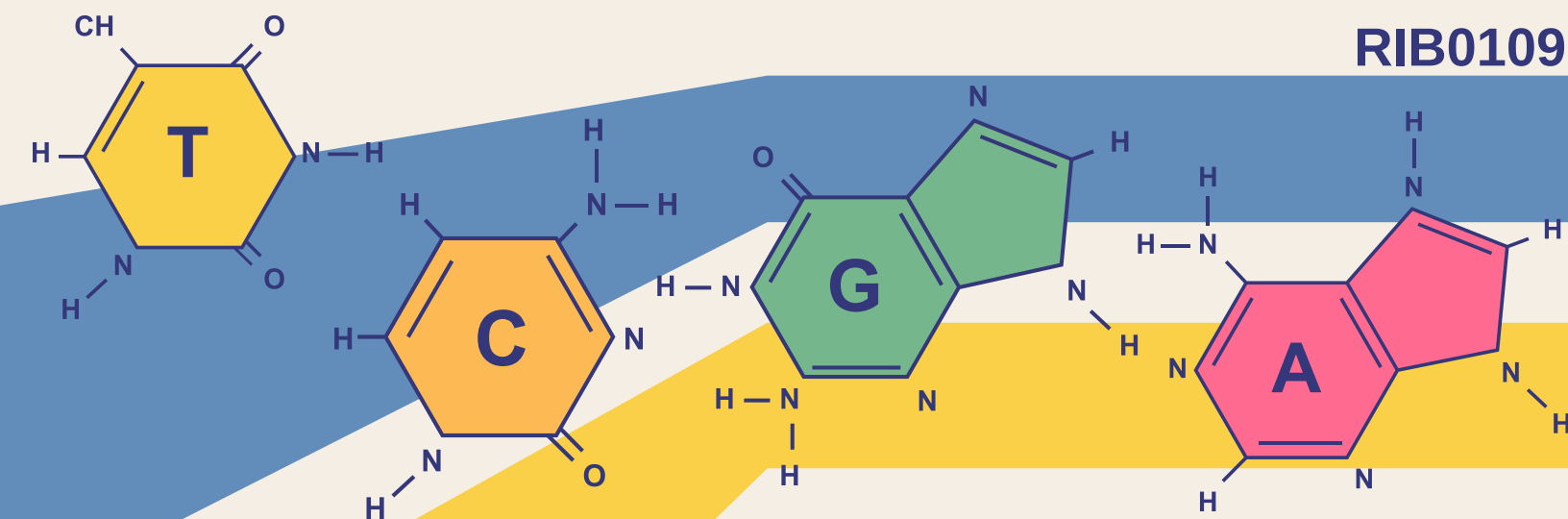




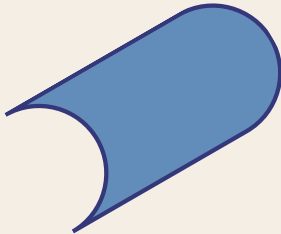
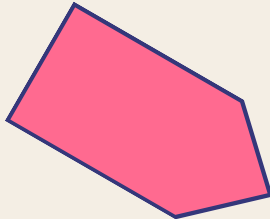
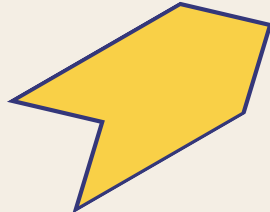
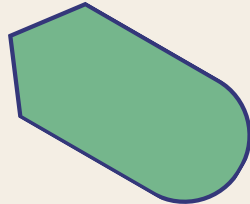
Diferenciação molecular e clínica entre Carcinossarcoma (TCGA-UCS) e Carcinoma Endometrial Uterino (TCGA-UCEC)

Análise de dados genômicos/moleculares

RIB0109 - Fundamentos de Informática em Bioinformática e Medicina Genômica



C
A
C
G
A
C
T
T
C
A
C
G
A
C
T
T



Integrantes

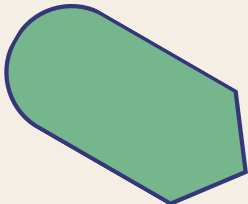
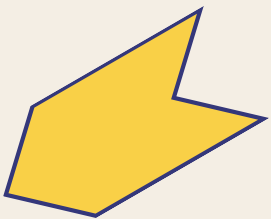
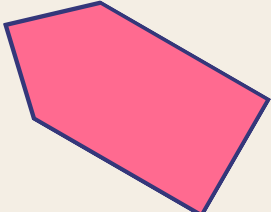
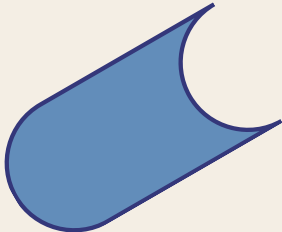
Ana Beatriz Monteiro Fioravanti
beatriz.fioravanti@usp.br
14556271

Beatriz de Oliveira Vieira
beatrizov@usp.br
14685352

Heloisa Ferian Soares
heloisaferian@usp.br
14577959

Maria Fernanda Maia de Oliveira
mariafmaia@usp.br
14595546

Vinícius de Souza Oliveira
vinicius-souza@usp.br
14761322



G
T
G
C
T
G
A
A
G
T
G
C
T
G
A
A



Sumário

- **Introdução**
- **Justificativa**
- **Objetivos**
- **Metodologia**
- **Resultados**
- **Conclusão**
- **Referências Bibliográficas**

Introdução

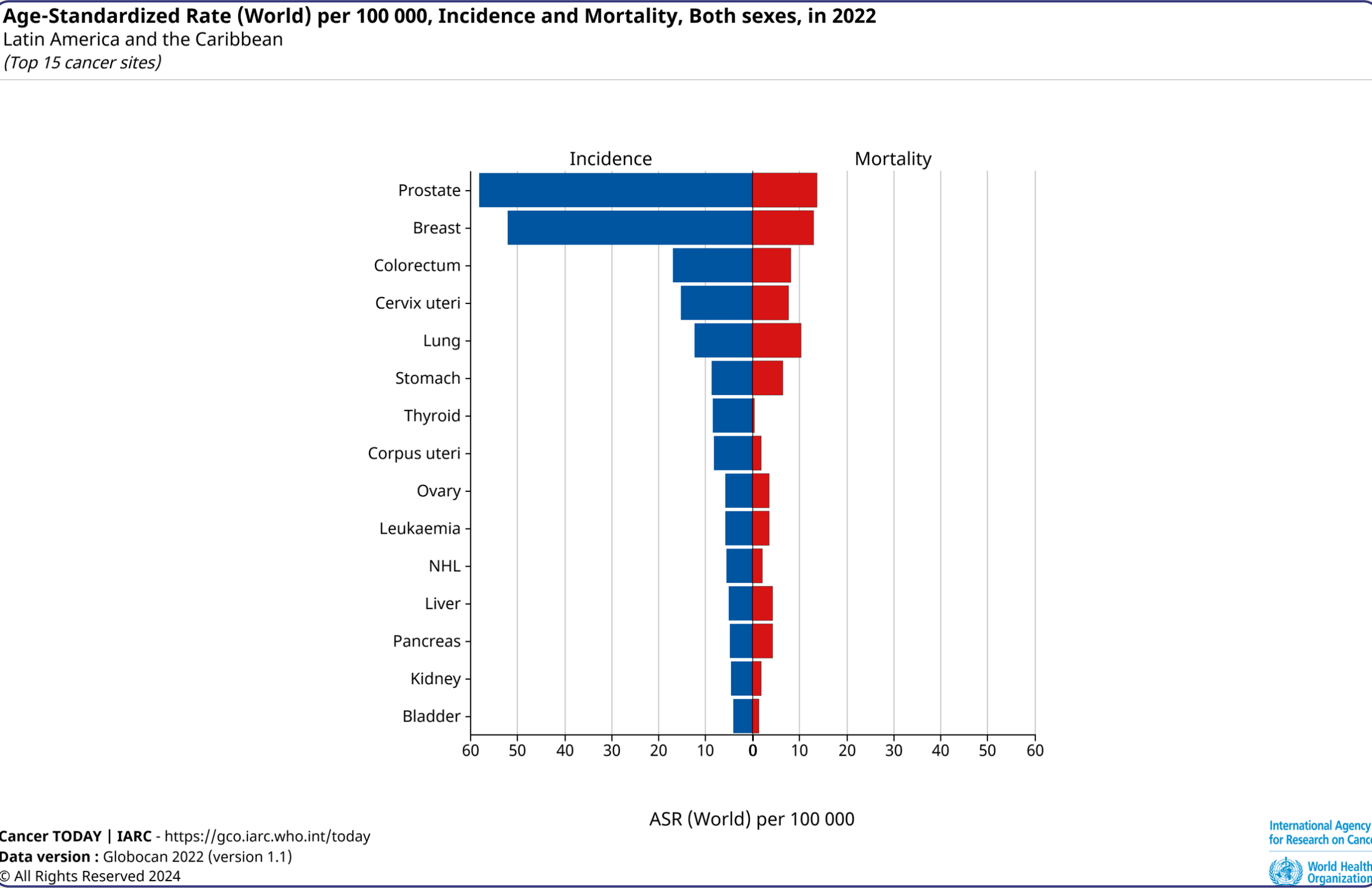


Figura 1- Classificação de câncer por incidência/mortalidade. Fonte: Global Cancer Observatory.

Introdução

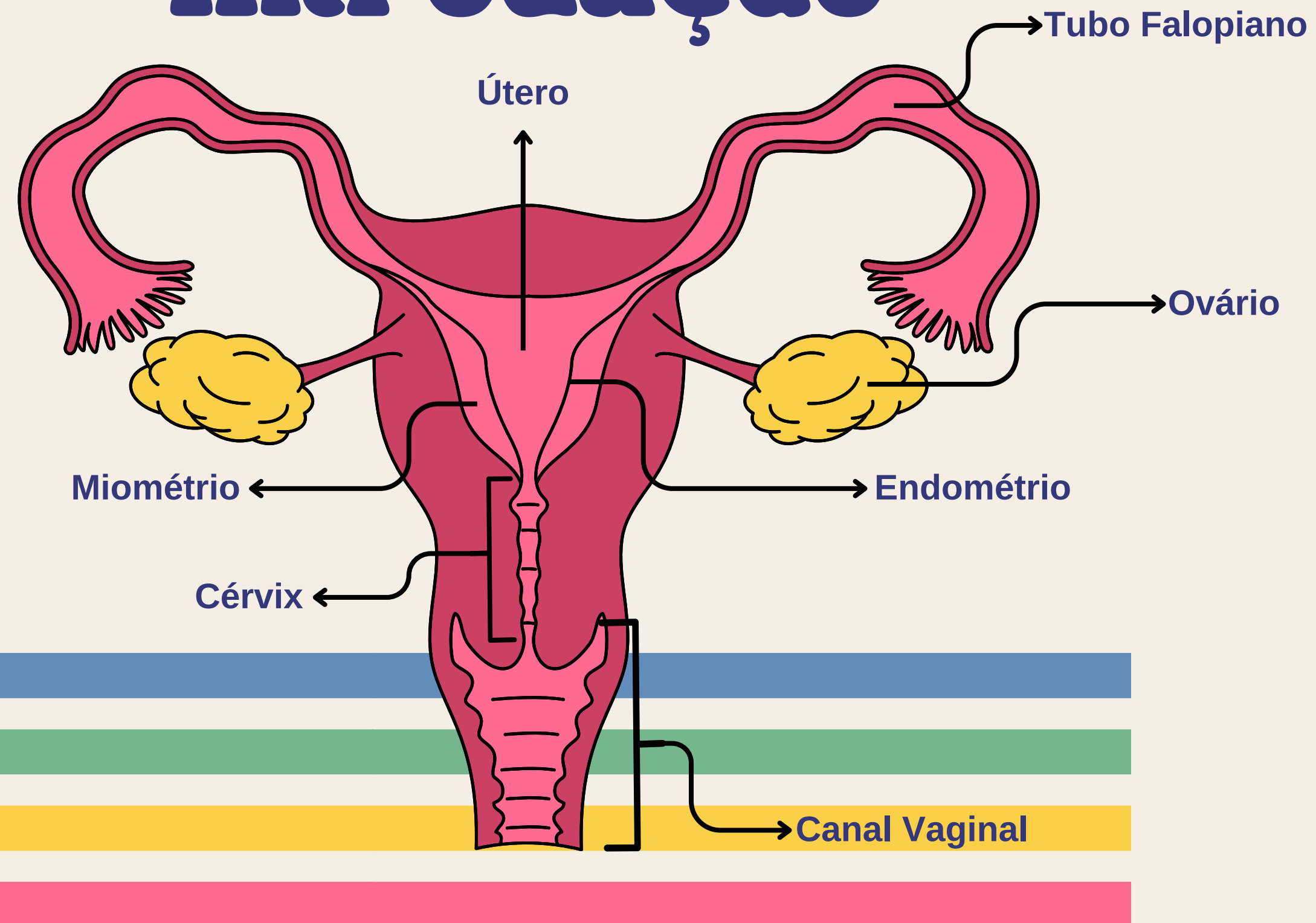
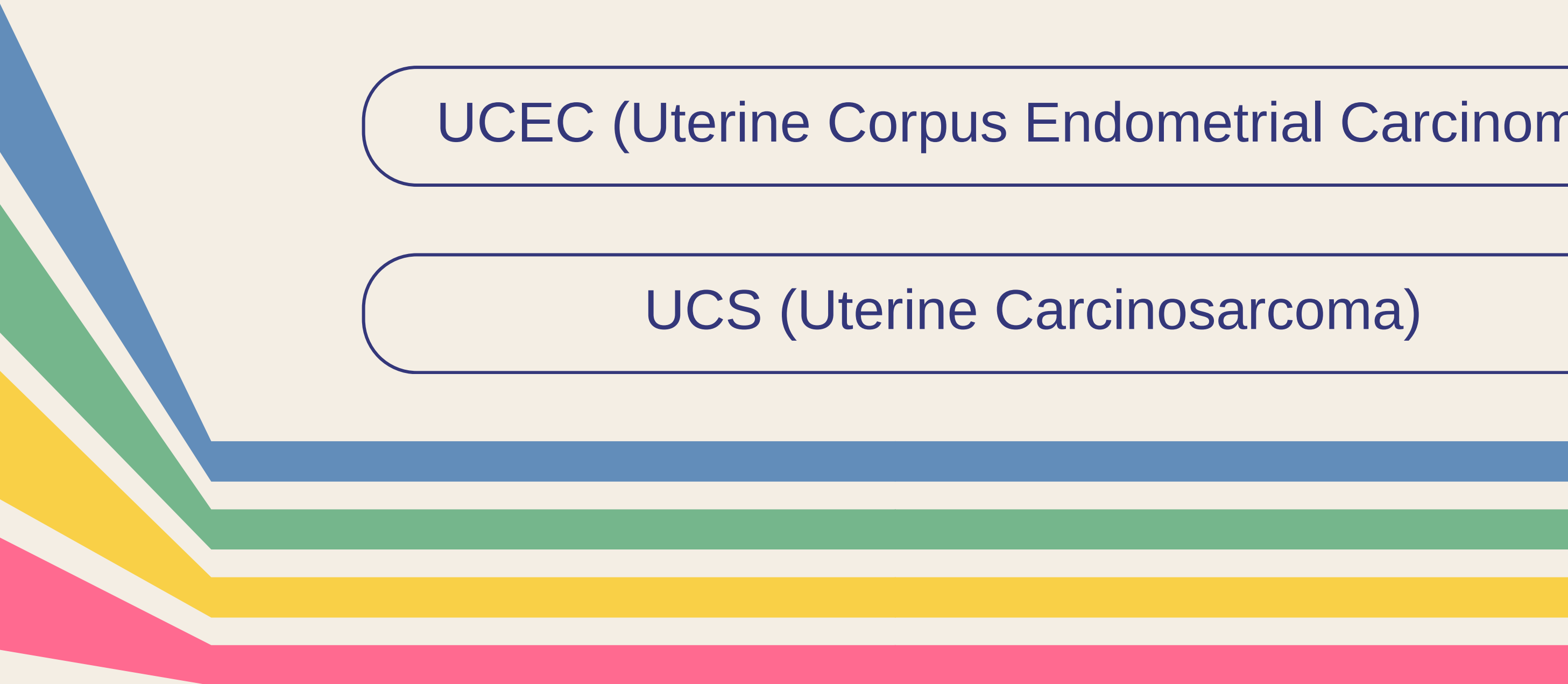


Figura 2- Sistema Reprodutivo Feminino. Fonte: autoria própria.

Introdução

UCEC (Uterine Corpus Endometrial Carcinoma)

UCS (Uterine Carcinosarcoma)

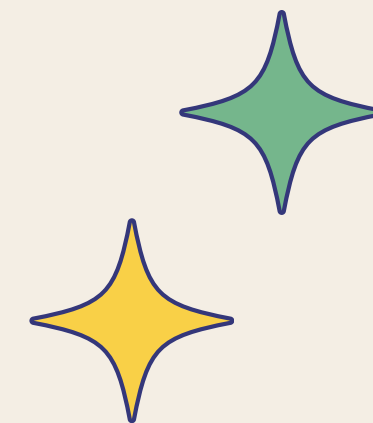
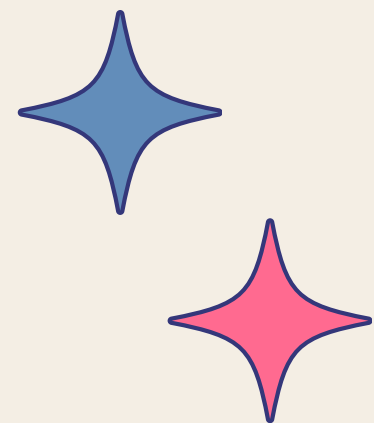


Justificativa

**Elaboração de novas
estratégias de terapias
gênicas**

**Definição de vias
moleculares que possam ser
importantes para prevenção
e tratamento do câncer**

**Compreensão dos
mecanismos e interações
biológicas que estão por
trás dessa doença**



Objetivos

Realizar análises comparativas entre os diferentes tipos de câncer endometrial (UCS e UCEC) à fim de inferir diferenças clínicas e moleculares

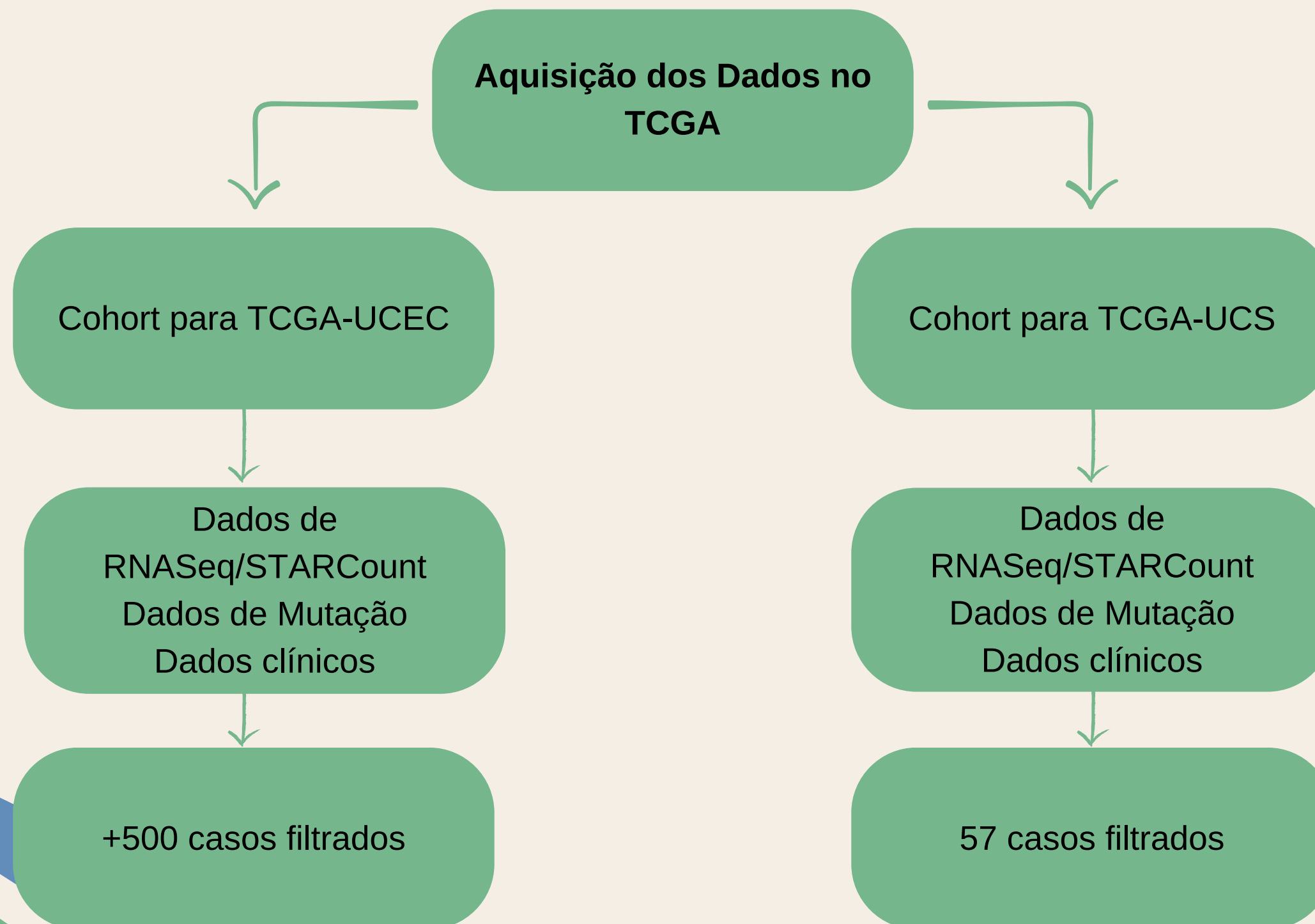
Entender o funcionamento da plataforma do
The Cancer Genome Atlas;

Obter estatísticas sobre os dados analisados;

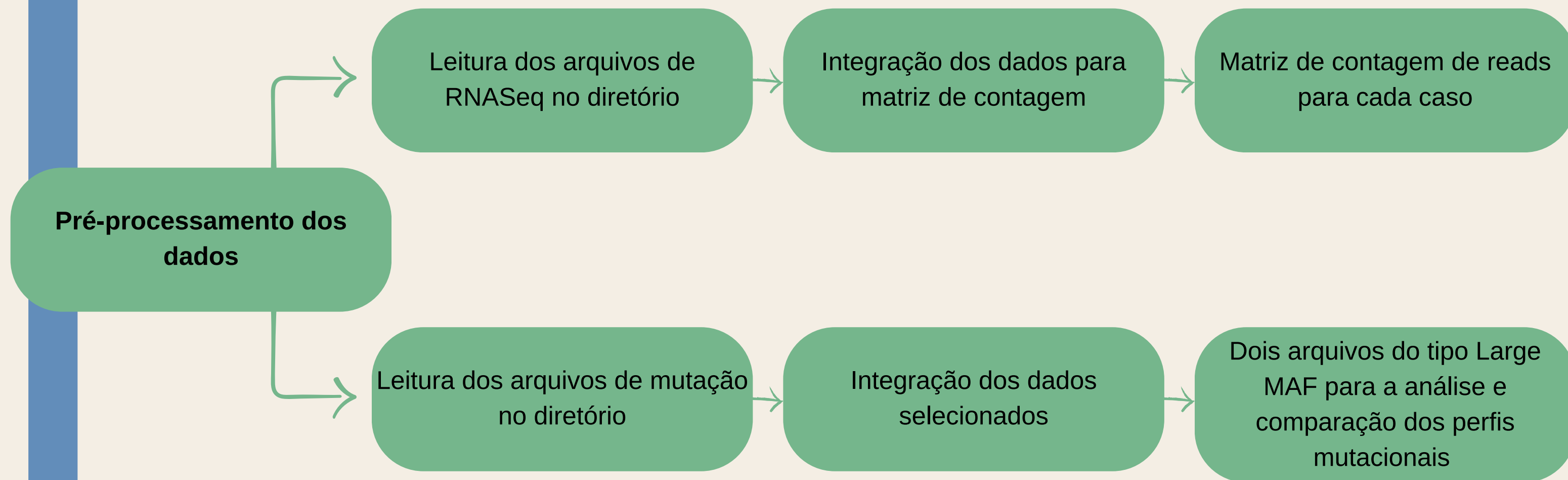
Realizar análises em arquivos de mutação de
exoma, expressão gênica e dados clínicos;

Observar diferenças na prevalência de
mutações.

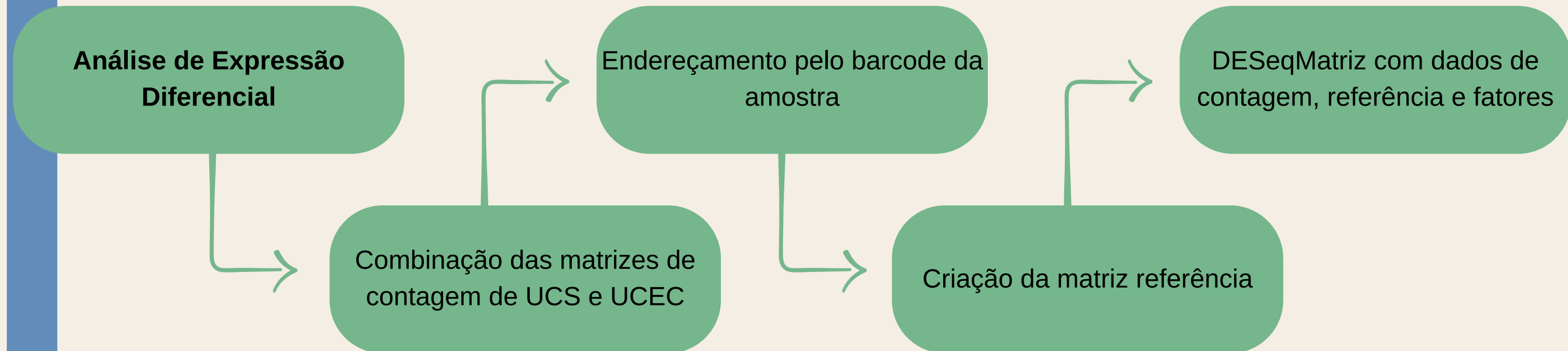
Metodologia



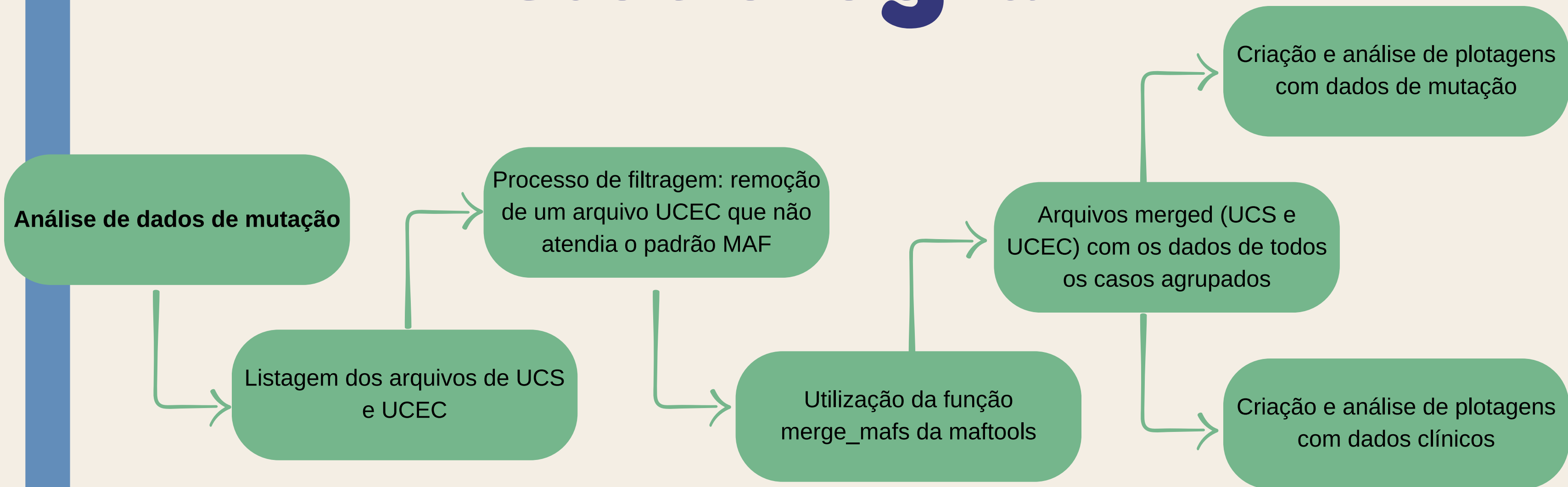
Metodologia



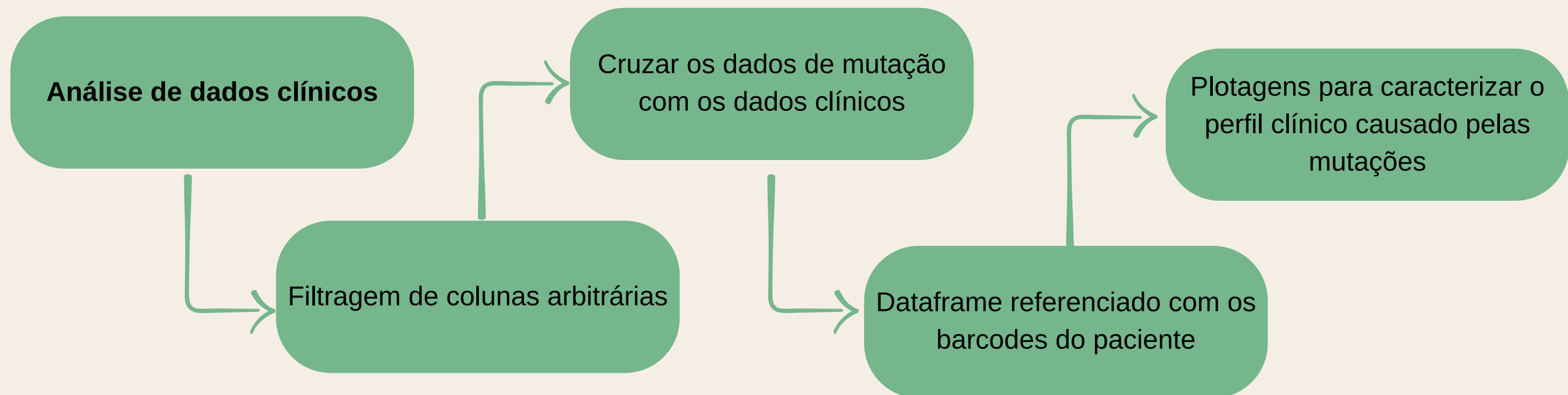
Metodologia



Metodologia



Metodologia



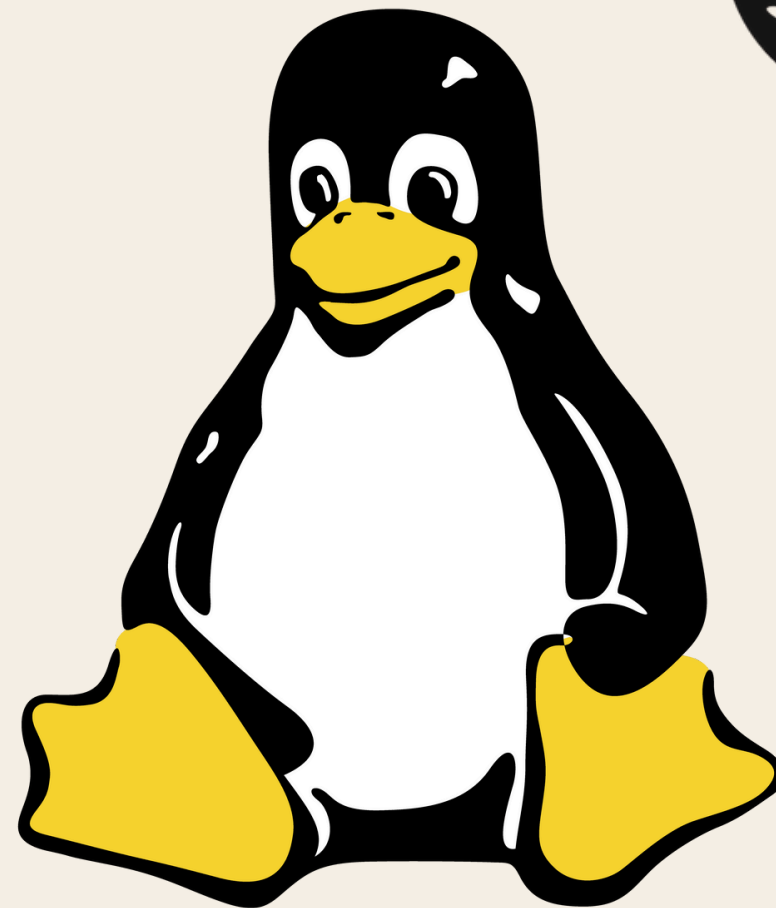
Metodologia



Studio[®]



GitHub



Resultados

script_server.R	script_final_ucs.R	análises_estat.R*	TCGA_UCS_RNASEQ	TCGA_UCS_RNASEQ_MATCHED	script_final_ucec.R	combined_counts	gene_names	raw_counts	res_df	
Filter	Cols: << 1 - 50 >>									
	TCGA-ND-A4WF-01A	TCGA-N5-A4R8-01A	TCGA-N6-A4VD-01A	TCGA-N9-A4Q8-01A	TCGA-N8-A4PN-01A	TCGA-N6-A4V9-01A	TCGA-N8-A4PO-01A	TCGA-NF-A4WX-01A	TCGA-QM-A5NM-01A	TCGA-N5-A4RM-01A
ENSG00000000003.15	4503	2600	6756	6163	3106	8801	6184	3836	2400	
ENSG00000000005.6	25225	548	101	3	55	39	42	112	201	
ENSG000000000419.13	1898	6690	2560	4671	3387	2158	4023	2366	1454	
ENSG000000000457.14	445	772	874	684	904	737	988	858	318	
ENSG000000000460.17	383	610	501	364	1240	497	1076	996	290	
ENSG000000000938.13	139	249	452	918	174	108	236	133	27	
ENSG000000000971.16	574	4046	1374	505	1020	10936	2330	939	907	
ENSG00000001036.14	2164	4161	3032	1754	1429	1964	3433	1748	2852	
ENSG00000001084.13	885	1517	2941	1847	1292	1638	1804	1342	399	
ENSG00000001167.14	2025	3768	2414	3681	4190	3330	4938	2752	1238	
ENSG00000001460.18	209	620	895	627	800	612	576	2070	557	
ENSG00000001461.17	642	828	1554	1401	1116	1054	1432	837	1448	
ENSG00000001497.18	1542	3528	4042	4103	4777	2765	2338	2515	2335	
ENSG00000001561.7	309	666	1220	1966	1494	1028	1010	759	358	
ENSG00000001617.12	12773	7667	4799	14103	4099	9739	6593	6175	4715	
ENSG00000001626.16	7	5	160	161	5	44	10	59	31	
ENSG00000001629.10	2447	2512	3759	2596	2294	5288	4014	2268	1806	
ENSG00000001630.17	54	10	57	2	42	85	50	34	4	
ENSG00000001631.16	323	478	328	207	308	315	597	406	550	
ENSG00000002016.18	174	1183	503	288	710	591	1919	1913	462	
ENSG00000002079.14	4	20	10	2	40	33	13	24	3	
ENSG00000002330.14	741	1252	631	747	656	481	669	935	1134	
ENSG00000002549.12	3034	5451	8549	2045	3315	4930	3026	2932	943	
ENSG00000002586.20	13278	10018	6754	11019	9922	20189	7621	4881	11850	
ENSG00000002586.20_PAR_Y	0	0	0	0	0	0	0	0	0	
Showing 1 to 25 of 60,660 entries, 57 total columns										

Figura 3 - Tabela de contagem de reads de UCS. Fonte: De autoria própria.

Resultados

alises_estat.R*

TCGA_UCS_RNASEQ*

TCGA_UCS_RNASEQ_MATCHED*

TCGA_UCS_MAF_MATCHED*

TCGA_UCS_MAF*

TCGA_UCS_MAF_filtered*

script_final

Filter

	ARID1A	CHD4	FBXW7	LRP2	PIK3CA	PPP2R1A	PTEN	RYR2	TP53	TTN
TCGA-ND-A4WF-01A	0	1	1	0	1	1	0	0	0	0
TCGA-N5-A4R8-01A	0	0	0	0	1	0	0	0	0	0
TCGA-N6-A4VD-01A	0	0	1	0	1	1	1	0	0	0
TCGA-N9-A4Q8-01A	0	0	0	0	0	0	0	0	1	0
TCGA-N8-A4PN-01A	0	0	1	0	0	1	0	1	1	0
TCGA-N6-A4V9-01A	0	0	0	0	0	0	0	0	1	0
TCGA-N8-A4PO-01A	0	0	0	0	0	0	0	1	1	0
TCGA-NF-A4WX-01A	0	0	1	0	0	0	0	0	0	0
TCGA-QM-A5NM-01A	0	0	0	0	0	1	1	0	1	0
TCGA-N5-A4RM-01A	0	0	0	0	0	0	0	0	1	0
TCGA-NA-A4QW-01A	0	0	0	0	0	1	0	0	1	0
TCGA-NF-A4X2-01A	1	0	0	0	0	1	1	0	1	0
TCGA-N8-A4PQ-01A	0	0	0	0	0	0	0	0	1	0
TCGA-NA-A4R1-01A	0	0	0	0	0	0	1	0	1	0
TCGA-N5-A4RO-01A	0	0	1	0	1	0	0	0	1	1
TCGA-N5-A4RJ-01A	0	0	1	0	1	0	0	0	1	0
TCGA-NF-A4WU-01A	0	0	0	0	0	0	0	0	1	0
TCGA-N9-A4Q7-01A	0	0	1	1	0	1	0	0	1	1
TCGA-N5-A4RS-01A	0	0	1	0	1	0	0	0	0	0
TCGA-N5-A59E-01A	0	0	0	0	0	1	0	0	1	0
TCGA-ND-A4W6-01A	1	0	0	0	1	0	0	0	1	0
TCGA-NG-A4VU-01A	0	0	1	0	0	1	0	0	1	0
TCGA-N8-A4PL-01A	0	0	0	0	0	1	0	0	1	0
TCGA-N5-A4RT-01A	0	1	1	0	1	0	0	1	1	0
TCGA-N9-A4Q3-01A	0	0	0	1	0	1	0	0	0	0
TCGA-N5-A4RN-01A	0	0	1	0	0	0	0	0	1	0

Showing 1 to 26 of 57 entries, 10 total columns

Figura 4 - Tabela de contagem de mutação para o top 10 genes de UCS. Fonte: De autoria própria.

Resultados

TCGA_UCS_RNASEQ_MATCHED	TCGA_UCS_MAF_MATCHED	TCGA_UCS_MAF	TCGA_UCS_MAF_filtered	script_final_ucec.R	combined_counts	gene_name
TCGA-BG-A3EW-11A	TCGA-EY-A1GI-01A	TCGA-BG-A0VZ-01A	TCGA-AX-A1CF-01A	TCGA-B5-A0K1-01A	TCGA-AX-A062-01A	TCGA-DF-A2KY-01A
ENSG000000000003.15	1461	2816	3650	1879	2430	1323
ENSG000000000005.6	117	0	4	0	1	4
ENSG0000000000419.13	1475	963	586	746	418	353
ENSG0000000000457.14	429	422	230	174	326	202
ENSG0000000000460.17	106	361	123	99	155	54
ENSG0000000000938.13	724	111	146	108	93	23
ENSG0000000000971.16	3596	1012	533	901	95	458
ENSG000000001036.14	1750	1959	1787	1395	1019	744
ENSG000000001084.13	1057	3436	304	1138	458	303
ENSG000000001167.14	766	936	644	1329	774	350
ENSG000000001460.18	366	343	445	287	769	80
ENSG000000001461.17	981	863	991	809	692	300
ENSG000000001497.18	2301	1082	1468	1631	1306	454
ENSG000000001561.7	647	991	1040	1872	797	305
ENSG000000001617.12	1026	692	2698	1694	1259	536
ENSG000000001626.16	64	2354	40	70	1174	304
ENSG000000001629.10	1273	2021	1015	2063	1084	719
ENSG000000001630.17	27	77	42	242	46	11
ENSG000000001631.16	341	208	177	137	135	118

Showing 1 to 19 of 60,660 entries, 646 total columns

TCGA_UCS_MAF_MATCHED	TCGA_UCS_MAF
Cancer_Type	
TCGA-BG-A3EW-11A_1	TCGA_UCEC
TCGA-EY-A1GI-01A_2	TCGA_UCEC
TCGA-BG-A0VZ-01A_3	TCGA_UCEC
TCGA-AX-A1CF-01A_4	TCGA_UCEC
TCGA-B5-A0K1-01A_5	TCGA_UCEC
TCGA-AX-A062-01A_6	TCGA_UCEC
TCGA-DF-A2KY-01A_7	TCGA_UCEC
TCGA-EO-A2CG-01A_8	TCGA_UCEC
TCGA-AX-A3FS-01A_9	TCGA_UCEC
TCGA-A5-A0G3-01A_10	TCGA_UCEC
TCGA-EO-A1Y7-01A_11	TCGA_UCEC
TCGA-D1-A3DA-01A_12	TCGA_UCEC
TCGA-EY-A1GE-01A_13	TCGA_UCEC
TCGA-BK-A4ZD-01A_14	TCGA_UCEC
TCGA-FL-A1YM-11A_15	TCGA_UCEC
TCGA-A5-A0R6-01A_16	TCGA_UCEC
TCGA-BG-A0VT-01A_17	TCGA_UCEC
TCGA-AW-A1PO-01A_18	TCGA_UCEC
TCGA-AP-A0LI-01A_19	TCGA_UCEC

Showing 1 to 20 of 646 entries, 1 total columns

Figura 5 - Tabela de combinação de reads de UCS e UCEC e relação entre barcode e tipo de câncer. Fonte: De autoria própria.

Resultados

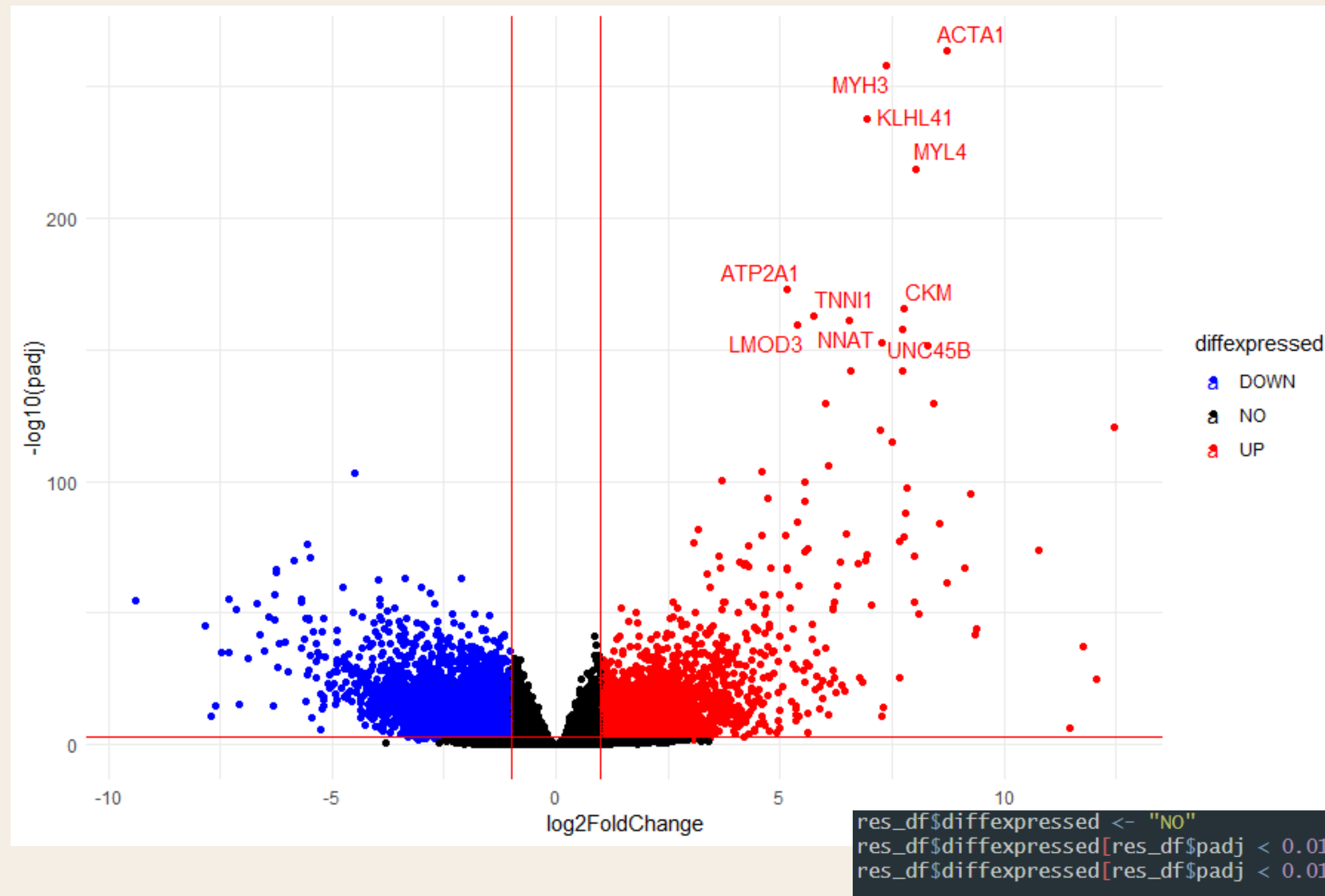


Figura 6 - Distribuição de Expressão Gênica Diferencial entre UCEC e UCS. Fonte: De autoria própria.

Resultados

```
> print(head(arrange(res_df, padj), 10))
```

	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj	diffexpressed
ENSG00000143632.14	891.19829	8.710737	0.2487342	35.02026	1.106326e-268	5.032344e-264	UP
ENSG00000109063.15	1501.20723	7.377359	0.2130192	34.63237	8.230560e-263	1.871917e-258	UP
ENSG00000239474.7	191.29560	6.949017	0.2090438	33.24192	2.670765e-242	4.049503e-238	UP
ENSG00000198336.9	281.54696	8.037547	0.2518801	31.91021	1.927060e-223	2.191404e-219	UP
ENSG00000196296.14	312.15835	5.166466	0.1817442	28.42713	9.345044e-178	8.501561e-174	UP
ENSG00000104879.5	646.60076	7.775407	0.2795306	27.81595	2.782443e-170	2.109416e-166	UP
ENSG00000159173.19	930.09361	5.735165	0.2079750	27.57622	2.146393e-167	1.394757e-163	UP
ENSG00000053438.11	274.02351	6.526441	0.2378162	27.44322	8.370805e-166	4.759535e-162	UP
ENSG00000163380.16	71.41501	5.377145	0.1971675	27.27196	9.126030e-164	4.612397e-160	UP
ENSG00000141161.12	123.95387	7.736444	0.2854053	27.10686	8.172690e-162	3.717511e-158	UP

Figura 7 - Tabela de genes mais significativos da expressão diferencial. Fonte: De autoria própria.

Resultados

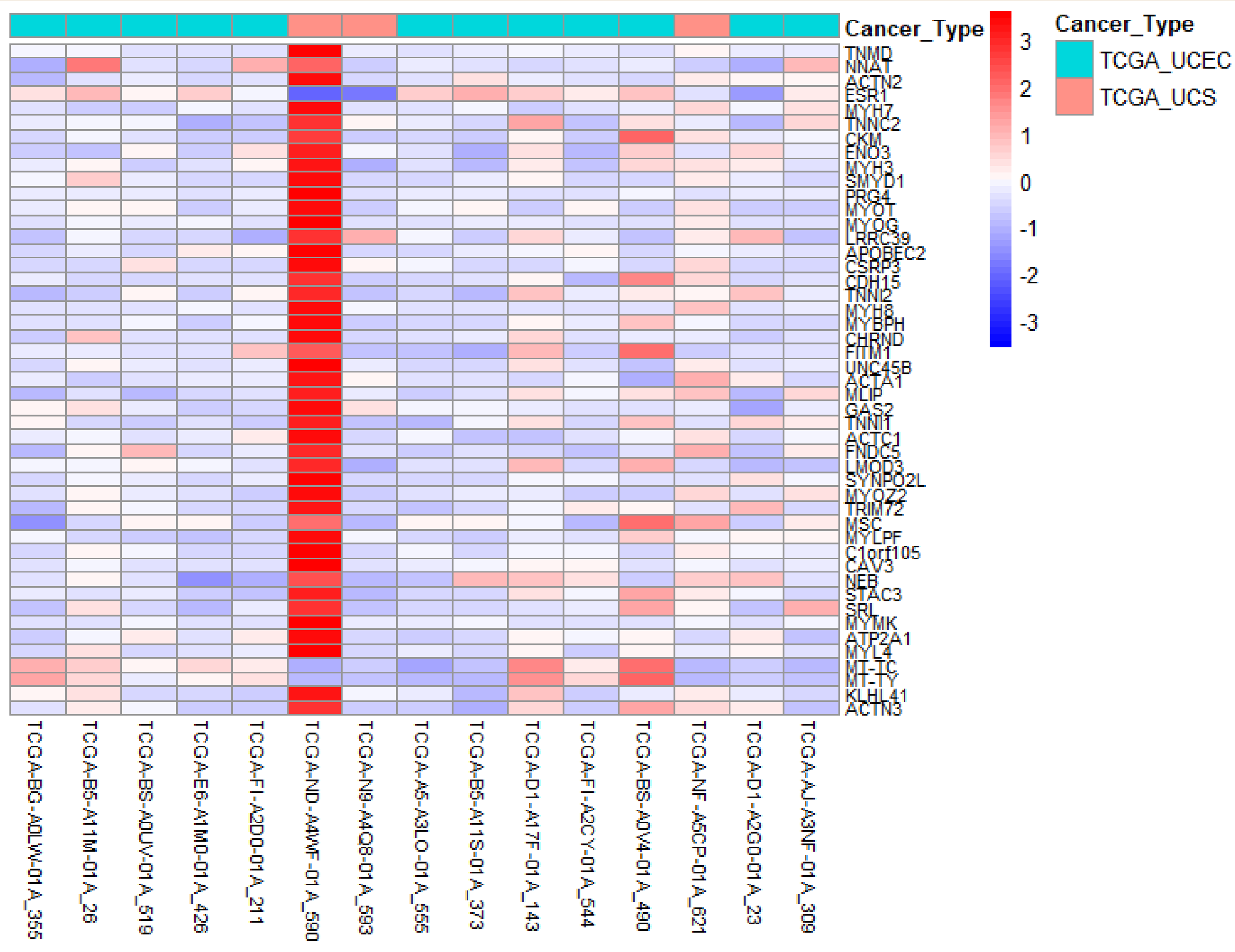


Figura 8 - Tabela de genes mais significativos da expressão diferencial. Fonte: De autoria própria.

Resultados

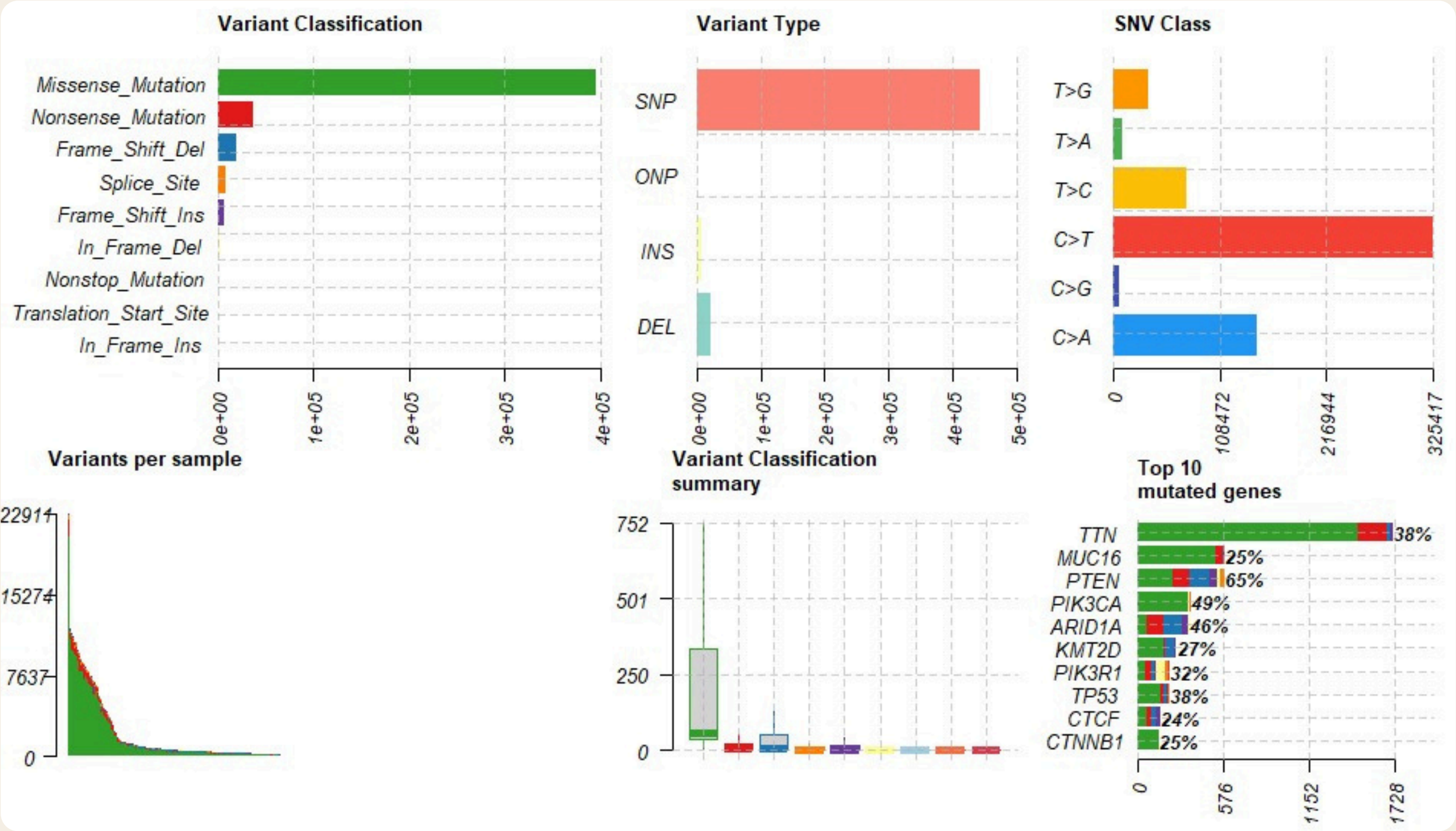


Figura 9 - Perfil genômico do UCEC (Uterine Corpus Endometrial Carcinoma). Fonte: De autoria própria.

Resultados

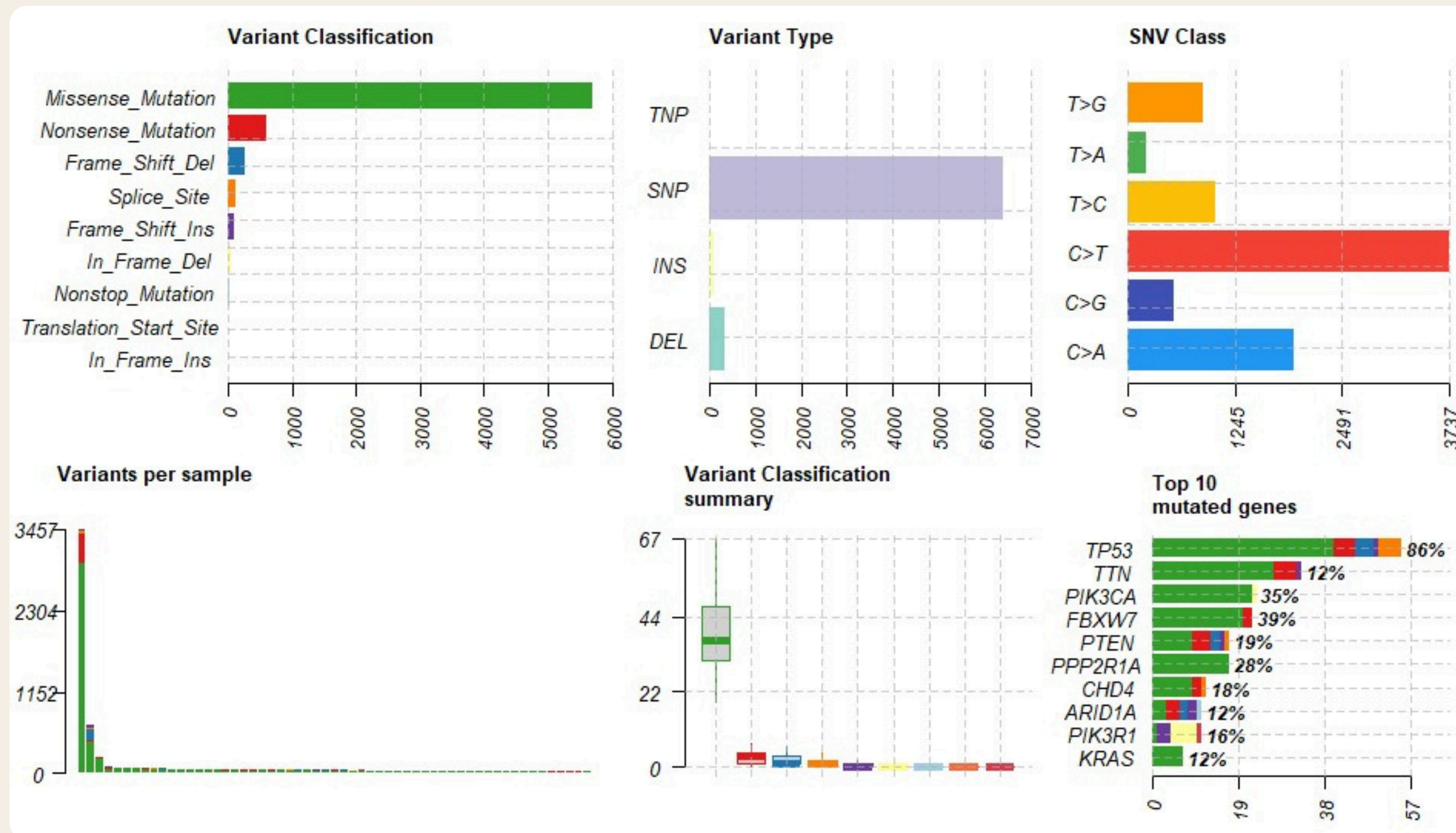


Figura 10 - Perfil genômico do UCS (Uterine Carcinosarcoma). Fonte: De autoria própria.

Resultados

```
compare <- mafCompare(m1 = merged_maf_UCEC, m2 = merged_maf_UCS, m1Name = "UCEC", m2Name = "UCS")
print(compare)
results
```

	Hugo_Symbol	UCEC	UCS	pval	or	ci.up	ci.low	adjPval
	<char>	<int>	<num>	<num>	<num>	<num>	<num>	<num>
1:	TP53	195	49	1.273316e-12	0.09891719	0.2164101	0.03958852	2.123255e-08
2:	PTEN	335	11	3.922901e-11	7.62794829	16.7492352	3.77775768	3.270719e-07
3:	CTNNB1	129	0	3.417613e-07	Inf	Inf	4.85853301	1.623048e-03
4:	ARID1A	236	7	3.893368e-07	5.96278339	15.8848037	2.62441730	1.623048e-03
5:	ZFHX4	105	0	7.075035e-06	Inf	Inf	3.71532000	2.359524e-02

6671:	TAF7L	26	2	1.000000e+00	1.45242678	12.9613353	0.34747268	1.000000e+00
6672:	TM9SF3	26	2	1.000000e+00	1.45242678	12.9613353	0.34747268	1.000000e+00
6673:	TSEN2	26	2	1.000000e+00	1.45242678	12.9613353	0.34747268	1.000000e+00
6674:	VPS35L	26	2	1.000000e+00	1.45242678	12.9613353	0.34747268	1.000000e+00
6675:	ZNF781	26	2	1.000000e+00	1.45242678	12.9613353	0.34747268	1.000000e+00

```
SampleSummary
Cohort SampleSize
<char>    <num>
:   UCEC      518
:   UCS       57
```

Figura 11 - Resultados da análise feita pela maf compare. Fonte: De autoria própria.

Resultados

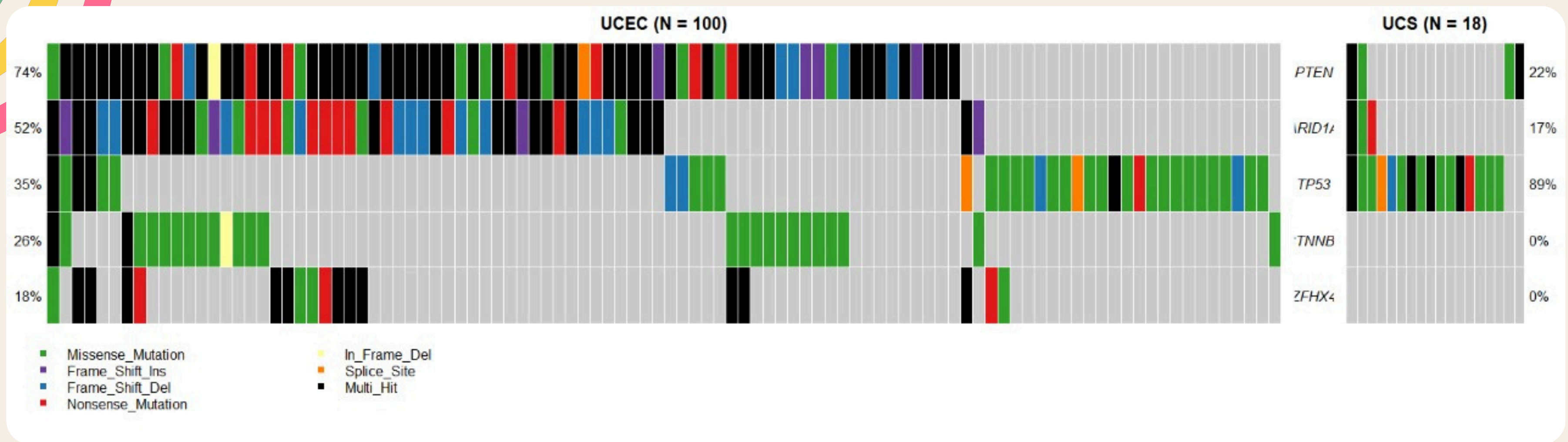


Figura 12 - Distribuição de mutações gênicas mais frequentes em UCEC e UCS, com os tipos de mutações representados por diferentes cores. Fonte: De autoria própria.

Resultados

```
> somaticInteractions(maf = merged_maf_UCS, top = 25, pvalue = c(0.05, 0.1))
```

	gene1 <char>	gene2 <char>	pValue <num>	oddsRatio <num>	00 <int>	01 <int>	11 <int>	10 <int>	pAdj <num>	Event <char>	pair <char>
1:	FER1L6	LRP2	0.0001841367	75.88721	50	2	4	1	0.004262424	Co_Occurence	FER1L6, LRP2
2:	FLNA	PTEN	0.0005983755	33.42179	45	6	5	1	0.012896023	Co_Occurence	FLNA, PTEN
3:	SYNE1	ARID1A	0.0012106397	27.63226	48	3	4	2	0.024408059	Co_Occurence	ARID1A, SYNE1
4:	RYR2	MUC16	0.0013290803	58.75027	51	2	3	1	0.025171975	Co_Occurence	MUC16, RYR2
5:	ARID1A	PTEN	0.0018902374	16.78657	44	6	5	2	0.033754240	Co_Occurence	ARID1A, PTEN

296:	ZBTB7B	MGAM	1.0000000000	0.00000	45	6	0	6	1.0000000000	Mutually_Exclusive	MGAM, ZBTB7B
297:	ZBTB7B	SYNE1	1.0000000000	0.00000	45	6	0	6	1.0000000000	Mutually_Exclusive	SYNE1, ZBTB7B
298:	DOP1A	ZBTB7B	1.0000000000	0.00000	46	6	0	5	1.0000000000	Mutually_Exclusive	DOP1A, ZBTB7B
299:	FER1L6	ZBTB7B	1.0000000000	0.00000	46	6	0	5	1.0000000000	Mutually_Exclusive	FER1L6, ZBTB7B
300:	VPS13B	ZBTB7B	1.0000000000	0.00000	46	6	0	5	1.0000000000	Mutually_Exclusive	VPS13B, ZBTB7B

	event_ratio <char>
1:	4/3
2:	5/7
3:	4/5
4:	3/3
5:	5/8

296:	0/12
297:	0/12
298:	0/11
299:	0/11
300:	0/11

Figura 13 - Interações somáticas para os genes UCS. Fonte: De autoria própria.

Resultados

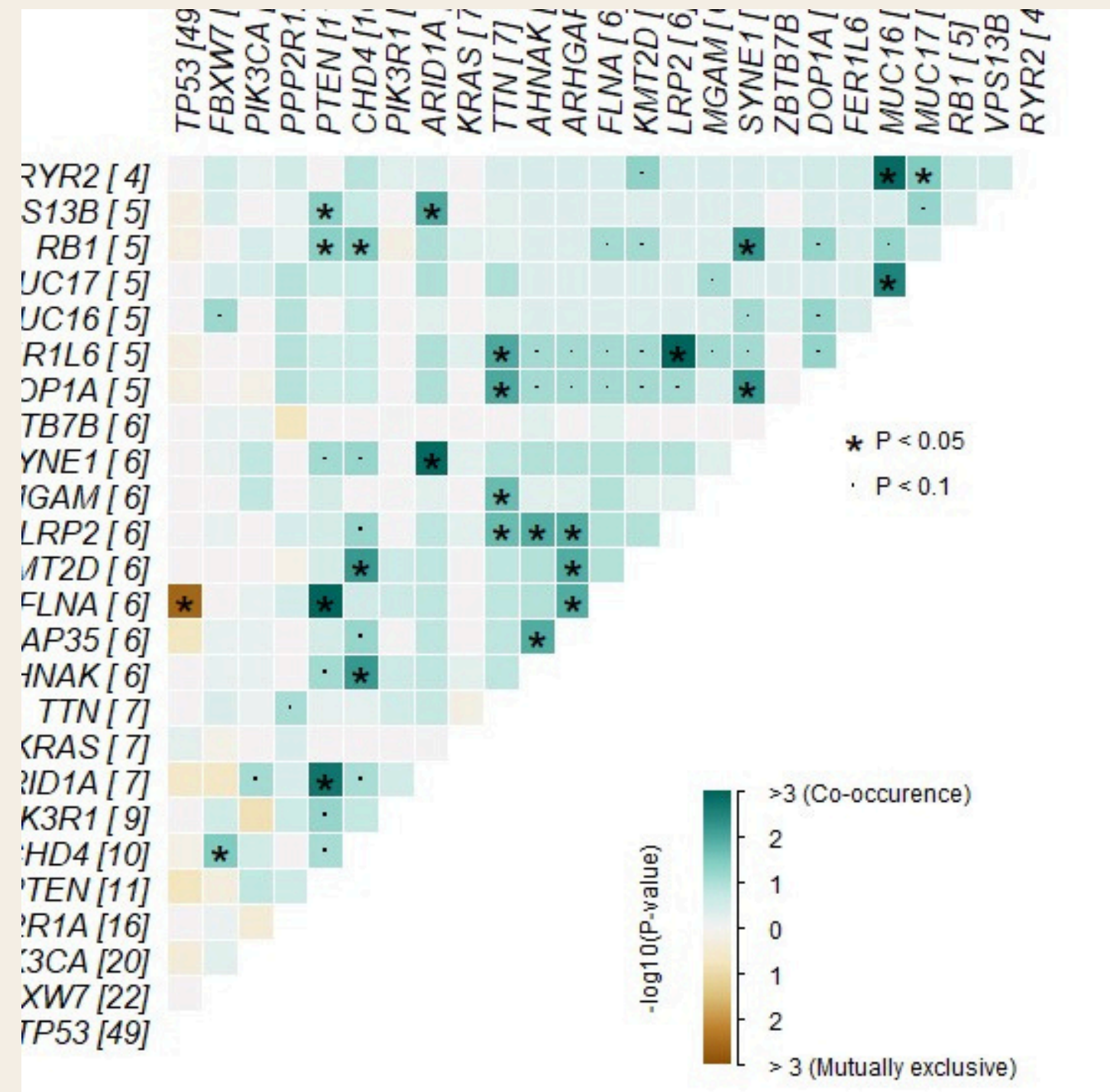


Figura 14 - Interações somáticas para os genes UCS. Fonte: De autoria própria.

Resultados

```
> somaticInteractions(maf = merged_maf_UCEC, top = 25, pvalue = c(0.05, 0.1))
Processed 3522 groups out of 3522. 100% done. Time elapsed: 5s. ETA: 0s.
```

	gene1 <char>	gene2 <char>	pValue <num>	oddsRatio <num>	00 <int>	01 <int>	11 <int>	10 <int>	pAdj <num>	Event <char>	pair <char>
1:	PTEN	TP53	1.602913e-36	0.07811274	48	135	60	275	3.710447e-35	Mutually_Exclusive	PTEN, TP53
2:	PCLO	SYNE1	5.683050e-33	21.31437443	382	36	67	33	1.224795e-31	Co_Occurence	PCLO, SYNE1
3:	PCLO	TTN	2.506901e-31	21.16690575	311	107	88	12	5.054236e-30	Co_Occurence	PCLO, TTN
4:	FAT4	PCLO	8.858849e-31	18.93107179	381	35	65	37	1.677812e-29	Co_Occurence	FAT4, PCLO
5:	PCLO	DST	3.160847e-30	17.54927384	375	43	67	33	5.644369e-29	Co_Occurence	DST, PCLO

296:	FAT1	TP53	4.317562e-01	0.81864167	253	159	36	70	4.373543e-01	Mutually_Exclusive	FAT1, TP53
297:	TP53	MUC16	4.612691e-01	0.84285199	240	83	44	151	4.657402e-01	Mutually_Exclusive	MUC16, TP53
298:	FAT4	TP53	6.486918e-01	0.88187455	257	159	36	66	6.528702e-01	Mutually_Exclusive	FAT4, TP53
299:	CHD4	TP53	9.099314e-01	1.04424123	259	155	40	64	9.128526e-01	Co_Occurence	CHD4, TP53
300:	KMT2B	CTNNB1	1.000000e+00	1.02172083	303	100	29	86	1.000000e+00	Co_Occurence	CTNNB1, KMT2B

	event_ratio <char>
1:	60/410
2:	67/69
3:	88/119
4:	65/72
5:	67/76

296:	36/229
297:	44/234
298:	36/225
299:	40/219
300:	29/186

Figura 15 - Interações somáticas para os genes UCEC. Fonte: De autoria própria.

Resultados

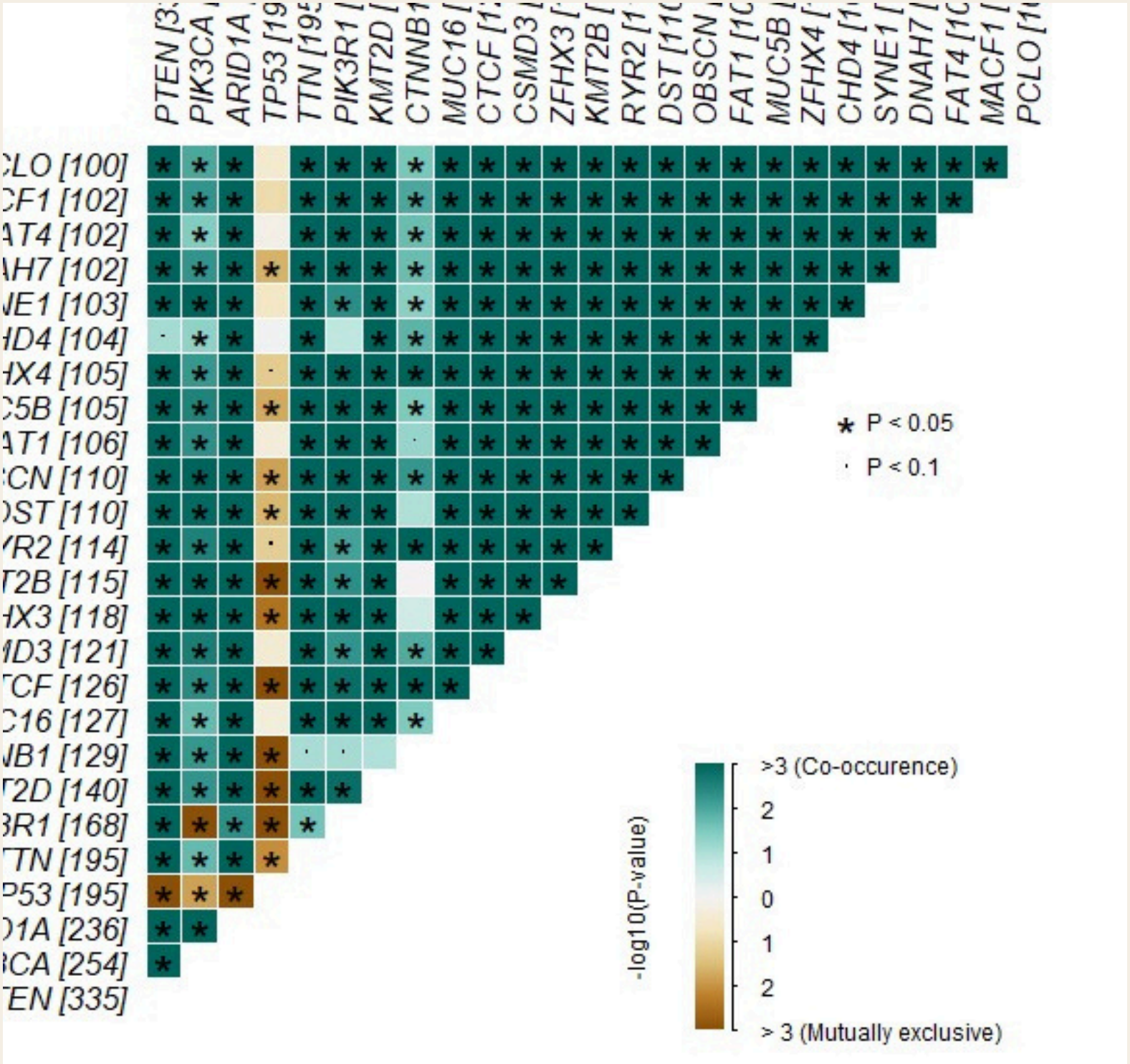


Figura 16 - Interações somáticas para os genes UCEC. Fonte: De autoria própria.

Resultados

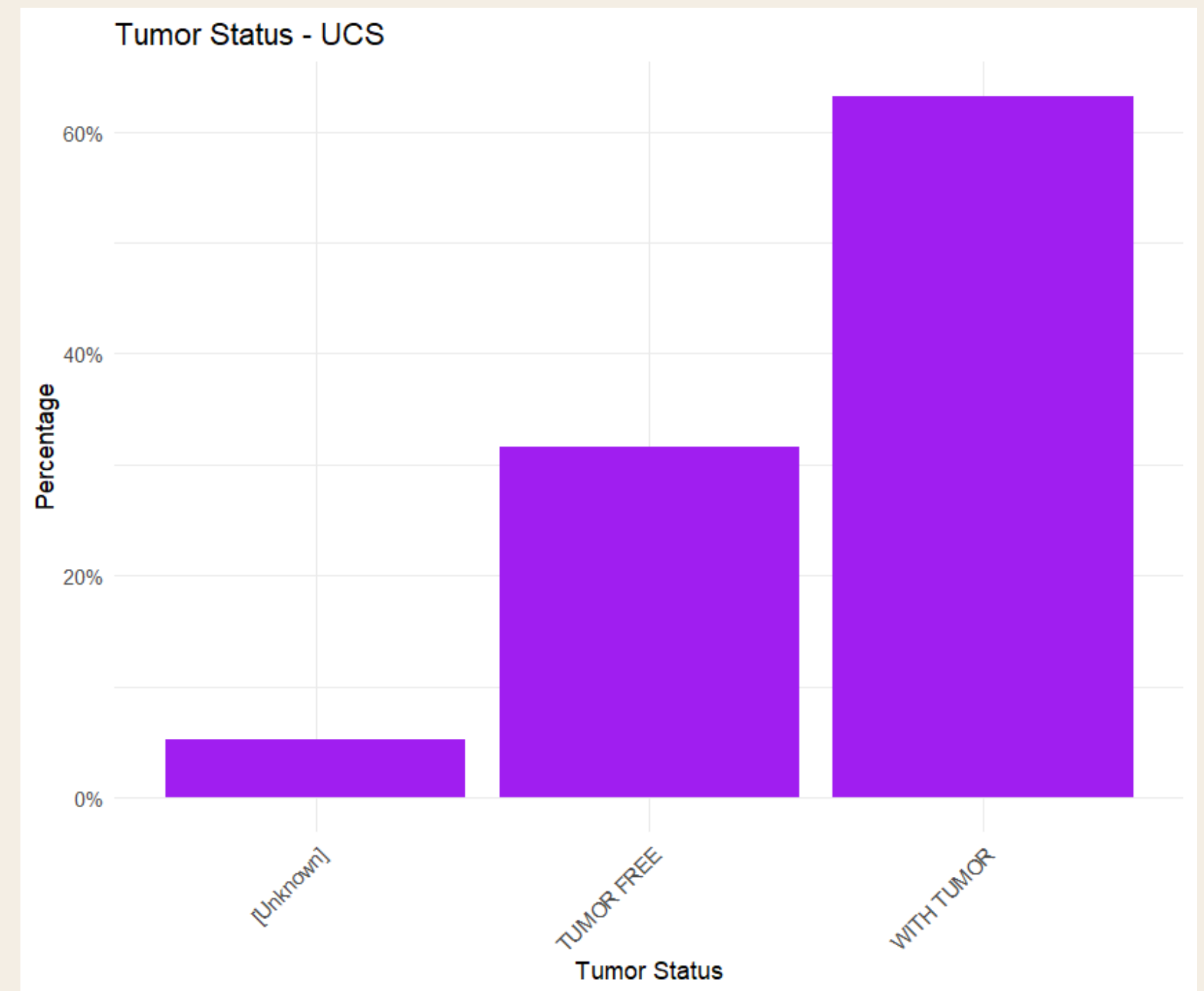
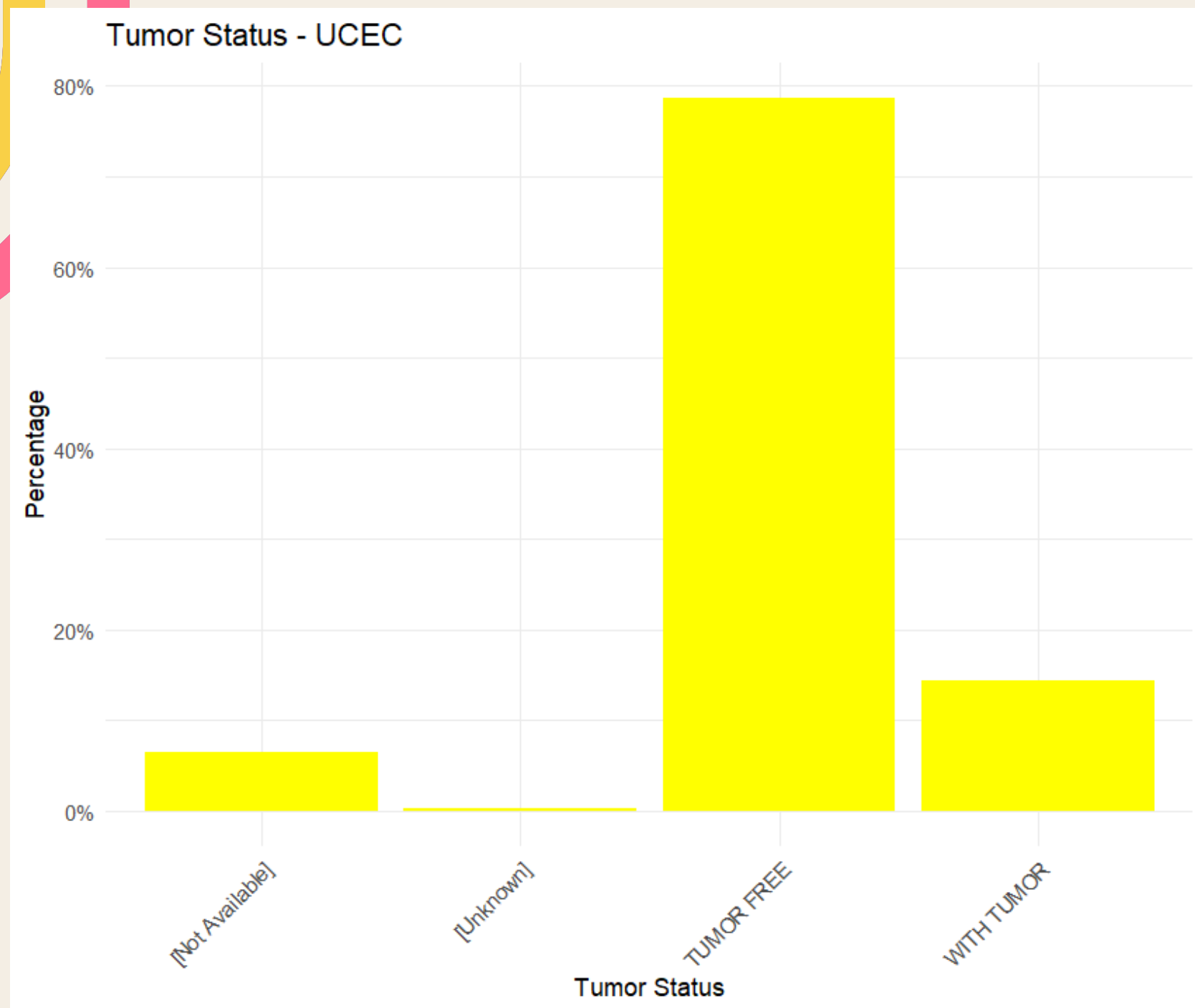


Figura 17 - Valores da Tumor Status para UCEC. Fonte: De autoria própria.

Figura 18 - Valores da Tumor Status para UCS. Fonte: De autoria própria.

Resultados

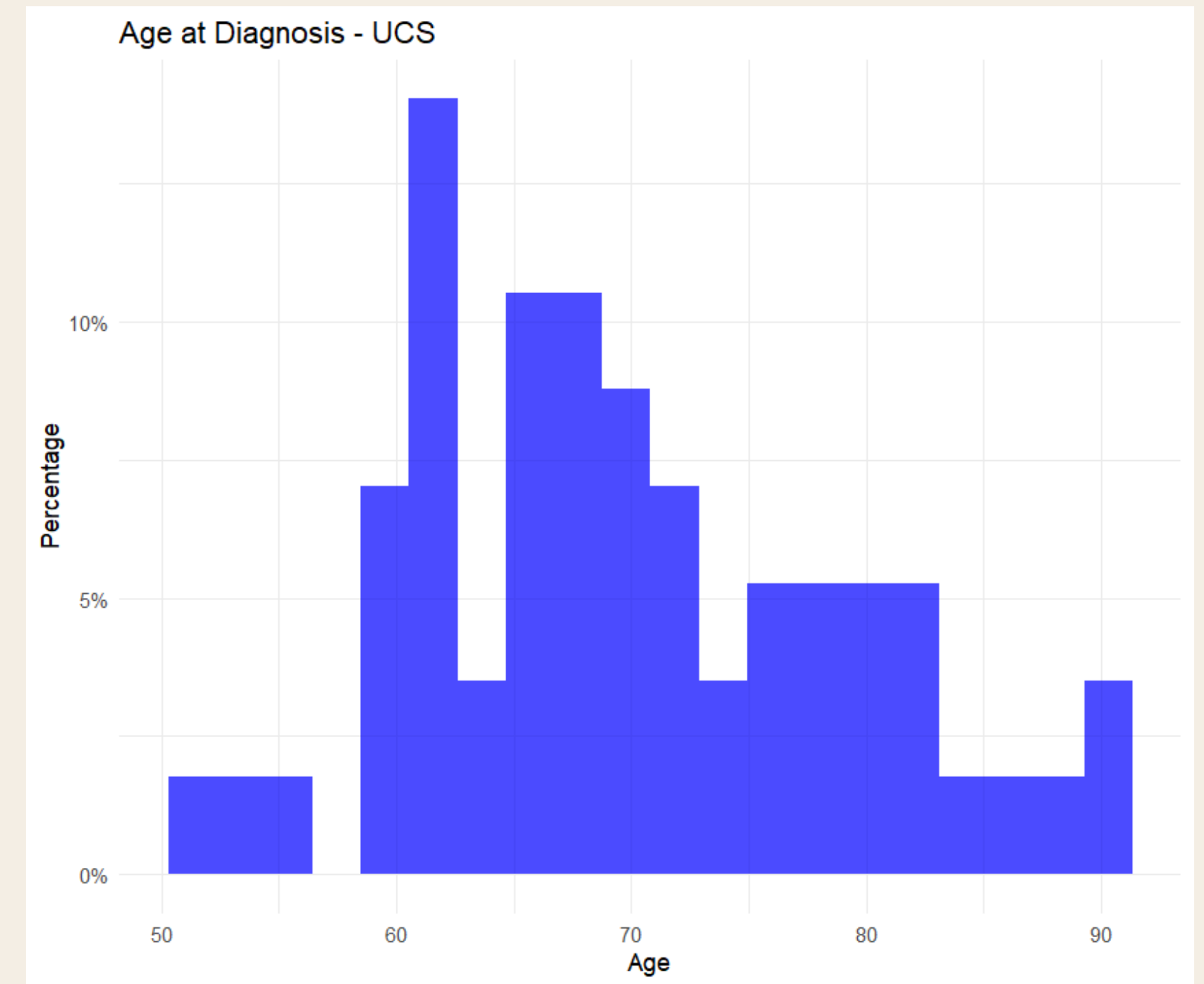
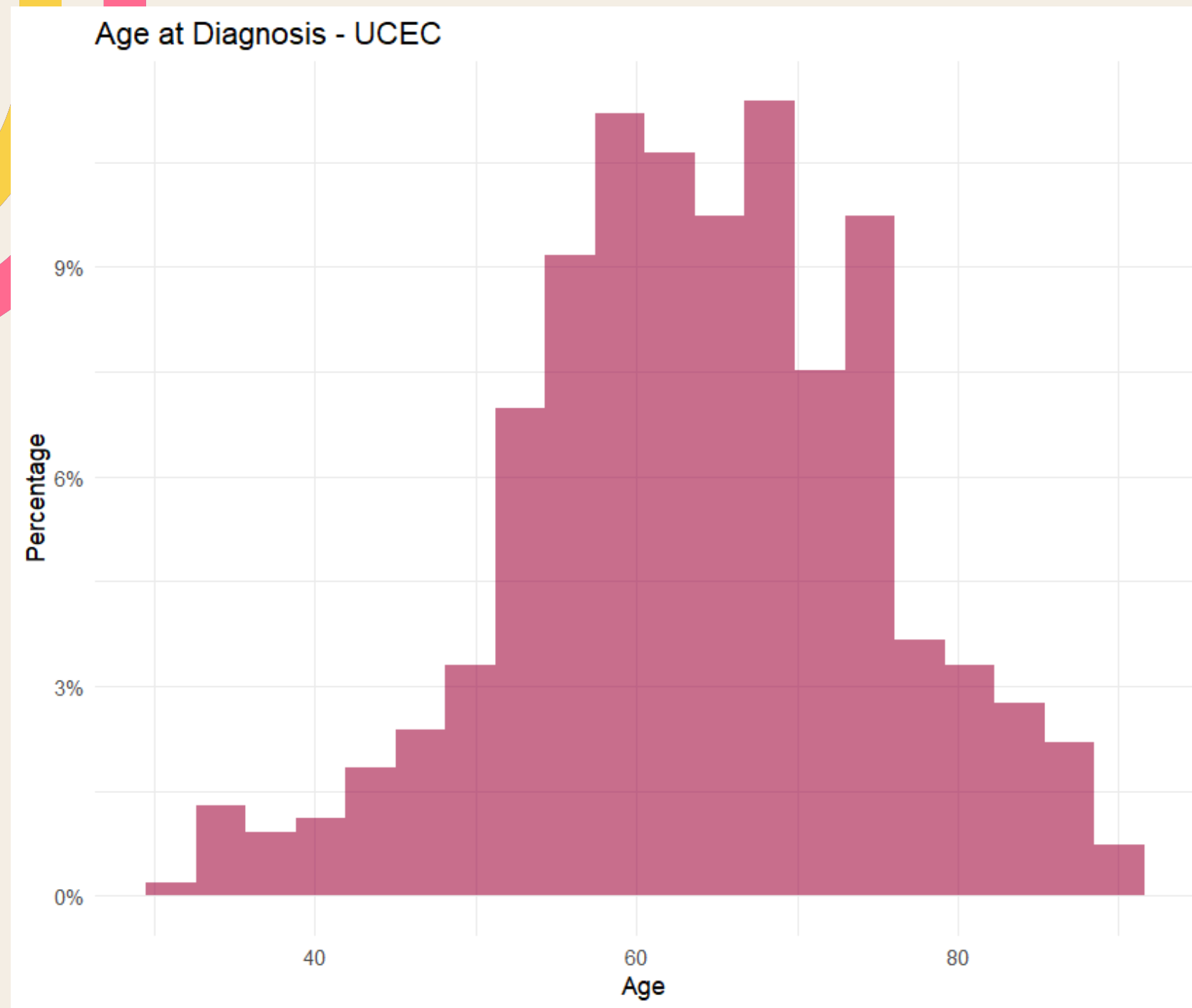


Figura 19 - Valores da Age at Diag. para UCS. Fonte: De autoria própria.

Figura 20 - Valores da Age at Diag. para UCEC. Fonte: De autoria própria.

Resultados

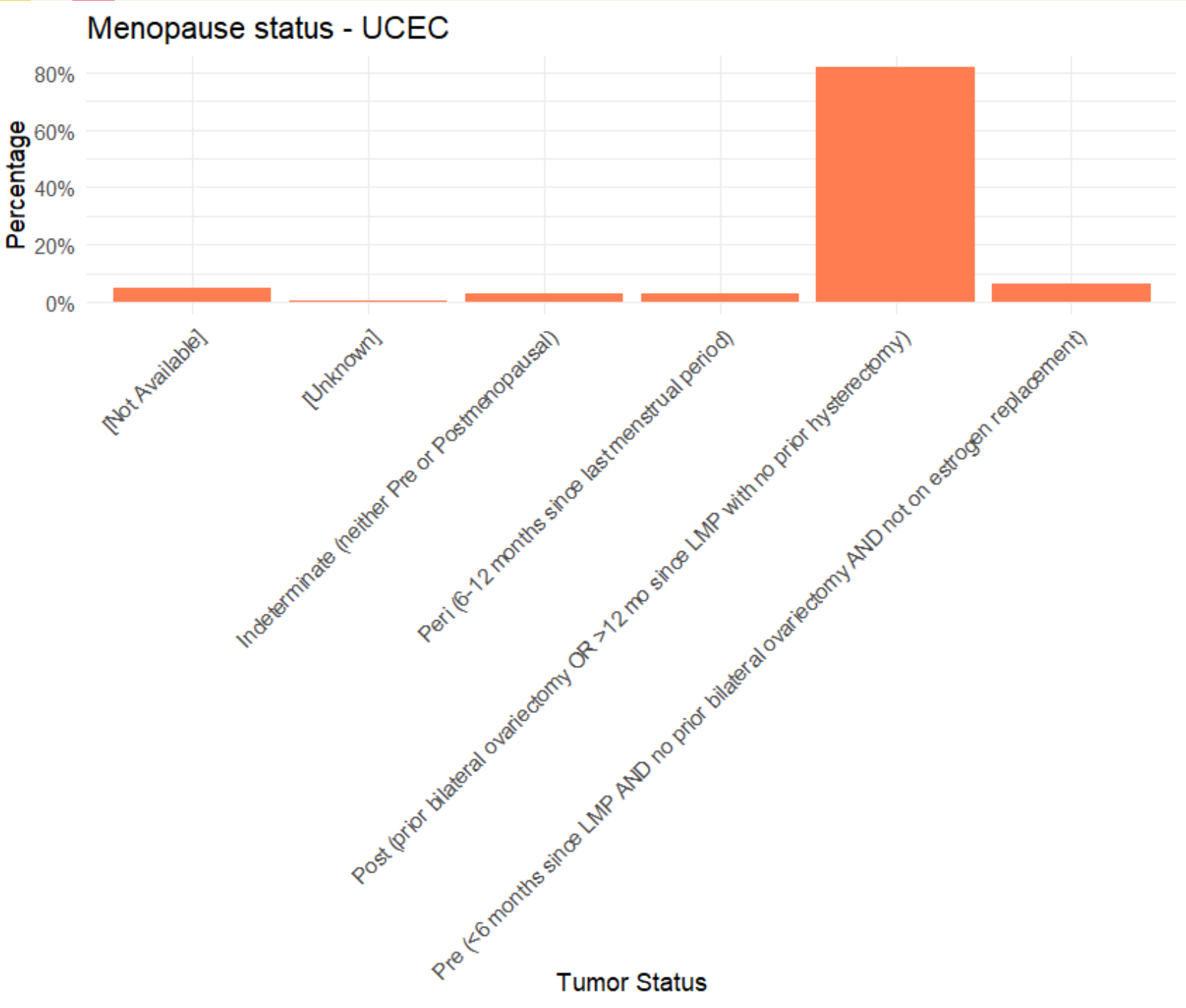


Figura 21 - Valores da Menopause Status para UCS. Fonte: De autoria própria.

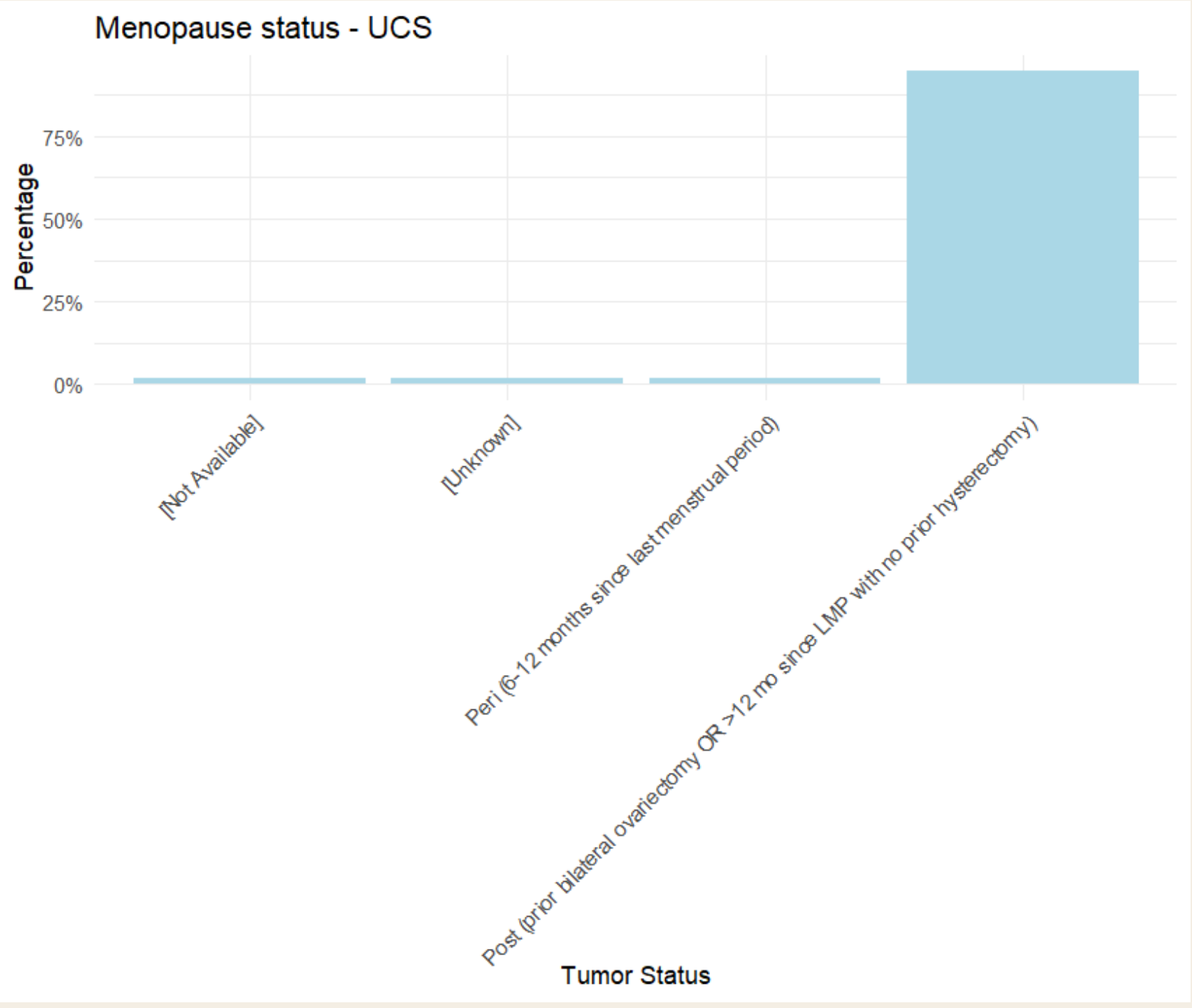


Figura 22 - Valores da Menopause Status para UCEC. Fonte: De autoria própria.

Resultados

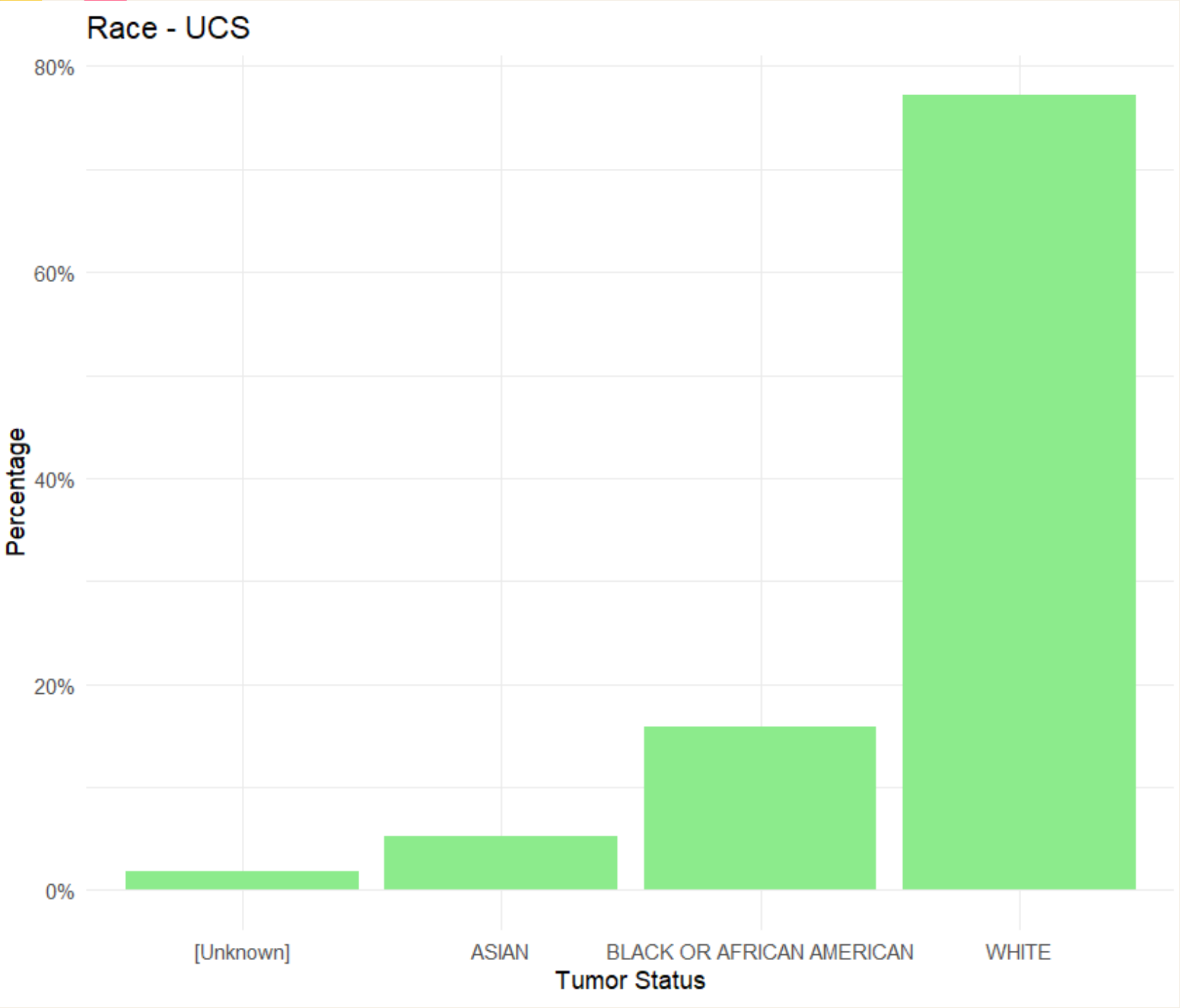


Figura 23 - Valores da Race para UCS. Fonte: De autoria própria.

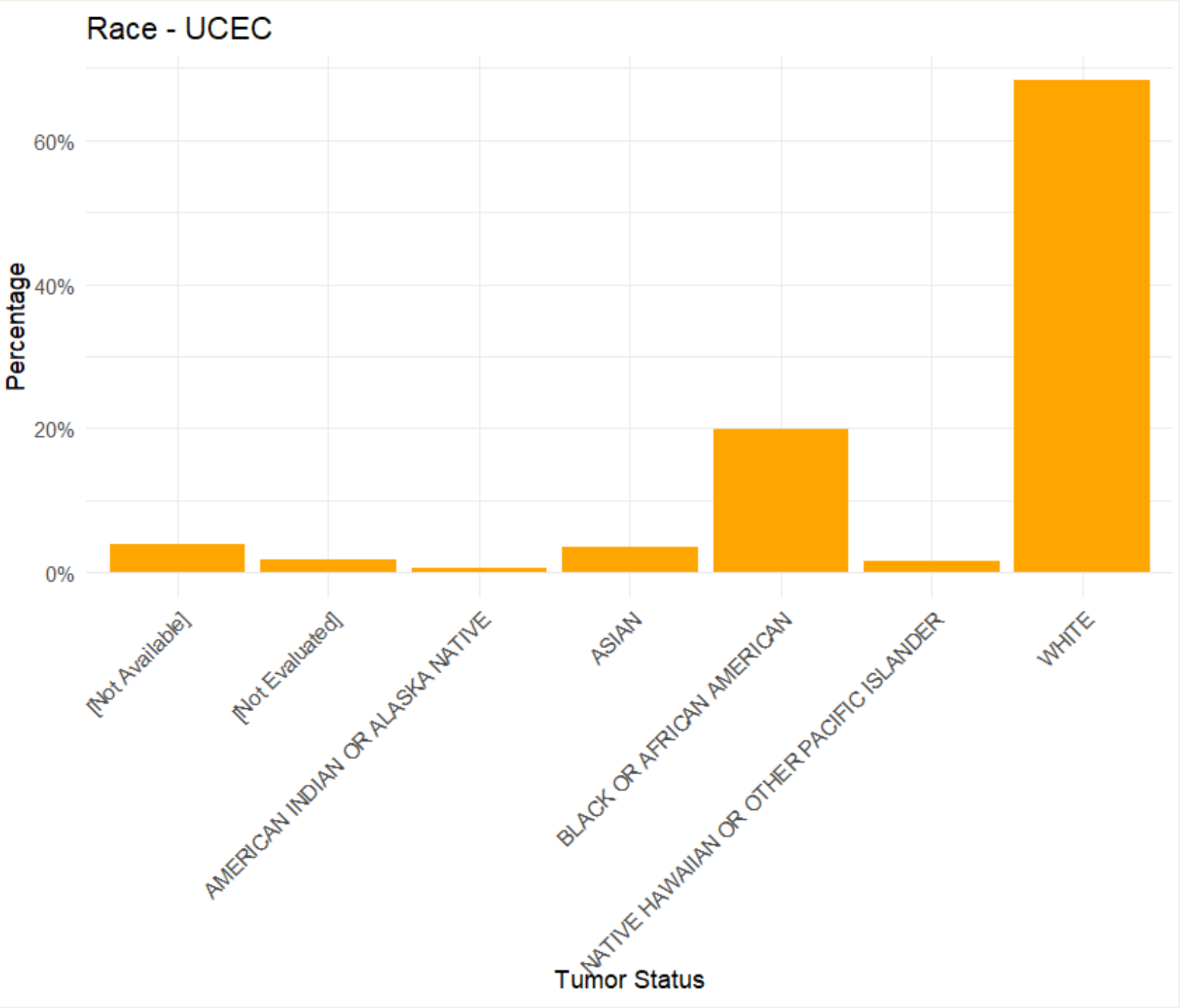


Figura 24 - Valores da Race para UCEC. Fonte: De autoria própria.

Resultados

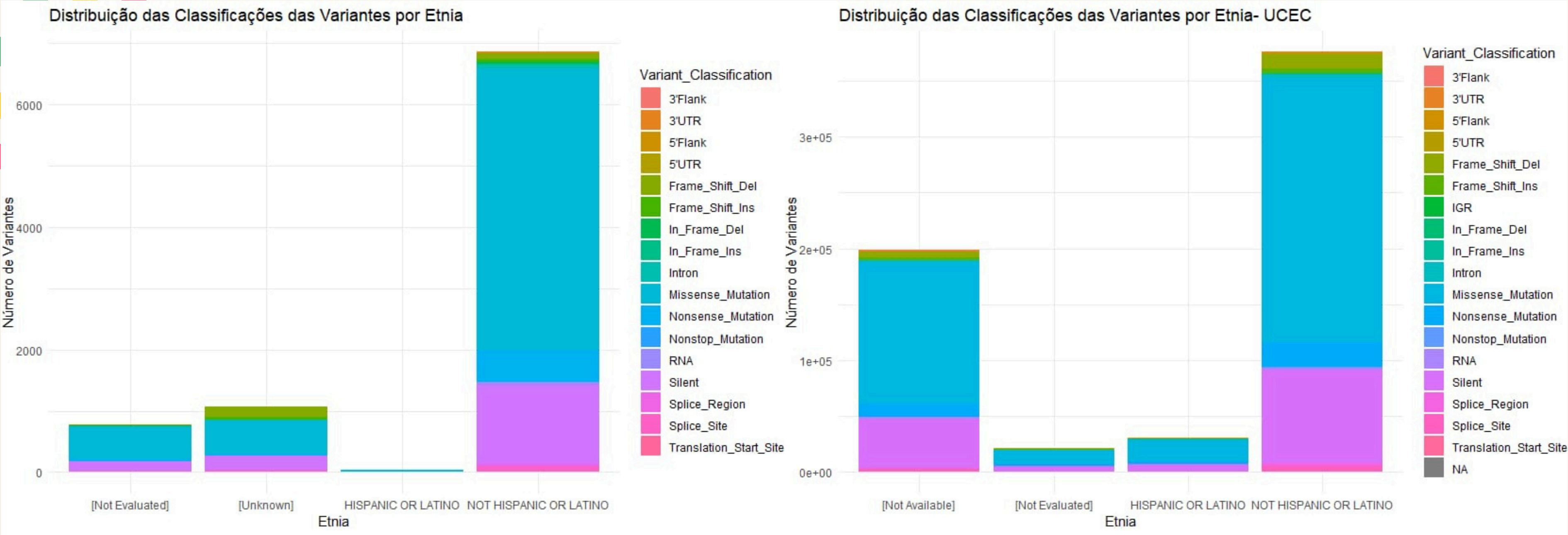


Figura 25 - Comparação da Distribuição das Classificações das Variantes por Etnia para UCS e UCEC.
Fonte: De autoria própria.

Resultados

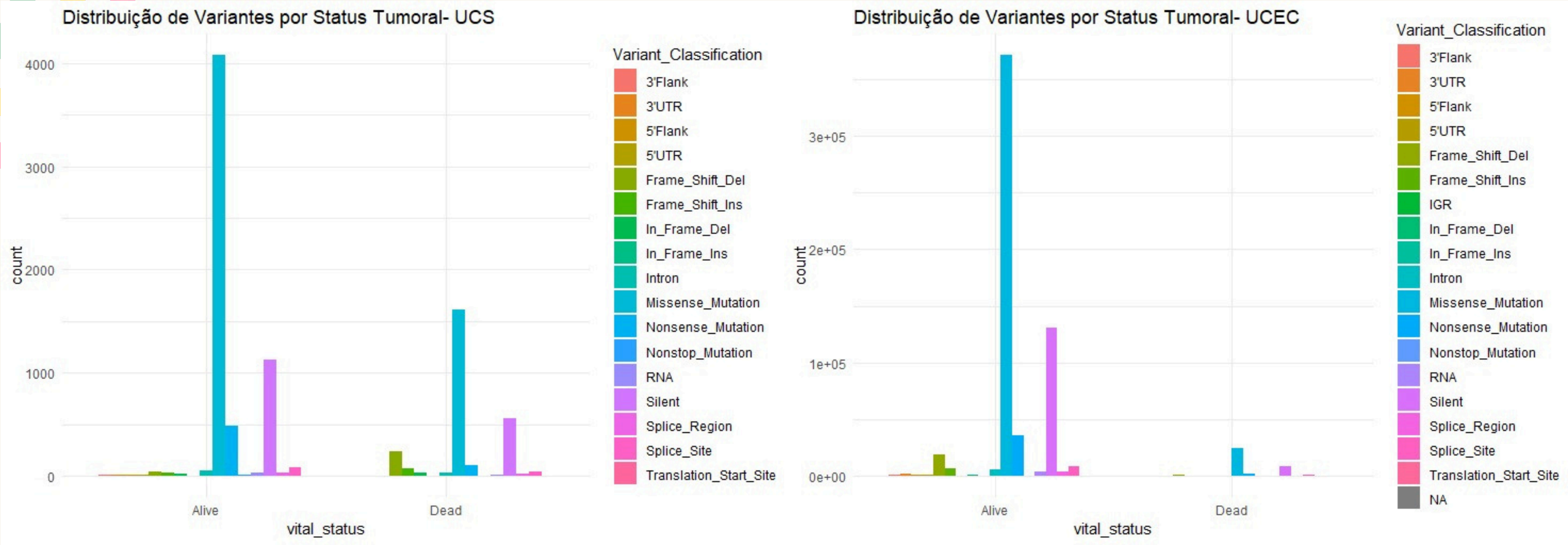


Figura 26 - Comparação da Distribuição das Classificações das Variantes por Status Tumoral para UCS e UCEC. Fonte: De autoria própria.

Conclusão

Diferenças significativas nos perfis moleculares e clínicos dos dois tipos de câncer.

UCEC

- Maior diversidade mutacional;
- Carga mutacional variada entre as amostras;
- Maior complexidade genética;
- Múltiplos caminhos de evolução tumoral;
- PTEN e PIK3CA;
- Terapias com abordagens mais amplas.

UCS

- Menor diversidade mutacional;
- Predominância de mutações no gene TP53 (86%);
- Terapias focadas e mais eficazes.



Referências Bibliográficas

1. AKBANI, R.; LEVINE, D. A.; THE CANCER GENOME ATLAS RESEARCH NETWORK. Abstract 133: Integrated molecular characterization of uterine carcinosarcoma in The Cancer Genome Atlas (TCGA) project. Molecular and Cellular Biology, Genetics. Anais...American Association for Cancer Research, 2016.
2. CANCER GENOME ATLAS RESEARCH NETWORK et al. Integrated genomic characterization of endometrial carcinoma. Nature, v. 497, n. 7447, p. 67–73, 2013.
3. HENLEY, S. J.; MILLER, J. W.; DOWLING, N. F.; BENARD, V. B.; RICHARDSON, L. C. Uterine Cancer Incidence and Mortality - United States, 1999-2016. MMWR Morb Mortal Wkly Rep. 2018 Dec 7;67(48):1333-1338. doi: 10.15585/mmwr.mm6748a1. PMID: 30521505; PMCID: PMC6329484.
4. HICKMAN, A. R. et al. Identification of condition-specific biomarker systems in uterine cancer. G3 (Bethesda, Md.), v. 12, n. 1, 2022.
5. LE GALLO, M. et al. The FOXA2 transcription factor is frequently somatically mutated in uterine carcinosarcomas and carcinomas. Cancer, v. 124, n. 1, p. 65–73, 2018.

Referências Bibliográficas

6. OKUDA, T. et al. Genetics of endometrial cancers. *Obstetrics and gynecology international*, v. 2010, n. 1, p. 984013, 2010.
7. PANE, K. et al. An integrative computational approach based on expression similarity signatures to identify protein-protein interaction networks in female-specific cancers. *Frontiers in genetics*, v. 11, p. 612521, 2020.
8. SILVA, T. C. et al. TCGA Workflow: Analyze cancer genomics and epigenomics data using Bioconductor packages. *F1000Research*, v. 5, p. 1542, 2016.
9. WANG, X. et al. Characteristics of The Cancer Genome Atlas cases relative to U.S. general population cancer cases. *British journal of cancer*, v. 119, n. 7, p. 885–892, 2018.



Obrigado!

beatriz.fioravanti@usp.br
beatrizov@usp.br
heloisaferian@usp.br
mariafmaia@usp.br
vinicius-souza@usp.br