Hybrid Summarization of Chemistry Patents

Darby Brown

W266: Natural Language Processing

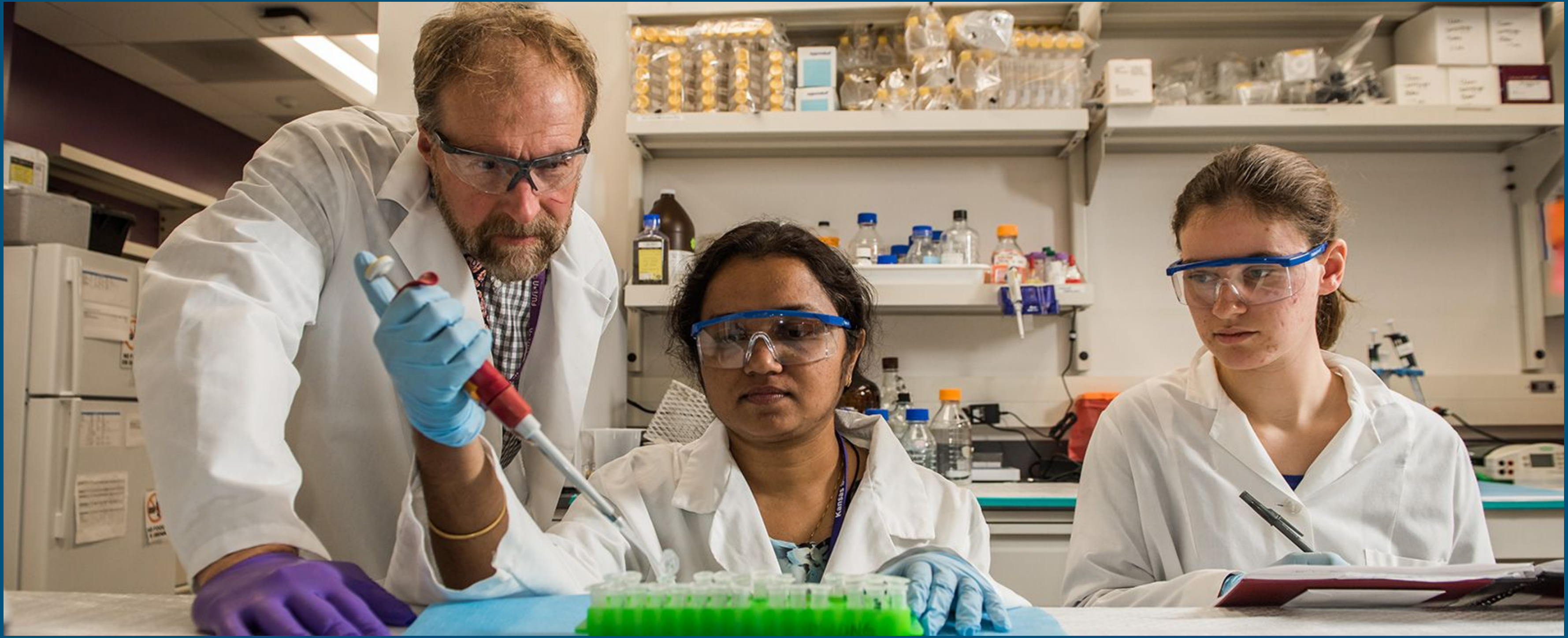
Dec 2023

Why patent summarization?

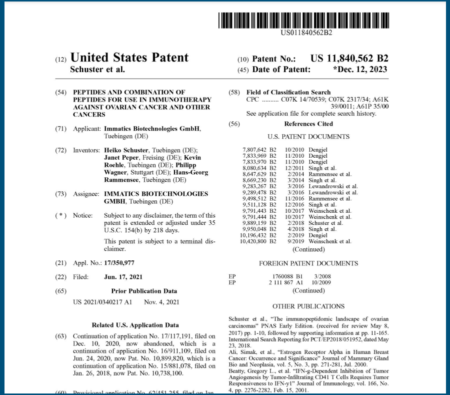
71%

of new technology published in patent literature
is not published elsewhere.

Reduce technology landscaping time to get scientists back into the lab, faster.



Working with Patents: Key Challenges



US011840562B2

(12) **United States Patent**
Schuster et al.

(10) **Patent No.: US 11,840,562 B2**
(45) **Date of Patent: *Dec. 12, 2023**

(54) **PEPTIDES AND COMBINATION OF PEPTIDES FOR USE IN IMMUNOTHERAPY AGAINST OVARIAN CANCER AND OTHER CANCERS**

(71) Applicant: **Immatics Biotechnologies GmbH**, Tuebingen (DE)

(72) Inventors: **Heiko Schuster**, Tuebingen (DE); **Janet Peper**, Freising (DE); **Kevin Roehle**, Tuebingen (DE); **Philipp Wagner**, Stuttgart (DE); **Hans-Georg Rammensee**, Tuebingen (DE)

(73) Assignee: **IMMATICS BIOTECHNOLOGIES GMBH**, Tuebingen (DE)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 218 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **17/350,977**

(22) Filed: **Jun. 17, 2021**

(65) **Prior Publication Data**

US 2021/0340217 A1 Nov. 4, 2021

Related U.S. Application Data

(63) Continuation of application No. 17/117,191, filed on Dec. 10, 2020, now abandoned, which is a continuation of application No. 16/911,109, filed on Jun. 24, 2020, now Pat. No. 10,899,820, which is a continuation of application No. 15/881,078, filed on Jan. 26, 2018, now Pat. No. 10,738,100.

(60) Provisional application No. 62/451,255, filed on Jan.

(58) **Field of Classification Search**
CPC C07K 14/70539; C07K 2317/34; A61K 39/0011; A61P 35/00
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

7,807,642	B2	10/2010	Dengjel
7,833,969	B2	11/2010	Dengjel
7,833,970	B2	11/2010	Dengjel
8,080,634	B2	12/2011	Singh et al.
8,647,629	B2	2/2014	Rammensee et al.
8,669,230	B2	3/2014	Singh et al.
9,283,267	B2	3/2016	Lewandrowski et al.
9,289,478	B2	3/2016	Lewandrowski et al.
9,498,512	B2	11/2016	Rammensee et al.
9,511,128	B2	12/2016	Singh et al.
9,791,443	B2	10/2017	Weinschenk et al.
9,791,444	B2	10/2017	Weinschenk et al.
9,889,159	B2	2/2018	Schuster et al.
9,950,048	B2	4/2018	Singh et al.
10,196,432	B2	2/2019	Dengjel
10,420,800	B2	9/2019	Weinschenk et al.

(Continued)

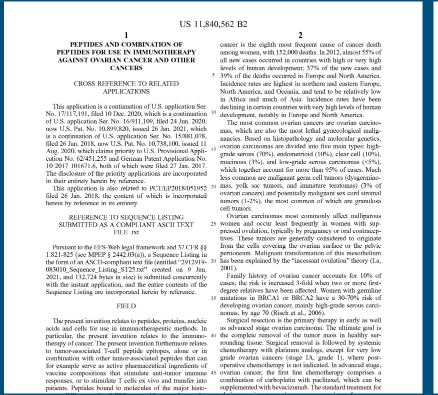
FOREIGN PATENT DOCUMENTS

EP	1760088	B1	3/2008
EP	2 111 867	A1	10/2009

(Continued)

OTHER PUBLICATIONS

Schuster et al., "The immunopeptidomic landscape of ovarian carcinomas" PNAS Early Edition. (received for review May 8, 2017) pp. 1-10, followed by supporting information at pp. 11-165.
International Search Reporting for PCT/EP2018/051952, dated May 23, 2018.
Ali, Simak, et al., "Estrogen Receptor Alpha in Human Breast Cancer: Occurrence and Significance" Journal of Mammary Gland Bio and Neoplasia, vol. 5, No. 3, pp. 271-281, Jul. 2000.
Beatty, Gregory L., et al. "IFN-g-Dependent Inhibition of Tumor Angiogenesis by Tumor-Infiltrating CD41 T Cells Requires Tumor Responsiveness to IFN-y1" Journal of Immunology, vol. 166, No. 4, pp. 2276-2282, Feb. 15, 2001.



US 11,840,562 B2

1

PEPTIDES AND COMBINATION OF PEPTIDES FOR USE IN IMMUNOTHERAPY AGAINST OVARIAN CANCER AND OTHER CANCERS

CROSS REFERENCE TO RELATED APPLICATIONS

This application is a continuation of U.S. application Ser. No. 17/117,191, filed 10 Dec. 2020, which is a continuation of U.S. application Ser. No. 16/911,109, filed 24 Jun. 2020, now U.S. Pat. No. 10,899,820, issued 26 Jan. 2021, which is a continuation of U.S. application Ser. No. 15/881,078, filed 26 Jan. 2018, now U.S. Pat. No. 10,738,100, issued 11 Aug. 2020, which claims priority to U.S. Provisional Application No. 62/451,255 and German Patent Application No. 10 2017 101671.6, both of which were filed 27 Jan. 2017. The disclosure of the priority applications are incorporated in their entirety herein by reference.

This application is also related to PCT/EP2018/051952 filed 26 Jan. 2018, the content of which is incorporated herein by reference in its entirety.

REFERENCE TO SEQUENCE LISTING SUBMITTED AS A COMPLIANT ASCII TEXT FILE .txt

Pursuant to the EFS-Web legal framework and 37 CFR §§ 1.821-825 (see MPEP § 2442.03(a)), a Sequence Listing in the form of an ASCII-compliant text file (entitled "2912919-083010_Sequence_Listing_ST25.txt" created on 9 Jun. 2021, and 132,724 bytes in size) is submitted concurrently with the instant application, and the entire contents of the Sequence Listing are incorporated herein by reference.

FIELD

The present invention relates to peptides, proteins, nucleic acids and cells for use in immunotherapeutic methods. In particular, the present invention relates to the immunotherapy of cancer. The present invention furthermore relates to tumor-associated T-cell peptide epitopes, alone or in combination with other tumor-associated peptides that can for example serve as active pharmaceutical ingredients of vaccine compositions that stimulate anti-tumor immune responses, or to stimulate T cells ex vivo and transfer into patients. Peptides bound to molecules of the major histo-

2

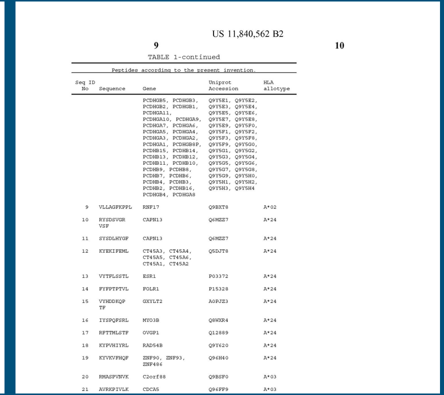
cancer is the eighth most frequent cause of cancer death among women, with 152,000 deaths. In 2012, almost 55% of all new cases occurred in countries with high or very high levels of human development; 37% of the new cases and 39% of the deaths occurred in Europe and North America. Incidence rates are highest in northern and eastern Europe, North America, and Oceania, and tend to be relatively low in Africa and much of Asia. Incidence rates have been declining in certain countries with very high levels of human development, notably in Europe and North America.

The most common ovarian cancers are ovarian carcinomas, which are also the most lethal gynecological malignancies. Based on histopathology and molecular genetics, ovarian carcinomas are divided into five main types: high-grade serous (70%), endometrioid (10%), clear cell (10%), mucinous (3%), and low-grade serous carcinomas (<5%), which together account for more than 95% of cases. Much less common are malignant germ cell tumors (dysgerminomas, yolk sac tumors, and immature teratomas) (3% of ovarian cancers) and potentially malignant sex cord stromal tumors (1-2%), the most common of which are granulosa cell tumors.

Ovarian carcinomas most commonly affect nulliparous women and occur least frequently in women with suppressed ovulation, typically by pregnancy or oral contraceptives. These tumors are generally considered to originate from the cells covering the ovarian surface or the pelvic peritoneum. Malignant transformation of this mesothelium has been explained by the "incessant ovulation" theory (La, 2001).

Family history of ovarian cancer accounts for 10% of cases; the risk is increased 3-fold when two or more first-degree relatives have been affected. Women with germline mutations in BRCA1 or BRCA2 have a 30-70% risk of developing ovarian cancer, mainly high-grade serous carcinomas, by age 70 (Risch et al., 2006).

Surgical resection is the primary therapy in early as well as advanced stage ovarian carcinoma. The ultimate goal is the complete removal of the tumor mass in healthy surrounding tissue. Surgical removal is followed by systemic chemotherapy with platinum analogs, except for very low grade ovarian cancers (stage IA, grade 1), where post-operative chemotherapy is not indicated. In advanced stage, ovarian cancer, the first line chemotherapy comprises a combination of carboplatin with paclitaxel, which can be supplemented with bevacizumab. The standard treatment for



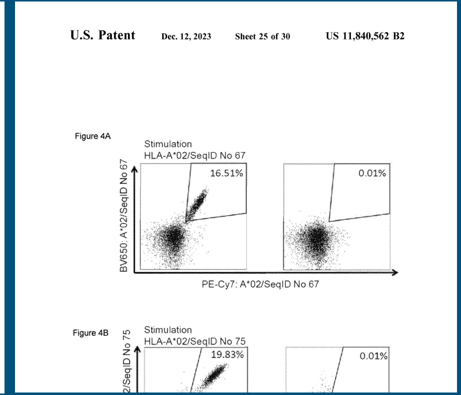
US 11,840,562 B2

9 10

TABLE 1-continued

Peptides according to the present invention.

Seq ID No	Sequence	Gene	Uniprot Accession	HLA allotype
		PCDHGB5, PCDHGB3, PCDHGB2, PCDHGB1, PCDHGA11, PCDHGA10, PCDHGA9, PCDHGA7, PCDHGA6, PCDHGA5, PCDHGA4, PCDHGA3, PCDHGA2, PCDHGA1, PCDHGB8P, PCDHB15, PCDHB14, PCDHB13, PCDHB12, PCDHB11, PCDHB10, PCDHB9, PCDHB8, PCDHB7, PCDHB6, PCDHB4, PCDHB3, PCDHB2, PCDHB16, PCDHGB4, PCDHGA8	Q9Y5E1, Q9Y5E2, Q9Y5E3, Q9Y5E4, Q9Y5E5, Q9Y5E6, Q9Y5E7, Q9Y5E8, Q9Y5E9, Q9Y5F0, Q9Y5F1, Q9Y5F2, Q9Y5F3, Q9Y5F8, Q9Y5F9, Q9Y5G0, Q9Y5G1, Q9Y5G2, Q9Y5G3, Q9Y5G4, Q9Y5G5, Q9Y5G6, Q9Y5G7, Q9Y5G8, Q9Y5G9, Q9Y5H0, Q9Y5H1, Q9Y5H2, Q9Y5H3, Q9Y5H4	
9	VLLAGFKPPL	RNF17	Q9BXT8	A*02
10	RYSDSVGRVSF	CAPN13	Q6MZZ7	A*24
11	SYSDLHYGF	CAPN13	Q6MZZ7	A*24
12	KYEKIFEML	CT45A3, CT45A4, CT45A5, CT45A6, CT45A1, CT45A2	Q5DJT8	A*24
13	VYTFLSSTL	ESR1	P03372	A*24
14	FYFPTPTVL	FOLR1	P15328	A*24
15	VYHDDKQPTF	GXYLT2	A0PJZ3	A*24
16	IYSPQFSRL	MYO3B	Q8WXR4	A*24
17	RFTTMLSTF	OVGP1	Q12889	A*24
18	KYPVHIYRL	RAD54B	Q9Y620	A*24
19	KYVKVFHQF	ZNF90, ZNF93, ZNF486	Q96H40	A*24
20	RMASPVNVK	C2orf88	Q9BSF0	A*03
21	AVRKPIVLK	CDCA5	Q96FF9	A*03



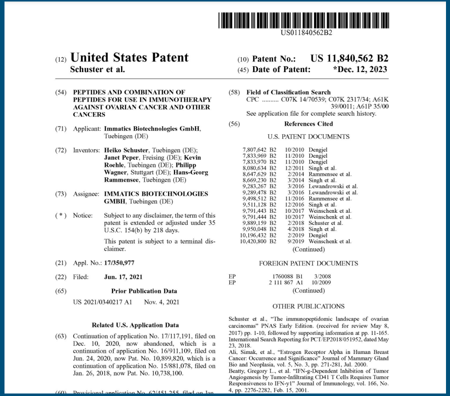
U.S. Patent **Dec. 12, 2023** **Sheet 25 of 30** **US 11,840,562 B2**

Figure 4A

Figure 4B

- Average document length: 24000 words
- Technical and legal jargon, often only interpretable by SMEs and attorneys
- Images, tables, and charts are frequently referred to in the text

Working with Patents: Key Challenges



US011840562B2

(12) **United States Patent**
Schuster et al.

(10) **Patent No.: US 11,840,562 B2**
(45) **Date of Patent: *Dec. 12, 2023**

(54) **PEPTIDES AND COMBINATION OF PEPTIDES FOR USE IN IMMUNOTHERAPY AGAINST OVARIAN CANCER AND OTHER CANCERS**

(71) Applicant: **Immatics Biotechnologies GmbH**, Tuebingen (DE)

(72) Inventors: **Heiko Schuster**, Tuebingen (DE); **Janet Peper**, Freising (DE); **Kevin Roehle**, Tuebingen (DE); **Philipp Wagner**, Stuttgart (DE); **Hans-Georg Rammensee**, Tuebingen (DE)

(73) Assignee: **IMMATICS BIOTECHNOLOGIES GMBH**, Tuebingen (DE)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 218 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **17/350,977**

(22) Filed: **Jun. 17, 2021**

(65) **Prior Publication Data**

US 2021/0340217 A1 Nov. 4, 2021

Related U.S. Application Data

(63) Continuation of application No. 17/117,191, filed on Dec. 10, 2020, now abandoned, which is a continuation of application No. 16/911,109, filed on Jun. 24, 2020, now Pat. No. 10,899,820, which is a continuation of application No. 15/881,078, filed on Jan. 26, 2018, now Pat. No. 10,738,100.

(60) Provisional application No. 62/451,255, filed on Jan.

(58) **Field of Classification Search**
CPC C07K 14/70539; C07K 2317/34; A61K 39/0011; A61P 35/00
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

7,807,642	B2	10/2010	Dengjel
7,833,969	B2	11/2010	Dengjel
7,833,970	B2	11/2010	Dengjel
8,080,634	B2	12/2011	Singh et al.
8,647,629	B2	2/2014	Rammensee et al.
8,669,230	B2	3/2014	Singh et al.
9,283,267	B2	3/2016	Lewandrowski et al.
9,289,478	B2	3/2016	Lewandrowski et al.
9,498,512	B2	11/2016	Rammensee et al.
9,511,128	B2	12/2016	Singh et al.
9,791,443	B2	10/2017	Weinschenk et al.
9,791,444	B2	10/2017	Weinschenk et al.
9,889,159	B2	2/2018	Schuster et al.
9,950,048	B2	4/2018	Singh et al.
10,196,432	B2	2/2019	Dengjel
10,420,800	B2	9/2019	Weinschenk et al.

(Continued)

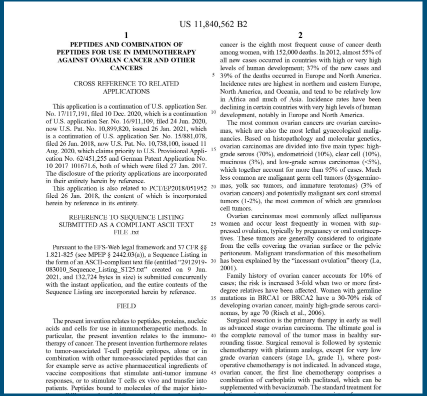
FOREIGN PATENT DOCUMENTS

EP	1760088 B1	3/2008
EP	2 111 867 A1	10/2009

(Continued)

OTHER PUBLICATIONS

Schuster et al., "The immunopeptidomic landscape of ovarian carcinomas" PNAS Early Edition. (received for review May 8, 2017) pp. 1-10, followed by supporting information at pp. 11-165.
International Search Reporting for PCT/EP2018/051952, dated May 23, 2018.
Ali, Simak, et al., "Estrogen Receptor Alpha in Human Breast Cancer: Occurrence and Significance" Journal of Mammary Gland Bio and Neoplasia, vol. 5, No. 3, pp. 271-281, Jul. 2000.
Beatty, Gregory L., et al. "IFN-g-Dependent Inhibition of Tumor Angiogenesis by Tumor-Infiltrating CD41 T Cells Requires Tumor Responsiveness to IFN-y1" Journal of Immunology, vol. 166, No. 4, pp. 2276-2282, Feb. 15, 2001.



US 11,840,562 B2

1

PEPTIDES AND COMBINATION OF PEPTIDES FOR USE IN IMMUNOTHERAPY AGAINST OVARIAN CANCER AND OTHER CANCERS

CROSS REFERENCE TO RELATED APPLICATIONS

This application is a continuation of U.S. application Ser. No. 17/117,191, filed 10 Dec. 2020, which is a continuation of U.S. application Ser. No. 16/911,109, filed 24 Jun. 2020, now U.S. Pat. No. 10,899,820, issued 26 Jan. 2021, which is a continuation of U.S. application Ser. No. 15/881,078, filed 26 Jan. 2018, now U.S. Pat. No. 10,738,100, issued 11 Aug. 2020, which claims priority to U.S. Provisional Application No. 62/451,255 and German Patent Application No. 10 2017 101671.6, both of which were filed 27 Jan. 2017. The disclosure of the priority applications are incorporated in their entirety herein by reference.

This application is also related to PCT/EP2018/051952 filed 26 Jan. 2018, the content of which is incorporated herein by reference in its entirety.

REFERENCE TO SEQUENCE LISTING SUBMITTED AS A COMPLIANT ASCII TEXT FILE .txt

Pursuant to the EFS-Web legal framework and 37 CFR §§ 1.821-825 (see MPEP § 2442.03(a)), a Sequence Listing in the form of an ASCII-compliant text file (entitled "2912919-083010_Sequence_Listing_ST25.txt" created on 9 Jun. 2021, and 132,724 bytes in size) is submitted concurrently with the instant application, and the entire contents of the Sequence Listing are incorporated herein by reference.

FIELD

The present invention relates to peptides, proteins, nucleic acids and cells for use in immunotherapeutic methods. In particular, the present invention relates to the immunotherapy of cancer. The present invention furthermore relates to tumor-associated T-cell peptide epitopes, alone or in combination with other tumor-associated peptides that can for example serve as active pharmaceutical ingredients of vaccine compositions that stimulate anti-tumor immune responses, or to stimulate T cells ex vivo and transfer into patients. Peptides bound to molecules of the major histo-

2

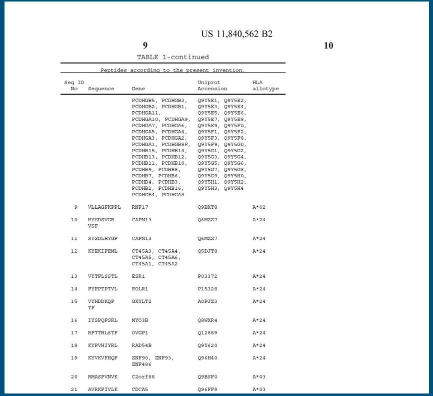
cancer is the eighth most frequent cause of cancer death among women, with 152,000 deaths. In 2012, almost 55% of all new cases occurred in countries with high or very high levels of human development; 37% of the new cases and 39% of the deaths occurred in Europe and North America. Incidence rates are highest in northern and eastern Europe, North America, and Oceania, and tend to be relatively low in Africa and much of Asia. Incidence rates have been declining in certain countries with very high levels of human development, notably in Europe and North America.

The most common ovarian cancers are ovarian carcinomas, which are also the most lethal gynecological malignancies. Based on histopathology and molecular genetics, ovarian carcinomas are divided into five main types: high-grade serous (70%), endometrioid (10%), clear cell (10%), mucinous (3%), and low-grade serous carcinomas (<5%), which together account for more than 95% of cases. Much less common are malignant germ cell tumors (dysgerminomas, yolk sac tumors, and immature teratomas) (3% of ovarian cancers) and potentially malignant sex cord stromal tumors (1-2%), the most common of which are granulosa cell tumors.

Ovarian carcinomas most commonly affect nulliparous women and occur least frequently in women with suppressed ovulation, typically by pregnancy or oral contraceptives. These tumors are generally considered to originate from the cells covering the ovarian surface or the pelvic peritoneum. Malignant transformation of this mesothelium has been explained by the "incessant ovulation" theory (La, 2001).

Family history of ovarian cancer accounts for 10% of cases; the risk is increased 3-fold when two or more first-degree relatives have been affected. Women with germline mutations in BRCA1 or BRCA2 have a 30-70% risk of developing ovarian cancer, mainly high-grade serous carcinomas, by age 70 (Risch et al., 2006).

Surgical resection is the primary therapy in early as well as advanced stage ovarian carcinoma. The ultimate goal is the complete removal of the tumor mass in healthy surrounding tissue. Surgical removal is followed by systemic chemotherapy with platinum analogs, except for very low grade ovarian cancers (stage IA, grade 1), where post-operative chemotherapy is not indicated. In advanced stage, ovarian cancer, the first line chemotherapy comprises a combination of carboplatin with paclitaxel, which can be supplemented with bevacizumab. The standard treatment for



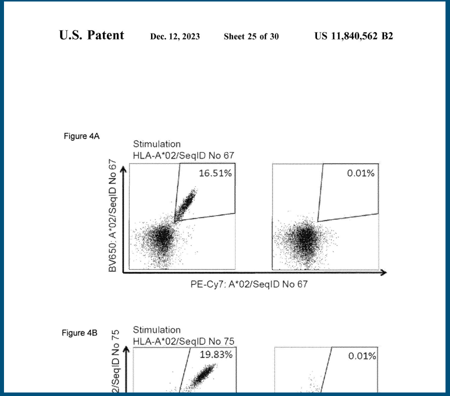
US 11,840,562 B2

9 10

TABLE 1-continued

Peptides according to the present invention.

Seq ID No	Sequence	Gene	Uniprot Accession	HLA allotype
		PCDHGB5, PCDHGB3, PCDHGB2, PCDHGB1, PCDHGA11, PCDHGA10, PCDHGA9, PCDHGA7, PCDHGA6, PCDHGA5, PCDHGA4, PCDHGA3, PCDHGA2, PCDHGA1, PCDHGB8P, PCDHB15, PCDHB14, PCDHB13, PCDHB12, PCDHB11, PCDHB10, PCDHB9, PCDHB8, PCDHB7, PCDHB6, PCDHB4, PCDHB3, PCDHB2, PCDHB16, PCDHGB4, PCDHGA8	Q9Y5E1, Q9Y5E2, Q9Y5E3, Q9Y5E4, Q9Y5E5, Q9Y5E6, Q9Y5E7, Q9Y5E8, Q9Y5E9, Q9Y5F0, Q9Y5F1, Q9Y5F2, Q9Y5F3, Q9Y5F8, Q9Y5F9, Q9Y5G0, Q9Y5G1, Q9Y5G2, Q9Y5G3, Q9Y5G4, Q9Y5G5, Q9Y5G6, Q9Y5G7, Q9Y5G8, Q9Y5G9, Q9Y5H0, Q9Y5H1, Q9Y5H2, Q9Y5H3, Q9Y5H4	
9	VLLAGFKPPL	RNF17	Q9BXT8	A*02
10	RYSDSVGRVSF	CAPN13	Q6MZZ7	A*24
11	SYSDLHYGF	CAPN13	Q6MZZ7	A*24
12	KYEKIFEML	CT45A3, CT45A4, CT45A5, CT45A6, CT45A1, CT45A2	Q5DJT8	A*24
13	VYTFLSSTL	ESR1	P03372	A*24
14	FYFPTPTVL	FOLR1	P15328	A*24
15	VYHDDKQPTF	GXYLT2	A0PJZ3	A*24
16	IYSPQFSRL	MYO3B	Q8WXR4	A*24
17	RFTTMLSTF	OVGP1	Q12889	A*24
18	KYPVHIYRL	RAD54B	Q9Y620	A*24
19	KYVKVFHQF	ZNF90, ZNF93, ZNF486	Q96H40	A*24
20	RMASPVNVK	C2orf88	Q9BSF0	A*03
21	AVRKPIVLK	CDCA5	Q96FF9	A*03



U.S. Patent Dec. 12, 2023 Sheet 25 of 30 US 11,840,562 B2

Figure 4A

Figure 4B

- **Average document length: 24000 words**
- **Technical and legal jargon,** often only interpretable by SMEs and attorneys
- Images, tables, and charts are frequently referred to in the text

Methods

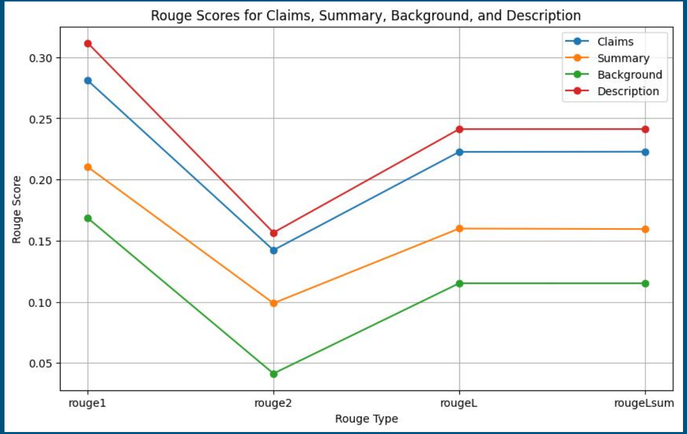
Intuition: **Document length:** Extractive prior to abstractive summarization should mitigate issue.

Technical jargon: Domain-specific fine-tuning might improve the model’s performance in that space.

Dataset	● Harvard USPTO Patent Dataset ● Organic Chemistry class (2% of overall dataset) ● Inputs: Claims, background, summary, description ● Labels: Abstract
Basemodel	● Pegasus-large
Improvements Tried	● Select best sections for input ids ● Hybrid summarization ○ SumBasic Pegasus-large → ● Fine-tuning on domain-specific data
Evaluation	● ROUGE (esp. ROUGE-L) ● Qualitative review

Results

- Best inputs: description and claims
- Hybrid summarization didn't improve quantitative results, but domain-specific fine-tuning did.



		Dataset	
Model	ROUGE	Description	Extracted
Pegasus -large	R1	31.16	22.51
	R2	15.66	8.57
	RL	24.11	15.81
	RLsum	24.15	15.8
Fine tuned Pegasus -large	R1	**44.68**	29.66
	R2	**27.94**	13.34
	RL	**37.52**	23.42
	RLsum	**40.42**	25.09

Baseline

Limitations & Future Research

Limitations:

- Memory constraints limited the dataset to <1000 observations
- Data leakage due to Pegasus’ training set
  - Overlap between BigPatent and HUPD

Future work:

- TextRank for extractive summarization
- Leverage newer patents and/or proprietary summaries from industrial stakeholders to reduce data leakage
- Multimodal models to incorporate images

Data Leakage Example	
Abstract	The present invention relates to an antibody construct comprising a first human binding domain specific for the extracellular part of the influenza envelope protein M2 (M2e) and a second domain specific for CD3....
Summary of Description	<pad>FIELD OF THE INVENTION The present invention relates to an antibody construct comprising a first human binding domain specific for the extracellular part of the influenza envelope protein M2 and a second domain specific for CD3....

Conclusion

- We established a proof-of-concept for domain-specific fine-tuning of patent summaries, which can serve as a starting point for R&D organizations or patent offices wishing to develop this capability.
- We explored the utility of hybrid summarization applied to patent documents and saw promising qualitative results.
- There is still much opportunity for optimization of patent summaries, both at the single- and multi- document level.