In [1]:

```
# os == operating system  # chdir == used to change the current working directory to the
import os
os.chdir(r"D:\Sabina\web Scrapping Project")
```

In [2]:

```
# ElementTree (represents the whole XML document as a tree) module  provides a simple an
# to parse(method of translating code into machine languageto investigate the appropriat
# widely used for data structuring), manipulate,and generate XML documents in Python

import xml.etree.ElementTree as ET
```

In [3]:

```
# parse==conert these structure to another structure
# Representing the entire XML structure & get root element of xml & converts it to a str

tree = ET.parse("769952.xml")
root = tree.getroot()
root=ET.tostring(root, encoding='utf8').decode('utf8')
```

In [4]:

```
print(root)
```

```
<?xml version='1.0' encoding='utf8'?>
<article>
    <front>
        <journal-meta>
            <journal-id journal-id-type="publication">0901c7918047d0e2</jo
urnal-id>
            <journal-id journal-id-type="publisher" />
            <journal-title>Orphan Drug Approvals</journal-title>
            <abbrev-journal-title abbrev-type="publication" />
            <issn />
            <publisher>
                <publisher-name>
                    <p>WebMD, LLC</p>
                </publisher-name>
            </publisher>
            <notes notes-type="support-page">
                <p>index</p>
            </notes>
        </journal-meta>
        <article-meta>
            <article-id>0901c79180555528</article-id>
            <article-categories>
                <subj-group>
                    <subject>News Alert</subject>
                </subj-group>
                <series-title />
            </article-categories>
            <title-group>
                <article-above-title />
                <article-title>FDA Grants Orphan Drug Status to Gevokizuma
b</article-title>
                <subtitle />
                <alt-title>The FDA has granted orphan drug designation to
gevokizumab for the treatment of noninfectious intermediate uveitis, poste
rior uveitis, or panuveitis, or chronic noninfectious anterior uveitis.</a
lt-title>
            </title-group>
            <contrib-group>
                <contrib contrib-type="Journalist">
                    <name>
                        <surname>Troy Brown</surname>
                    </name>
                    <role>Journalist</role>
                    <bio>
                        <p>Troy Brown is a freelance writer for Medscape.
</p>
                    </bio>
                    <author-comment>
                        <title>Disclosure</title>
                        <p>Troy Brown has disclosed no relevant financial
relationships.</p>
                    </author-comment>
                    <author-comment>
                        <title>Title</title>
                        <p />
                    </author-comment>
                </contrib>
            </contrib-group>
            <pub-date>
                <day>29</day>
                <month>08</month>
```

```
                <year>2012</year>
            </pub-date>
            <volume />
            <issue />
            <fpage />
            <lpage />
            <copyright-year />
            <copyright-statement />
            <kwd-group>
                <kwd>choroiditis,cyclitis,intermediate uveitis,orphan drug
s,pars planitis,posterior uveitis</kwd>
            </kwd-group>
            <history>
                <date date-type="posting">
                    <day>29</day>
                    <month>08</month>
                    <year>2012</year>
                </date>
            </history>
        </article-meta>
    </front>
    <body>
        <sec sec-type="page">
            <title />
            <sec sec-type="Default">
                <title />
                <sec sec-type="section">
                    <title />
                    <p>August 29, 2012 — The US Food and Drug Administrati
on (FDA) has granted orphan drug status to gevokizumab (<italic>Xoma 052</
italic>, Xoma Corp), a monoclonal antibody that binds strongly to interleu
kin 1β (IL-1β), for the treatment of noninfectious intermediate uveitis, p
osterior uveitis, or panuveitis, or chronic noninfectious anterior uveiti
s.</p>
<p>The Orphan Drug Act of 1983 was passed to encourage companies to develo
p treatments for rare diseases (diseases that affect fewer than 200,000 pe
ople in the United States). Because the market is so small, such treatment
s can be unprofitable to develop. Companies that develop orphan drugs rece
ive a 50% tax credit for the cost of conducting human clinical trials, 7-y
ear marketing exclusivity, and other incentives.</p>
<p>Behçet's disease is a rare multisystem disease that causes blood vessel
inflammation throughout the body. Common symptoms are mouth sores, genital
sores, and a type of panuveitis known as Behçet's uveitis, an inflammation
of the uvea, retina, and vitreous humor that can lead to retinal detachmen
t, vitreous hemorrhage, glaucoma, and blindness.</p>
<p>"A genetic association has been shown between Behçet's disease and the
IL-1 gene cluster, and IL-1β has been implicated as a mediator in Behçet's
disease pathogenesis," Christine Kay, MD, the director of Retinal Clinical
Research and the director of the Electrophysiology Service in the Vitreore
tinal Division of the Department of Ophthalmology at the University of Flo
rida in Gainesville, told <italic>Medscape Medical News</italic>. Dr. Kay
is a clinical correspondent for the American Academy of Ophthalmology.</p>
<p>"Gevokizumab regulates the activation of IL-1 receptors and can be intr
avenously or subcutaneously administered," Dr. Kay added.</p>
<p>Patients with Behçet's uveitis have few treatment options. "There are c
urrently only 2 drugs FDA-approved for the treatment of chronic noninfecti
ous intermediate, posterior, and panuveitis (<italic>Retisert</italic> [Ba
usch &amp;amp; Lomb] and <italic>Ozurdex</italic> [Allergan]), and both are ex
tended-release corticosteroid ocular implants," Dr. Kay said.</p>
<p>Results of a proof-of-concept phase 2 trial of intravenous gevokizumab
in 7 patients with Behçet's uveitis were published in the April issue of t
```

```
he <italic>Annals of Rheumatic Diseases</italic>. In that trial patients w
ere given a single infusion of gevokizumab (0.3 mg/kg), and all patients e
xperienced complete reduction of intraocular inflammation in between 4 and
21 days (median, 14 days). There were no treatment-related adverse events.
</p>
<p>"In clinical trials, so far, gevokizumab has been studied in nearly 500
patients. The studies have shown that gevokizumab is well-tolerated, and n
o drug-related adverse events have been reported," Fred Kurland, chief fin
ancial officer of Xoma, said in an email interview with <italic>Medscape M
edical News</italic>.</p>
<p>Although it appears that gevokizumab "may offer a viable treatment opti
on in Behçet's disease, it remains to be seen if an IL-1 antibody will hav
e an effect in other forms of noninfectious uveitis. A phase 3 clinical tr
ial to evaluate the efficacy of [gevokizumab] in the treatment of noninfec
tious uveitis is in the recruitment process," Dr. Kay said.</p>
<p>"Gevokizumab does offer the possibility of a pathophysiology-driven tar
geted therapy for IL-1 related uveitis, and if proven safe and effective i
n a phase 3 trial, this could provide a valuable option in the treatment o
f noninfectious intermediate uveitis, posterior uveitis, and panuveitis. E
ven if this drug is only shown to be effective in Behçet's disease, this c
ould provide a useful and targeted treatment for an extremely aggressive c
ondition, perhaps limiting broader and more toxic immunosuppression," Dr.
Kay said.</p>
<p>
                        <bold>Other Potential Indications</bold>
                    </p>
<p>"As an IL-1β inhibitor, gevokizumab has potential in a very large numbe
r of indications that are driven by inflammation, such as noninfectious uv
eitis.... [W]e are also engaged in 2 proof-of-concept phase 2 trials using
gevokizumab in patients with moderate to severe acne vulgaris and in erosi
ve osteoarthritis of the hand, and we will initiate a third proof-of-conce
pt trial in another indication later this year," Kurland explained.</p>
<p>"With respect to the [noninfectious uveitis] market specifically, we es
timate that there are approximately 150,000 patients in the [United States
who have noninfectious uveitis]," Kurland added, noting they are not discu
ssing the drug's pricing yet.</p>
<p>
                        <italic>Dr. Kay has disclosed no relevant financia
l relationships.</italic>
                    </p>
                </sec>
            </sec>
        </sec>
    </body>
    <back>
        <ref-list>
            <title>References</title>
            <list>
                <list-item>
                    <p />
                </list-item>
            </list>
        </ref-list>
        <ack>
            <title>Acknowledgements</title>
            <p />
        </ack>
        <fn-group>
            <fn fn-type="bkmtr_front">
                <p />
            </fn>
```

```
        </fn-group>
    </back>
</article>
```

In [5]:

```
#re =regular expression used for pattern matching ,manipulating string

import re, string, unicodedata
```

In [6]:

```
# use bs4 library for extract information text from xml or html documents

from bs4 import BeautifulSoup
```

In [7]:

```
# remove HTML tags from the text  <.*?>

def strip_html(text):
    soup = BeautifulSoup(text, "html.parser")
    return soup.get_text()
```

In [8]:

```
# remove text between square brackets & substitute into empty string
#  ^ == begining of the line [*]== removes [] with or witout text (where as []+ == remvo

def remove_between_square_brackets(text):
    return re.sub('\[[^]]*\]', '', text)
```

In [9]:

```
print('')
```

In [10]:

```
# calls the above two functions & removes extra spaces
def denoise_text(text):
    text = strip_html(text)
    text = remove_between_square_brackets(text)
    text=re.sub('  ','',text)
    return text
```

In [11]:

```
# denoise_text() function applies to the XML data (root)
sample = denoise_text(root)
print(sample)
```

0901c7918047d0e2

Orphan Drug Approvals

WebMD, LLC

index

0901c79180555528

News Alert

FDA Grants Orphan Drug Status to Gevokizumab

The FDA has granted orphan drug designation to gevokizumab for the treatme
nt of noninfectious intermediate uveitis, posterior uveitis, or panuveiti
s, or chronic noninfectious anterior uveitis.

Troy Brown

Journalist

Troy Brown is a freelance writer for Medscape.

Disclosure
Troy Brown has disclosed no relevant financial relationships.

Title

29
08
2012

choroiditis,cyclitis,intermediate uveitis,orphan drugs,pars planitis,poste
rior uveitis

29
08
2012

August 29, 2012 — The US Food and Drug Administration (FDA) has granted or
phan drug status to gevokizumab (Xoma 052, Xoma Corp), a monoclonal antibo
dy that binds strongly to interleukin 1β (IL-1β), for the treatment of non
infectious intermediate uveitis, posterior uveitis, or panuveitis, or chro
nic noninfectious anterior uveitis.
The Orphan Drug Act of 1983 was passed to encourage companies to develop t
reatments for rare diseases (diseases that affect fewer than 200,000 peopl
e in the United States). Because the market is so small, such treatments c
an be unprofitable to develop. Companies that develop orphan drugs receive
a 50% tax credit for the cost of conducting human clinical trials, 7-year
marketing exclusivity, and other incentives.
Behçet's disease is a rare multisystem disease that causes blood vessel in
flammation throughout the body. Common symptoms are mouth sores, genital s
ores, and a type of panuveitis known as Behçet's uveitis, an inflammation
of the uvea, retina, and vitreous humor that can lead to retinal detachmen
t, vitreous hemorrhage, glaucoma, and blindness.
"A genetic association has been shown between Behçet's disease and the IL-
1 gene cluster, and IL-1β has been implicated as a mediator in Behçet's di
sease pathogenesis," Christine Kay, MD, the director of Retinal Clinical R
esearch and the director of the Electrophysiology Service in the Vitreoret
inal Division of the Department of Ophthalmology at the University of Flor
ida in Gainesville, told Medscape Medical News. Dr. Kay is a clinical corr
espondent for the American Academy of Ophthalmology.
"Gevokizumab regulates the activation of IL-1 receptors and can be intrave
nously or subcutaneously administered," Dr. Kay added.
Patients with Behçet's uveitis have few treatment options. "There are curr
ently only 2 drugs FDA-approved for the treatment of chronic noninfectious
intermediate, posterior, and panuveitis (Retisertand Ozurdex), and both a
re extended-release corticosteroid ocular implants," Dr. Kay said.
Results of a proof-of-concept phase 2 trial of intravenous gevokizumab in
7 patients with Behçet's uveitis were published in the April issue of the
Annals of Rheumatic Diseases. In that trial patients were given a single i
nfusion of gevokizumab (0.3 mg/kg), and all patients experienced complete
reduction of intraocular inflammation in between 4 and 21 days (median, 14
days). There were no treatment-related adverse events.
"In clinical trials, so far, gevokizumab has been studied in nearly 500 pa
tients. The studies have shown that gevokizumab is well-tolerated, and no
drug-related adverse events have been reported," Fred Kurland, chief finan

cial officer of Xoma, said in an email interview with Medscape Medical News.

Although it appears that gevokizumab "may offer a viable treatment option in Behçet's disease, it remains to be seen if an IL-1 antibody will have an effect in other forms of noninfectious uveitis. A phase 3 clinical trial to evaluate the efficacy ofin the treatment of noninfectious uveitis is in the recruitment process," Dr. Kay said.

"Gevokizumab does offer the possibility of a pathophysiology-driven targeted therapy for IL-1 related uveitis, and if proven safe and effective in a phase 3 trial, this could provide a valuable option in the treatment of noninfectious intermediate uveitis, posterior uveitis, and panuveitis. Even if this drug is only shown to be effective in Behçet's disease, this could provide a useful and targeted treatment for an extremely aggressive condition, perhaps limiting broader and more toxic immunosuppression," Dr. Kay said.

Other Potential Indications

"As an IL-1β inhibitor, gevokizumab has potential in a very large number of indications that are driven by inflammation, such as noninfectious uveitis.... e are also engaged in 2 proof-of-concept phase 2 trials using gevokizumab in patients with moderate to severe acne vulgaris and in erosive osteoarthritis of the hand, and we will initiate a third proof-of-concept trial in another indication later this year," Kurland explained.

"With respect to themarket specifically, we estimate that there are approximately 150,000 patients in the ," Kurland added, noting they are not discussing the drug's pricing yet.

Dr. Kay has disclosed no relevant financial relationships.

References

Acknowledgements

```
C:\Users\SAMEER\anaconda3\lib\site-packages\bs4\builder\__init__.py:545: X
MLParsedAsHTMLWarning: It looks like you're parsing an XML document using
an HTML parser. If this really is an HTML document (maybe it's XHTML?), yo
u can ignore or filter this warning. If it's XML, you should know that usi
ng an XML parser will be more reliable. To parse this document as XML, mak
e sure you have the lxml package installed, and pass the keyword argument
`features="xml"` into the BeautifulSoup constructor.
  warnings.warn(
```

In [12]:

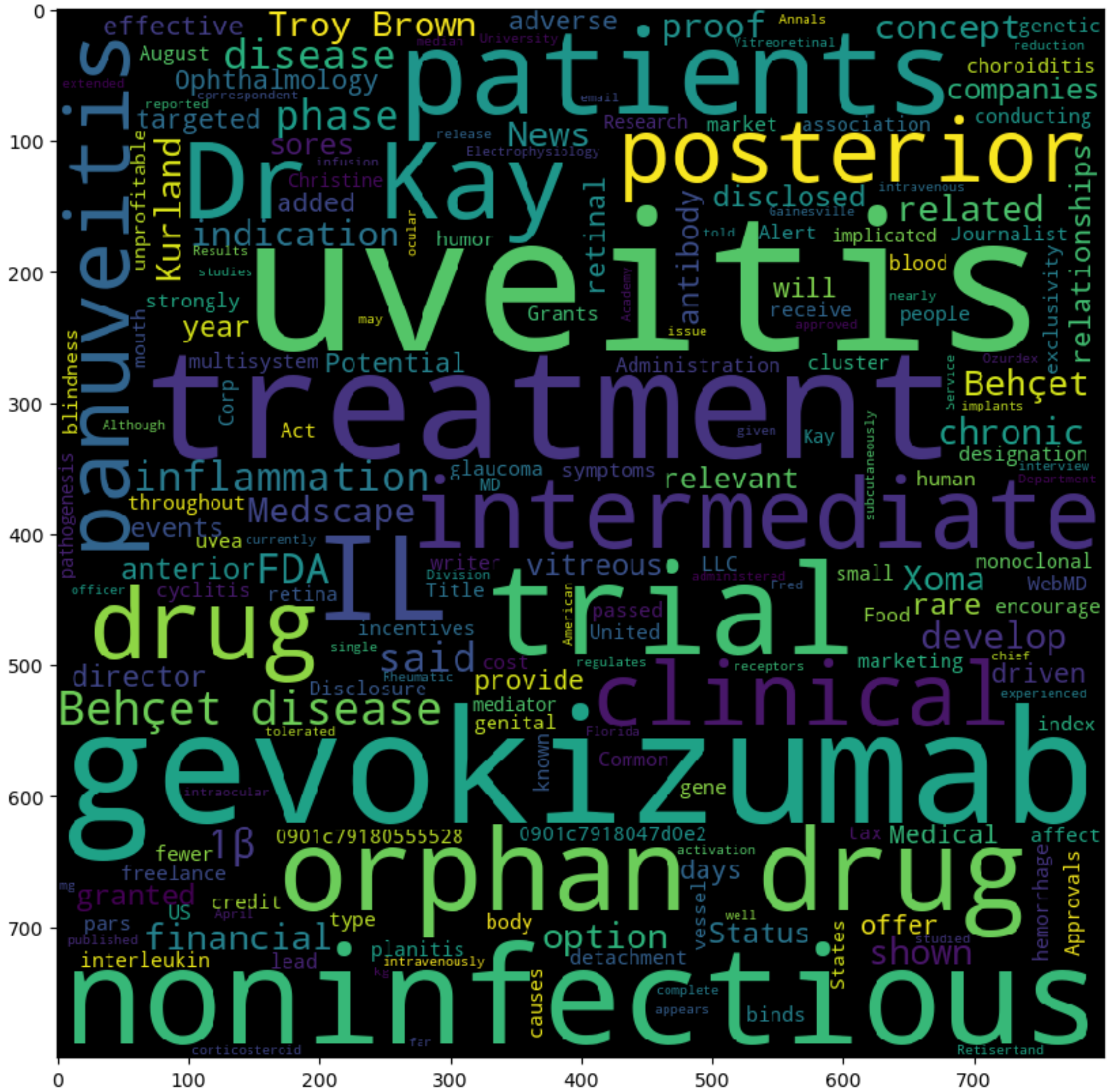
```python
# visualize the most frequent words in the text
from wordcloud import WordCloud
import matplotlib.pyplot as plt
```

In [13]:

```python
wordcloud = WordCloud(width=800, height=800, margin=2).generate(sample)
```

In [14]:

```python
plt.figure(figsize=(10,10))
plt.imshow(wordcloud, interpolation='nearest')
plt.axis("on")
plt.margins(x=0, y=0)
plt.show()
```



In [ ]: