regulation on medical devices and proposal related to the use of the regulatory procedure with scrutiny
  - In 2019 & 2020: research funding (Horizon Europe, InvestEU, EU4Health) and a regulation related to the use of the regulatory procedure with scrutiny
- Several keywords that do not appear a lot for a long time have clearly appeared around 2020: therapeutic innovation, genome editing, nanomedicine:
  - Genome editing and nanomedicine in same 2019 text > research funding (7th framework programme)
  - Therapeutic innovation in Regulation on HTA

## Findings 2: Networks between documents



In accordance with our objectives, we looked at all instruments cited in every document of our corpus containing at least one of our keywords.

The graph on the left is the result of a network analysis conducted on these instruments. Its principle is straightforward: when a document comprised another as an instrument cited in the Eur-lex metadata, a link was declared between both. The documents having more relations are thus those either that cited the most documents, or were cited by most, or both.

Graph 2: Network of instruments cited

### Main results

- Among the 7 resolutions the most cited, 3 lead to a directive and 4 to a regulation.
- Only one of the most cited texts relates to a final text that has not been adopted yet (52022AP0317 on promotion of energy from renewable sources).
- The topics covered by descending order are the following: Animal Health (transmissible animal diseases) > European Maritime, Fisheries and Aquaculture Fund > medical devices > promotion of energy from renewable sources > quality of water intended for human consumption > Internal Security Fund > industrial emissions (integrated pollution prevention and control)
- Regarding the dates of the most cited texts: 3 are from 2019, 2 from 2014, 1 from 2022, and 1 from 2009

### Main conclusions

- The legal nature of the act to which the resolutions relate does not seem to have an impact on its importance regarding its citation quantity.
- The citation quantity is not linked to the outcome of the legal procedure.
- The topics appear various.
- Nevertheless, they are mainly crossing health and environmental fields.
- Among the 7 different topics, 2 relate to a fund (Fisheries and Internal security)
- Only one text directly relates to pharmaceutical and it covers medical devices and not medicinal products
- The dates of texts range from 2009 to 2022, and mainly from 2014 to 2019.

The analysis of the most cited legal basis has no interest for our studies as it highlights the legal rules for the involvement of the European Parliament in the adoption of legal acts. Indeed, the most cited legal basis are the ones that empowered and framed the involvement of the European Parliament in the legal procedures: mainly Articles of the Treaty relating to the ordinary legislative procedure (current and previous versions), European Parliament's rules of procedures, and adaptation of legal acts for the use of the regulatory procedure with scrutiny. Thus, it relates to the legal technique only and not to the topic of our study.

## Findings 3: Exploratory analysis

Once all documents containing our selected vocabulary were identified, we wished to conduct an automated research of associated documents. The method consists in producing a vector representation of each document in the collection in order to be able to compute a similarity metric. To produce these vector representations, we can define a common vocabulary for all documents and compute a weight for each term in a given document to obtain its vector. Regardless of the method used to produce these vectors, finding new relevant documents consists in looking for the nearest neighbours of target documents in the document vector space. We performed a preliminary experiment following this methodology using the TF/IDF weighting method. Unfortunately, this analysis did not yield interesting results so far, as documents with similar vectors did not seem to match our theme of innovation in biotechnologies. Future research may comprise the use of more advanced weighting methods based on embedding as well as considering document graphs. However, the very strong homogeneity of Parliament resolutions' texts makes this sort of tool difficult to operate.

## Findings 1.2: Clustering of keywords

We then set out to create a series of groups or clusters, based on terms which were more often present in the same documents. Below is the list of every cluster:

### Cluster A: Biotechnology/ Stem cell/Cloning

Common texts for these 3 keywords: 7th framework programme, Horizon 2020, InvestEU & Horizon Europe, and directive on legal protection of biotechnological inventions (For Biotechnology and cloning only)

=> The most common texts relate to research funding.

### Cluster B: Advanced therapy medicinal product/ Advanced therapy/ Gene therapy/ Innovative medicine

There is not 1 common text for these 4 keywords. But : 1 for Advanced therapy, Innovative medicine: Horizon 2020 (2013) and 1 for Advanced therapy, Gene therapy, Advanced therapy medicinal product: Horizon Europe (2019)

=> They all relate to research funding.

### Cluster C: Nanomedicine/ Genome editing/ Regenerative Medicine/ Cell therapy

Not 1 common text for these 4 keywords. But : 1 for Nanomedicine, RM, Cell Therapy: Horizon 2020 (2013) and 1 for Nanomedicine, Genome editing: Horizon Europe (2019)

=> They all relate to research funding.

### Cluster D: Orphan medicinal product/ Orphan medicine/ Therapeutic innovation

2 texts related to Regulation on Health Technology Assessment (2019 & 2018) are common to these 3 keywords, and 1 text related to Action on rare diseases (2009) is common to Orphan medicinal product, Orphan medicine

=> The most common texts relate to Regulation on Health Technology Assessment.

### On the occurrences of keywords

- As Biomedical innovation, Tissue engineering, Advanced Medicine and Biological Medicine have not been found and are not legal terms, it may be thought that the EP is using preferentially the legal term. Nevertheless, although Orphan medicinal product has been more used than orphan medicine, the latter has been found, even if it is not the legal term.
- Biotechnology has appeared the most both in general texts on research funding and in more specific texts (legal protection of biotechnological inventions, Nagoya Protocol on Access to Genetic Resources, GMOs).
- Cloning has appeared a lot both in general texts on research funding and in more specific texts (proposal of a directive on the cloning of animals, legal protection of biotechnological inventions, novel foods and on organic production).
- Stem cell has also appeared a lot but only in general texts on research funding.
- Orphan medicinal product has also appeared a lot mainly in texts related directly to these products or to rare diseases, but also to Health Technology Assessment.
- ATMP has not appeared many times but in more different documents than Orphan medicinal product (9 vs.7). They relate to ATMP specifically but also to a specific stage of their development phase (pharmacovigilance) and to medical device as a potentially closed legal category and relevant regulation (e.g. Combined ATMP). They also relate to more general topics such as research funding and the use of the regulatory procedure with scrutiny.

### On the networks between documents

- The apparition of our keywords in the most cited texts is not linked to the importance of the final legal act (legal nature) or to the outcome of the legal procedure (adoption or not of the final legal act). Thus, the quantity of citation of a text containing our keywords is more linked to its topic than to its technico-legal context.
- The topics of the most cited texts are transversal to health and environmental issues.
- Contrary to our expectations, our keywords had very little to do with pharmaceuticals, as only one text (on medical devices and not on medicines) in this field stood out.
- There may be an unforeseen parameter on the Eur-Lex website that promotes the most recent texts. This may have biased our results.

While the European Parliament has been known to tackle ethical questions, it has not yet identified organoids as a field to be considered. In contrast, Biotechnology (e.g., directives on legal protection of biotechnological inventions and on GMOs), Orphan medicinal product (e.g., Regulation on HTA) and Cloning (e.g., withdrawn proposal of a directive on the cloning of animals) have been highly found, and the most found among European Parliament's opinions proposing amendments due to ethical, legal, and societal issues at stake.

Both the analysis by occurrence of keywords and the analysis by clusters of keywords found together show that Biomedical Innovation related keywords appear more in texts related to research funding. Nevertheless, cluster 4 is distinct and specific as none of its keywords (Orphan medicinal product, Orphan medicine, Therapeutic innovation) appear in texts related to research funding.

Both Orphan medicinal product and Advanced Therapy Medicinal Product appear in more various texts (beyond general research funding for the latter) but also texts that have more specific topics as linked to these particularly regulated types of medicinal products.

## Conclusion