

TRABAJO FIN DE GRADO

Automatización de extracción de datos
de informes de secuenciación masiva
del software *Oncomine Reporter*

Lucía Vítores López

ÍNDICE

01

Introducción

02

Objetivos

03

Metodología

04

Resultados

05

Líneas futuras

06

Conclusiones

1. INTRODUCCIÓN

**Oncome
Reporter**



**Tablas
manuales**

**¿VENTAJA O
INCONVENIENTE?**

Google Académico

Artículos

Cualquier momento
Desde 2023

Su consulta, **"PDF scraping" ADN "Oncomine Reporter"** , no ha obtenido ningún artículo como resultado.

☰ Google Académico

2. OBJETIVOS

Software/Hardware

1. Código capaz de facilitar la extracción de datos de ficheros PDF.
2. Interfaz sencilla para que cualquier persona sea capaz de usarla correctamente.

Técnicos

1. Revisión bibliográfica PDFscraping.
2. Código lectura para PDF.
3. Código de selección de variables de interés.
4. Tablas específicas.
5. Exportar tablas a .xlsx.
6. Material complementario para ayudar a su implementación.



3. METODOLOGÍA



PDFscraping: técnica de extracción automática de los datos y del contenido almacenado en archivos PDF.

Python VS R

Fitz
Pandas
Numpy
Os
Re

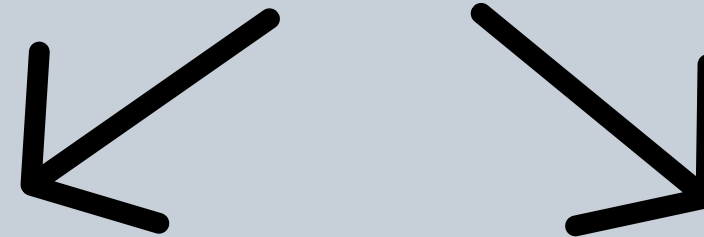


Pdftools
Tools
Tidyverse
Purrr
Readxl

4. RESULTADOS

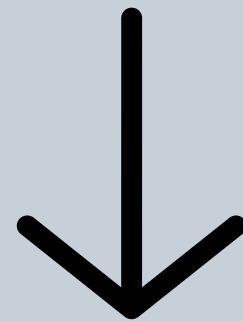


Ficheros resultantes en formato .xlsx

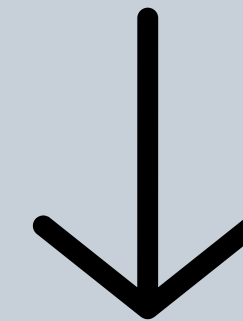


Histórico general

Histórico patológico



Parcial general {i}



Parcial patológico {i}

Tipo de cáncer de la muestra: Carcinoma pulmonar no microcítico

Resultados del informe

1 Biomarcadores relevantes

6 Tratamientos disponibles

6 Ensayos clínicos

Biomarcadores relevantes

Nivel	Alteración genómica	Tratamientos relevantes (En este tipo de cáncer)	Tratamientos relevantes (En otro tipo de cáncer)	Ensayos clínicos
IA	<i>EML4-ALK fusion</i> EMAP like 4 - ALK receptor tyrosine kinase	alectinib ^{1, 2} brigatinib ^{1, 2} ceritinib ^{1, 2} crizotinib ^{1, 2} lorlatinib ^{1, 2} next-generation ALK inhibitor	crizotinib ¹ alectinib brigatinib ceritinib lorlatinib	6

Fuentes de datos públicos incluidas en tratamientos pertinentes: FDA1, NCCN, EMA2, ESMO

Detalles de la variante

Variantes de secuencia de ADN									
Gen	Cambio de aminoácidos	Codificación	Exón	Frecuencia de alelo	Locus	Transcripto	ID de la variante	Efecto de la variante	ClinVar ¹
ALK	p.(D1529E)	c.4587C>G	29	50.98%	chr2:29416366	NM_004304.4	.	missense	Benign
ALK	p.(I1461V)	c.4381A>G	29	99.85%	chr2:29416572	NM_004304.4	.	missense	Benign
FGFR4	p.(P136L)	c.407C>T	4	97.85%	chr5:176517797	NM_213647.2	.	missense	
MYC	p.(N26S)	c.77A>G	2	49.56%	chr8:128750540	NM_002467.4	.	missense	

1 Basado en la versión ClinVar 20180225

Fusiones de genes (ARN)				
Genes	ID de la variante	Locus	Numero de lecturas	ClinVar ¹
EML4-ALK	EML4-ALK.E6aA20.AB374361	chr2:42491871 - chr2:29446394	404	
EML4-ALK	EML4-ALK.E6bA20.AB374362	chr2:42492091 - chr2:29446394	196	

1 Basado en la versión ClinVar 20180225

Fdo. Dra. Smith Peterson, Palm (Bióloga Molecular) / Dr. García García, Juan (Jefe de Servicio)

Declaración: Los datos aquí presentados proceden de una base de conocimientos seleccionada con información de acceso público, pero puede que no sean exhaustivos. La versión de datos es 2023.02(005).El contenido de este informe no ha sido evaluado ni aprobado por la FDA ni por otras autoridades de registro sanitario.



Lectura de dos tipos de informes distintos

Indicación y alcance

Estudio de secuenciación masiva (NGS-Oncomine Focus), validado para muestras FFPE. Este caso ha sido seleccionado por el Comité Molecular de Tumores del HUBU, siendo **responsable de la propuesta e integración de la información obtenida el facultativo solicitante**, quien informa al paciente sobre las limitaciones y objetivos del estudio para obtener su consentimiento informado. Los datos presentados sólo deben interpretarse por profesionales cualificados y deben valorarse en el contexto clínico del paciente.

Tipo de cáncer de la muestra: Carcinoma pulmonar no microcítico

Datos de la muestra:

- Diagnóstico: Adenocarcinoma de pulmón
- % de células tumorales: 100 / Sexo: FEM
- Tipo de muestra: ADN y ARN obtenidos de BAG ADENOPATÍA

Resultados del informe

1 Biomarcadores relevantes

6 Tratamientos disponibles

6 Ensayos clínicos

Biomarcadores relevantes

Nivel	Alteración genómica	Tratamientos relevantes (En este tipo de cáncer)	Tratamientos relevantes (En otro tipo de cáncer)	Ensayos clínicos
IA	<i>EML4-ALK fusion</i> EMAP like 4 - ALK receptor tyrosine kinase	alectinib ^{1, 2} brigatinib ^{1, 2} ceritinib ^{1, 2} crizotinib ^{1, 2} lorlatinib ^{1, 2} atezolizumab + bevacizumab + chemotherapy	crizotinib ¹ alectinib brigatinib ceritinib lorlatinib	6

Fuentes de datos públicos incluidas en tratamientos pertinentes: FDA1, NCCN, EMA2, ESMO

Detalles de la variante

Variantes de secuencia de ADN									
Gen	Cambio de aminoácidos	Codificación	Exón	Frecuencia de alelo	Cobertura	Locus	ClinVar ¹	Transcripto	Efecto de la variante
ALK	p.(D1529E)	c.4587C>G	29	50.98%	1999	chr2:29416366	Benign	NM_004304.4	missense
ALK	p.(I1461V)	c.4381A>G	29	99.85%	2000	chr2:29416572	Benign	NM_004304.4	missense
FGFR4	p.(P136L)	c.407C>T	4	97.85%	2000	chr5:176517797		NM_213647.2	missense
MYC	p.(N26S)	c.77A>G	2	49.56%	1461	chr8:128750540		NM_002467.4	missense

1 Basado en la versión ClinVar 20180225

Resultados obtenidos de los ficheros proporcionados.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Número de p de p			NHC	Número de biopsia	so	fecha de inform	Diagnóstico	o del diag	utaciones detectad	la mutació	úmero de	de frecuencia al	Fusiones ID	sayos clíni	/NO ensay	naco aprob	NO fármacc
2	0	100	1	21	23B00021-A1	1	24-mar-2023	Carcinoma pulmona	15,1	['MYC']	[40]	1	['49.56']	['EML4-ALK.E6aA	6	1	6	1
3	1	100	2	22	22B00022-A2	1	24-mar-2023	Carcinoma pulmona	15,1	['KRAS']	[35]	1	['66.50']	[]	5	1	1	1
4	2	100	3	23	21B00023-A1	1	24-mar-2023	Carcinoma pulmona	15,1	[]	[]	0	[]	[]	0	0	0	0
5	3	100	4	24	23B00024-A1	1	24-mar-2023	Carcinoma pulmona	15,1	['FGFR1', 'PIK3CA']	[22, 47]	2	[]	[]	0	0	0	0
6	4	100	5	25	23C00025-A1	3	24-mar-2023	Carcinoma pulmona	15,1	['CDK6', 'MET']	[10, 38]	2	['21.75', '21.75']	[]	3	1	3	1
7	5	100	6	26	23P00026	2	24-mar-2023	Carcinoma pulmona	15,1	['CTNNB1', 'EGFR']	[11, 13]	2	['27.51', '35.40']	[]	5	1	0	0
8	6	100	7	27	22B00027-C4	1	24-mar-2023	Cáncer tiroideo	30	['HRAS']	[28]	1	['46.17']	[]	1	1	0	0
9	7	100	8	28	23B00028-A2	1	24-mar-2023	Cáncer gástrico	11	['KIT', 'MYC']	[34, 40]	2	['53.12', '60.64']	[]	0	0	1	1
10	8	100	1	1234567	23B000000-A1/CH	1	25-may-2023	Carcinoma pulmona	15,1	['MYC']	[40]	1	['49.56']	['EML4-ALK.E6aA	6	1	6	1
11	9	100	2	1234567	23B00000-A1/CHI	1	25-may-2023	Carcinoma pulmona	15,1	['KRAS']	[35]	1	['66.50']	[]	4	1	1	1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Número de o de p		NHC	Número de biopsia		opsia sólida	fecha de informe	Diagnóstico	o del diagn		es patogén	la mutación	cuencia al	utaciones	sayos clíni	/NO ensay	naco aprob	NO fármacos
2	0	100	1	21	23B00021-A1	1	24-mar-2023	Carcinoma pulmonar no microcítico	15,1	[]	[]	[]	0	6	1	6	1	
3	1	100	2	22	22B00022-A2	1	24-mar-2023	Carcinoma pulmonar no microcítico	15,1	['KRAS']	[35]	['66.50']	1	5	1	1	1	
4	2	100	3	23	21B00023-A1	1	24-mar-2023	Carcinoma pulmonar no microcítico	15,1	[]	[]	[]	0	0	0	0	0	
5	3	100	4	24	23B00024-A1	1	24-mar-2023	Carcinoma pulmonar no microcítico	15,1	[]	[]	[]	0	0	0	0	0	
6	4	100	5	25	23C00025-A1	3	24-mar-2023	Carcinoma pulmonar no microcítico	15,1	[]	[]	[]	0	3	1	3	1	
7	5	100	6	26	23P00026	2	24-mar-2023	Carcinoma pulmonar no microcítico	15,1	['CTNNB1']	[11]	['27.51']	1	5	1	0	0	
8	6	100	7	27	22B00027-C4	1	24-mar-2023	Cáncer tiroideo	30	[]	[]	[]	0	1	1	0	0	
9	7	100	8	28	23B00028-A2	1	24-mar-2023	Cáncer gástrico	11	[]	[]	[]	0	0	0	1	1	
10	8	100	1	1234567	23B000000-A1/CH	1	25-may-2023	Carcinoma pulmonar no microcítico	15,1	[]	[]	[]	0	6	1	6	1	
11	9	100	2	1234567	23B000000-A1/CH	1	25-may-2023	Carcinoma pulmonar no microcítico	15,1	['KRAS']	[35]	['66.50']	1	4	1	1	1	

5. LÍNEAS FUTURAS

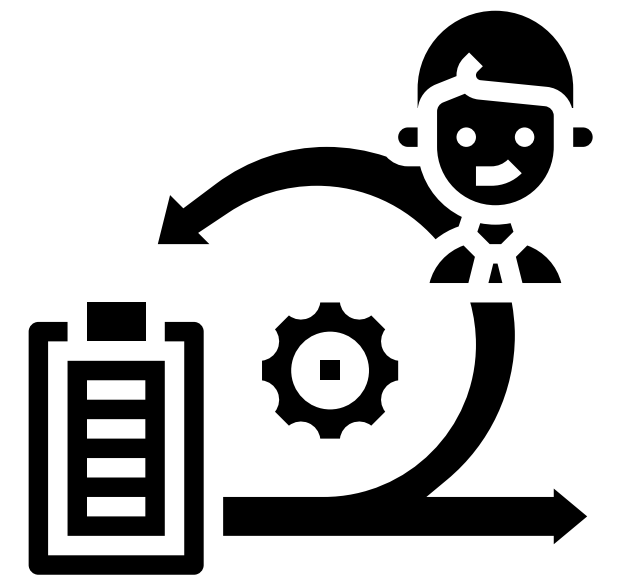
CORTO PLAZO

Código que determine la ruta sin meterla a mano

LARGO PLAZO

Implementar dicho código en otros lenguajes de programación como R. Una vez obtenido este código, se pueda implementar una aplicación alojada en la web como servicio usando por ejemplo el framework Shiny

6. CONCLUSIONES



- Agilizar y simplificar el proceso de extracción.
- Solución personalizada adaptada a las necesidades.
- Reducción del tiempo.
- Disminución del error humano.

MUCHAS GRACIAS POR SU ATENCIÓN.

¿PREGUNTAS O DUDAS?