**RIB 0109 - Fundamentos de Informática em Bioinformática e Medicina Genômica**

**Proposta de Projeto de análise de dados genômicos/moleculares**

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

**Proponentes:**

Ana Beatriz Monteiro Fioravanti - 14556271

Beatriz de Oliveira Vieira - 14685352

Heloisa Ferian Soares - 14577959

Maria Fernanda Maia de Oliveira - 14595546

Vinícius de Souza Oliveira - 14761322

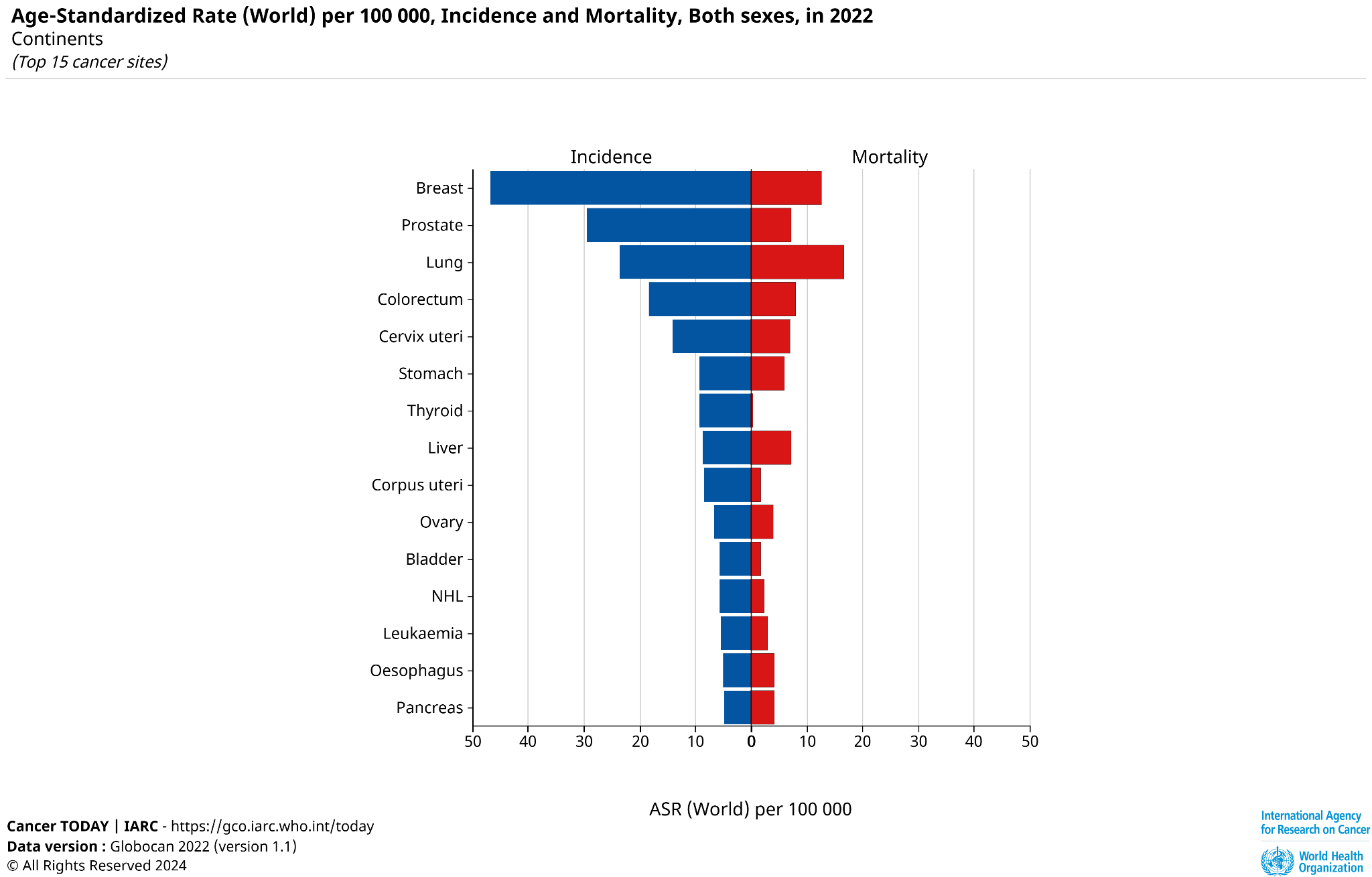
**Sumário**

1. **Introdução**
   * Contextualização sobre a relevância do câncer uterino em termos de incidência e mortalidade nos EUA.
   * Diferenciação entre os subtipos de câncer uterino: carcinoma endometrial (UCEC) e carcinossarcoma uterino (UCS).
   * Dados epidemiológicos e características gerais.
2. **Justificativa**
   * Importância de entender diferenças genômicas entre UCEC e UCS.
   * Impacto na definição de estratégias terapêuticas e na compreensão dos mecanismos biológicos.
3. **Objetivos**
   * **Principal**: Analisar comparativamente os subtipos UCEC e UCS para identificar diferenças clínicas e moleculares.
   * **Complementares**:
     + Familiarizar-se com a plataforma TCGA.
     + Analisar dados de mutações, expressão gênica e informações clínicas.
     + Obter estatísticas e observar padrões de prevalência de mutações.
4. **Metodologia**
   * **Descrição do Estudo**: Abordagem de bioinformática com dados do TCGA.
   * **Banco de Dados**:
     + Amostras de UCEC (>500 casos) e UCS (57 casos).
     + Dados de RNA-Seq, mutações somáticas (MAF) e dados clínicos.
   * **Pré-processamento**:
     + Normalização de RNA-Seq e filtragem de mutações somáticas.
     + Armazenamento e análise em ambiente HPC.
   * **Análises Realizadas**:
     + Genômica (mutações, carga mutacional).
     + Expressão gênica diferencial.
     + Comparação dos dados clínicos.
     + Visualizações: gráficos de mutações, volcano plots e tabelas estatísticas.
5. **Resultados**
   * **UCEC**:
     + Alta heterogeneidade genética e diversidade mutacional.
     + Genes mais alterados: PTEN, PIK3CA, TTN.
     + Predominância de mutações missense e substituições nucleotídicas C>T.
   * **UCS**:
     + Perfil genômico menos diverso.
     + Foco em mutações no gene TP53 (89%).
     + Menor carga mutacional por amostra.
   * Comparações detalhadas: gráficos, tabelas e volcano plots destacando genes diferencialmente expressos.
6. **Conclusão**
   * Diferenças significativas entre UCEC e UCS nos perfis genômicos e clínicos.
   * UCEC: maior complexidade genética, múltiplos genes afetados.
   * UCS: foco em TP53, potencial para terapias direcionadas.
   * Implicações para tratamentos personalizados e melhoria no prognóstico das pacientes.
7. **Referências Bibliográficas**
   * Listagem de artigos e fontes utilizadas para fundamentação teórica e metodológica.

**1. Introdução**

O câncer uterino possui taxas de incidência e mortalidade crescentes nos Estados Unidos, sendo considerado como o quarto tipo de câncer mais comum entre os diagnósticos e o sétimo mais observado nas causas de óbito entre mulheres da mesma região. De acordo com análises realizadas pelo *Centers for Disease Control and Prevention,* o câncer uterino teve um acréscimo de 0,7% por ano, entre 1999 a 2015, nas taxas de incidência, e aumento de 1,1% nas taxas de mortalidade, sendo mais comum entre mulheres negras (9/100.000) em comparação com outras etnias (4-5/100.00). (1- <https://pmc.ncbi.nlm.nih.gov/articles/PMC6329484/pdf/mm6748a1.pdf>)

Figura 1- Classificação de câncer por incidência/mortalidade de acordo com o GCO.



Fonte: *Global Cancer Observatory*. (<https://gco.iarc.fr/today/en/dataviz/tables?mode=cancer&cancers=24&multiple_cancers=0&sexes=0&group_populations=1&multiple_populations=1>)

A origem do câncer uterino pode ser classificada pela região do útero: pode se originar no miométrio (sarcoma uterino), a camada musculosa intermediária, ou no endométrio (endometrial), definido como o revestimento interno do útero, com frequência de 85% dos casos de câncer uterino (Bokhman JV. 1983. Two pathogenetic types of endometrial carcinoma. Gynecol Oncol. 15:10–17. doi:10.1016/0090–8258(83)901 11-7.). A classificação de câncer endometrial é uma categoria abrangente, catalogada como UCEC (Uterine Corpus Endometrial Carcinoma) no TCGA. Dentro dessa categoria há o Carcinossarcoma uterino, localizado como UCS (Uterine Carcinosarcoma) no TCGA, com 3-10% de frequência entre os casos de câncer endometrial, porém, de acordo com estudos recentes, a origem de mutação inicia-se no endométrio e se estende ao miométrio, com prognóstico de aproximadamente 35% de taxa de sobrevivência após 5 anos (Cantrell LA, Blank SV, Duska LR. 2015. Uterine carcinosarcoma: a review of the literature. Gynecol Oncol. 137:581–588. doi: 10.1016/j.ygyno.2015.03.041.)

Portanto, torna-se evidente a necessidade analisar quais mutações são comuns entre UCEC e UCS, observando aspectos que auxiliam na caracterização de cada tipo de câncer, além de observar as mutações exclusivas do carcinossarcoma.

**2. Justificativa**

Este relatório visa o entendimento sobre as diferenças genômicas entre os tipos de câncer e sobre as mutações presentes em cada um auxiliam na elaboração de novas estratégias de terapias gênicas, definição de vias moleculares importantes para prevenção e tratamento do câncer, compreensão dos mecanismos e interações biológicas, entre outros parâmetros.

**3. Objetivos**

O objetivo principal do trabalho é realizar análises comparativas entre os diferentes tipos de câncer endometrial (UCS e UCEC) à fim de inferir diferenças clínicas e moleculares, além dos objetivos complementares:

* entender o funcionamento da plataforma do The Cancer Genome Atlas;
* realizar análises em arquivos de mutação de exoma, expressão gênica e dados clínicos;
* obter estatísticas sobre os dados analisados;
* observar diferenças na prevalência de mutações.

**4. Metodologia**

Descrição do Estudo:  
Este estudo utiliza uma abordagem de bioinformática para analisar dados genômicos e clínicos provenientes de carcinomas endometriais uterinos (UCEC) e carcinossarcomas uterinos (UCS). O objetivo é identificar diferenças moleculares e clínicas entre esses dois subtipos de câncer, concentrando-se na caracterização de mutações somáticas, expressão gênica diferencial e correlações com dados clínicos, permitindo inferir mecanismos biológicos e potenciais alvos terapêuticos.

Banco de Dados:  
Os dados analisados foram obtidos do repositório público *The Cancer Genome Atlas* (TCGA), que disponibiliza uma vasta gama de dados genômicos e clínicos. As coortes utilizadas incluem:

* UCEC (Uterine Corpus Endometrial Carcinoma): Mais de 500 casos selecionados.
* UCS (Uterine Carcinosarcoma): 57 casos selecionados.

Os dados extraídos incluem:

* Arquivos de RNA-Seq (STAR Count): Para análise de expressão gênica.
* Mutações Somáticas (formato MAF): Anotação de variantes genéticas.
* Dados Clínicos: Informações associadas às características dos pacientes.

Os dados podem ser acessados no portal do TCGA (<https://portal.gdc.cancer.gov>).

Pré-processamento e Armazenamento de Dados:  
O pré-processamento foi realizado em duas etapas principais:

1. RNA-Seq (Expressão Gênica):
   * Integração dos arquivos STAR Count para gerar uma matriz de contagem de reads.
   * Normalização dos dados utilizando ferramentas como DESeq2 para minimizar o viés técnico.
2. Mutações Somáticas:
   * Processamento dos arquivos MAF para extração de variantes relevantes, como mutações do tipo Missense, Nonsense, SNPs, inserções e deleções.
   * Filtragem das mutações por tipo de variante e frequência, criando uma matriz estruturada de mutações.

Os dados processados foram armazenados em um ambiente de computação de alto desempenho (HPC) e estruturados em formatos compatíveis com ferramentas estatísticas, como R e Python.

Análise de Dados e Estatística:  
Para atender aos objetivos propostos, foram realizadas as seguintes análises:

1. Análise Genômica:
   * Identificação e caracterização de variantes genéticas, com destaque para padrões de substituição nucleotídica (SNVs) e os genes mais frequentemente alterados.
   * Avaliação da carga mutacional por amostra, evidenciando a maior heterogeneidade do UCEC e a predominância de mutações no gene TP53 no UCS.
2. Análise de Expressão Gênica:
   * Comparação diferencial de expressão gênica utilizando DESeq2.
   * Visualização dos resultados em Volcano Plots, destacando genes diferencialmente expressos por significância estatística (*padj*) e magnitude de alteração (*log2 Fold Change*).
3. Integração de Dados Clínicos:
   * Avaliação das correlações entre mutações específicas e características clínicas, como estágio tumoral e agressividade.
4. Visualizações e Comparações:
   * Geração de gráficos como *oncoplots* para análise de distribuição mutacional.
   * Comparação do perfil genômico de UCS e UCEC, destacando diferenças na diversidade mutacional e alterações em genes-chave.

Ferramentas e Métodos Estatísticos:  
As análises foram conduzidas utilizando R, com os pacotes especializados *maftools* e *DESeq2*. Métricas estatísticas foram aplicadas, incluindo ajuste de valores p para múltiplas comparações.

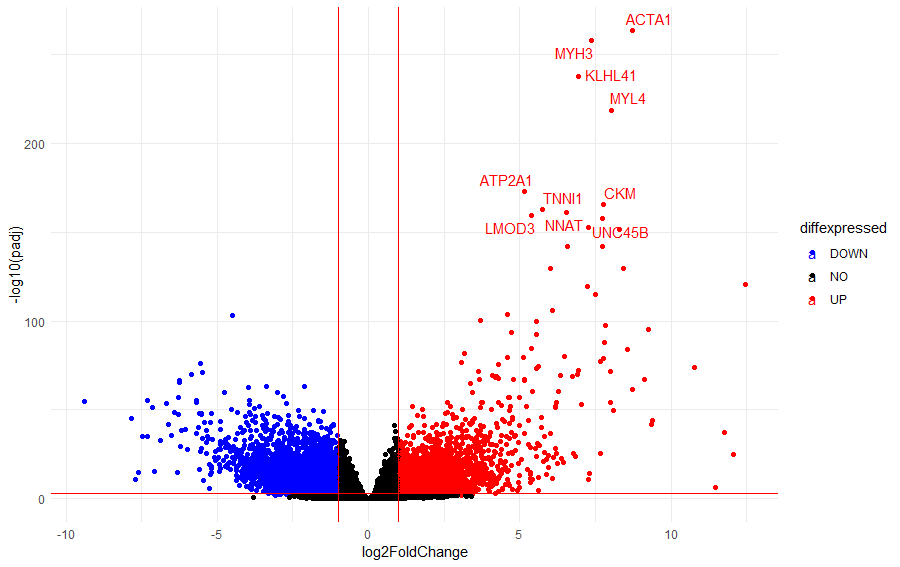
Essas abordagens permitiram identificar padrões moleculares distintos entre UCS e UCEC, auxiliando na priorização de biomarcadores e no desenvolvimento de estratégias terapêuticas personalizadas.

**5. Resultados**

Nesta seção, são apresentados os resultados das análises genômicas realizadas com os dados de mutações somáticas dos grupos UCEC (*Uterine Corpus Endometrial Carcinoma*) e UCS (*Uterine Carcinosarcoma*) obtidos do The Cancer Genome Atlas (TCGA). As análises visam identificar e caracterizar as variantes genéticas, avaliar a distribuição da carga mutacional, e investigar os genes mais frequentemente alterados, destacando as diferenças genômicas entre esses dois tipos de câncer. Além disso, serão discutidas as correlações entre mutações e dados clínicos, bem como outros aspectos relevantes que emergem das análises.

Análises Genômicas:

Figura 2- Distribuição de Expressão Gênica Diferencial entre UCEC e UCS.



Fonte: De autoria própria

O *Volcano Plot* (Imagem 2) oferece uma análise visual dos genes diferencialmente expressos entre os subtipos de câncer endometrial UCEC (*Uterine Corpus Endometrial Carcinoma*) e UCS (*Uterine Carcinosarcoma*). Ele combina a magnitude das alterações na expressão gênica (*log2 Fold Change*) com a significância estatística (*-log10 padj*), facilitando a identificação dos genes mais relevantes. Genes com expressão aumentada (em vermelho) podem estar associados à progressão tumoral, sendo potenciais biomarcadores ou alvos terapêuticos, especialmente no UCS, que é mais agressivo. Por outro lado, genes com expressão reduzida (em azul) podem estar envolvidos na supressão de vias tumorais. Já os genes em preto não apresentam alterações significativas.

Figura 3- Tabela de genes mais significativos da expressão diferencial.

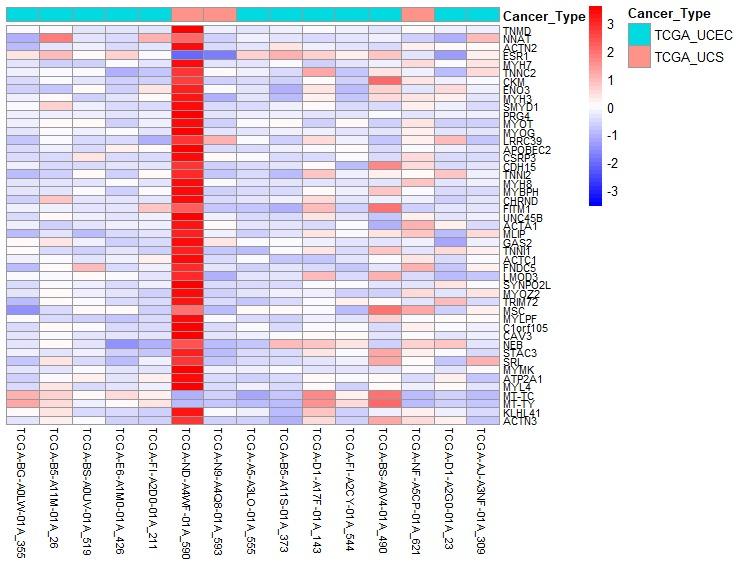


Fonte: De autoria própria

A Tabela (Imagem 3) complementa o *Volcano Plot*, listando genes com alta significância estatística e maior expressão diferencial. Exemplos como o gene ENSG00000143632.14, com log2 Fold Change de 8,71 e *padj* extremamente baixo (5,03e-264), reforçam a relevância desses genes. Essa análise permite detalhar características moleculares e potenciais alvos terapêuticos para o estudo.

A combinação do *Volcano Plot* e da tabela apresenta uma visão abrangente sobre a expressão gênica diferencial, destacando diferenças moleculares entre UCEC e UCS. Tais achados sugerem perfis genômicos distintos, com genes diferencialmente expressos que podem contribuir para estratégias terapêuticas personalizadas.

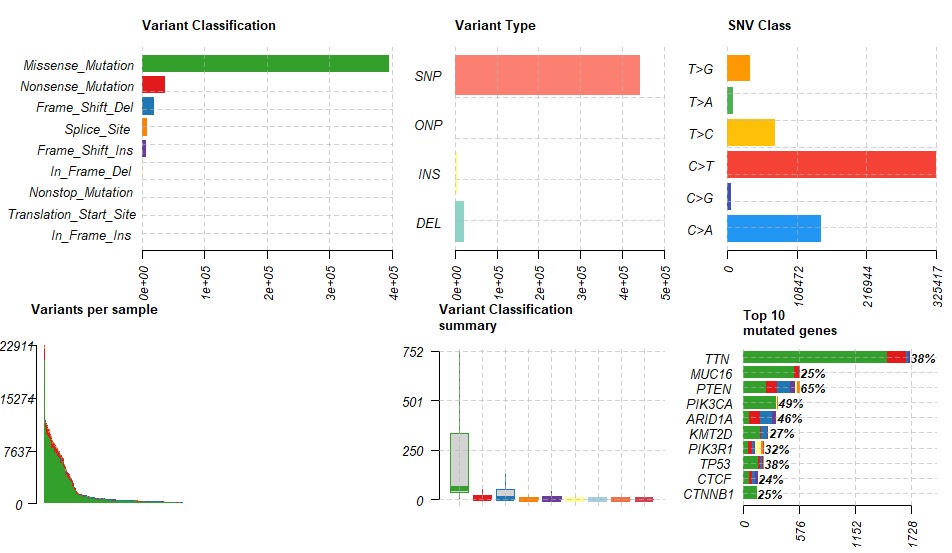
Figura 4- Tabela de genes mais significativos da expressão diferencial.



Fonte: De autoria própria

O gene TTN demonstra alta relevância em amostras associadas ao TCGA UCS, destacando-se como um possível marcador característico deste subtipo de câncer. Em contraste, as amostras de TCGA UCEC exibem maior homogeneidade nos padrões de expressão para a maioria dos genes, sugerindo menos variabilidade molecular. Genes como MUC16 e PIK3CA apresentam padrões semelhantes entre os dois subtipos, indicando que podem ter papéis comuns em ambos os contextos. Alguns genes, como FAT3 e FAT4, mostram níveis consistentemente baixos de expressão ou relevância, sem grande variação entre os grupos. Além disso, observa-se que os genes com expressão diferencial mais pronunciada, como os marcados por colorações vermelhas intensas ou azuis profundas, podem ser candidatos relevantes para diferenciação molecular entre os subtipos. Essa diferenciação é particularmente evidente em genes relacionados à função tumoral ou à sinalização celular, possivelmente refletindo vias biológicas específicas de cada subtipo.

Figura 5- Perfil Genômico do UCEC (Uterine Corpus Endometrial Carcinoma).

****

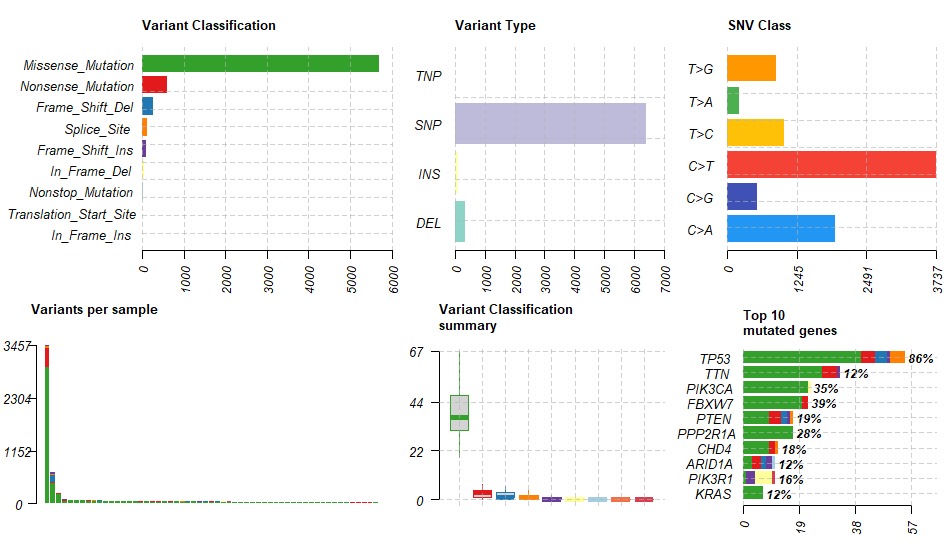
Fonte: De autoria própria.

No carcinoma endometrial uterino (UCEC), a análise genômica revelou que a maior parte das variantes detectadas pertence à categoria de *Missense Mutations*, seguida por *Nonsense Mutations* and *Frame Shift Del*. Essas alterações possuem grande potencial de impactar a funcionalidade protéica, indicando desregulação em vias moleculares essenciais para o desenvolvimento do tumor.

Os *Single Nucleotide Polymorphisms* (SNPs) são o tipo de variante mais prevalente, consistindo em uma característica comum a muitos tipos de câncer. Inserções (INS) e deleções (DEL), embora menos frequentes, podem ter relevância funcional importante. Entre as substituições de nucleotídeos (SNVs), a troca C>T é a mais comum, seguida por C>A e T>C, sugerindo mecanismos de mutagênese como desaminação de citosina e danos ao DNA relacionados ao estresse oxidativo.

A análise por amostra mostrou alta heterogeneidade na carga mutacional, com algumas amostras apresentando um número elevado de mutações, caracterizando a diversidade genética do UCEC. Os genes mais frequentemente mutados incluem TTN, MUC16, PTEN e PIK3CA. Enquanto TTN pode ser um gene "passageiro" devido ao seu tamanho, mutações em PTEN e PIK3CA estão associadas a processos críticos como a sinalização celular e o crescimento tumoral.

Figura 6- Perfil Genômico do UCS (Uterine Carcinosarcoma).

****

Fonte: De autoria própria.

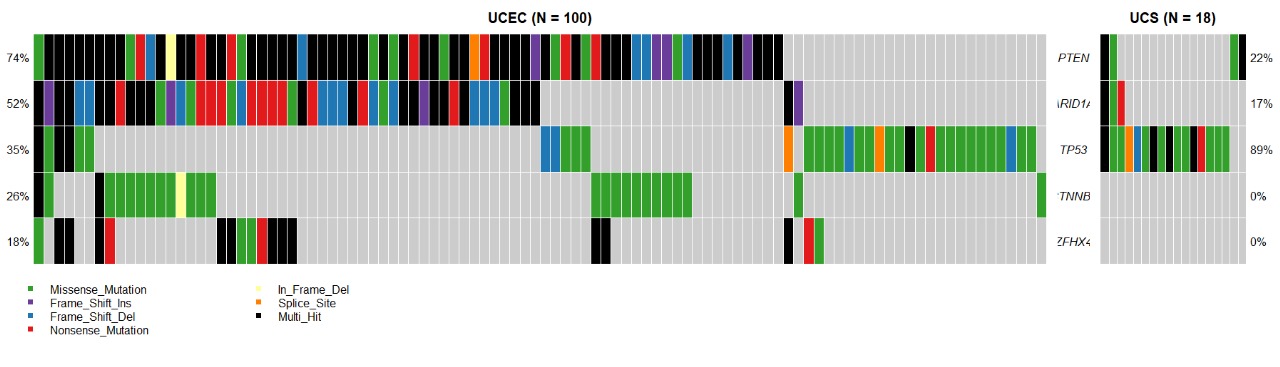
O carcinossarcoma uterino (UCS) apresenta um perfil genômico distinto do UCEC. Assim como no UCEC, as *Missense Mutations* são as variantes mais frequentes, mas o número total de mutações detectadas no UCS é significativamente menor, refletindo uma carga mutacional reduzida nesse tipo de tumor.

Os SNPs também predominam no UCS, enquanto inserções e deleções são menos comuns. O padrão de substituições de nucleotídeos (SNVs) é semelhante ao do UCEC, com C>T como a substituição mais prevalente, sugerindo mecanismos de mutagênese compartilhados, como estresse oxidativo.

A carga mutacional no UCS é consideravelmente menor quando avaliada por amostra, com a maioria apresentando poucas mutações. Entre os genes mais frequentemente mutados, TP53 se destaca, com uma frequência de mutação de 86%, evidenciando seu papel central na tumorigênese desse tipo de câncer. Outros genes como PIK3CA e FBXW7 aparecem em menor proporção, sugerindo que o UCS é impulsionado principalmente por mutações críticas em genes-chave, particularmente TP53.

Os resultados apresentados em Figura 2 e Figura 3 destacam diferenças significativas entre os perfis genômicos de UCEC e UCS. O UCEC apresenta maior diversidade mutacional, refletindo sua complexidade genética, enquanto o UCS é caracterizado por um foco em alterações pontuais, especialmente em TP53. Essas diferenças têm implicações importantes no desenvolvimento de terapias personalizadas, com estratégias mais amplas necessárias para o UCEC e abordagens direcionadas para o UCS.

Figura 7- Distribuição de mutações gênicas mais frequentes em UCEC e UCS, com os tipos de mutações representados por diferentes cores.



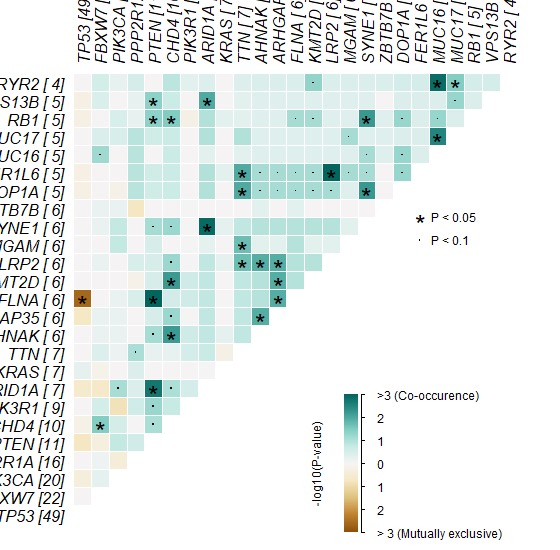
Fonte: De autoria própria.

A partir da análise de maf compare ilustrada na imagem, é possível fazer inferências relevantes para o maior entendimento das características do carcinossarcoma (UCS) quando comparado ao grupo de cânceres de endométrio (UCEC).

No caso do UCEC, nota-se uma distribuição mais heterogênea nas mutações, enquanto no UCS 89% das mutações ocorrem no gene TP53. Isso sugere que esse gene pode desempenhar uma função crucial na patogênese deste tipo de câncer.

Além disso, percebe-se que há uma maior diversidade nos tipos de mutação em UCEC, representadas nesse gráfico por faixas de diferentes cores. Em contraste, em UCS, há uma predominância clara de mutações do tipo missense e nonsense, com poucas ocorrências de outros tipos. Essa diferença indica maior complexidade de UCEC quando comparado a UCS, que pode ser atribuída ao fato de UCEC consistir de uma categoria que abrange todos os cânceres endometriais.

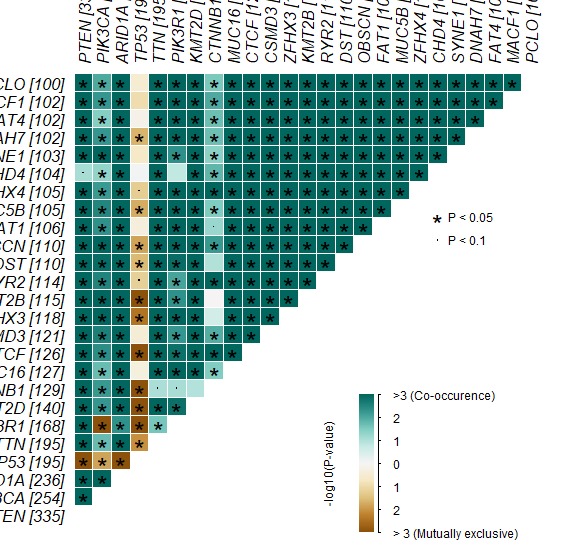
Figura 8- Interações somáticas para os genes UCS.



Fonte: De autoria própria.

O gráfico apresentado na Figura 5 é uma representação visual que ilustra a relação entre as mutações em diferentes genes. As cores indicam se as mutações são co-ocorrências (quando ocorrem juntas) ou mutuamente exclusivas (isto é, nunca ocorrem simultaneamente). No gráfico, o símbolo da estrela (★) sinaliza alta significância estatística, enquanto o ponto (ᐧ) representa baixa significância estatística. O gene TP53 se destaca por apresentar um grande número de associações com outros genes, sugerindo que ele pode estar diretamente envolvido nos processos biológicos relacionados à formação do UCS.

Figura 9- Interações somáticas para os genes UCEC.

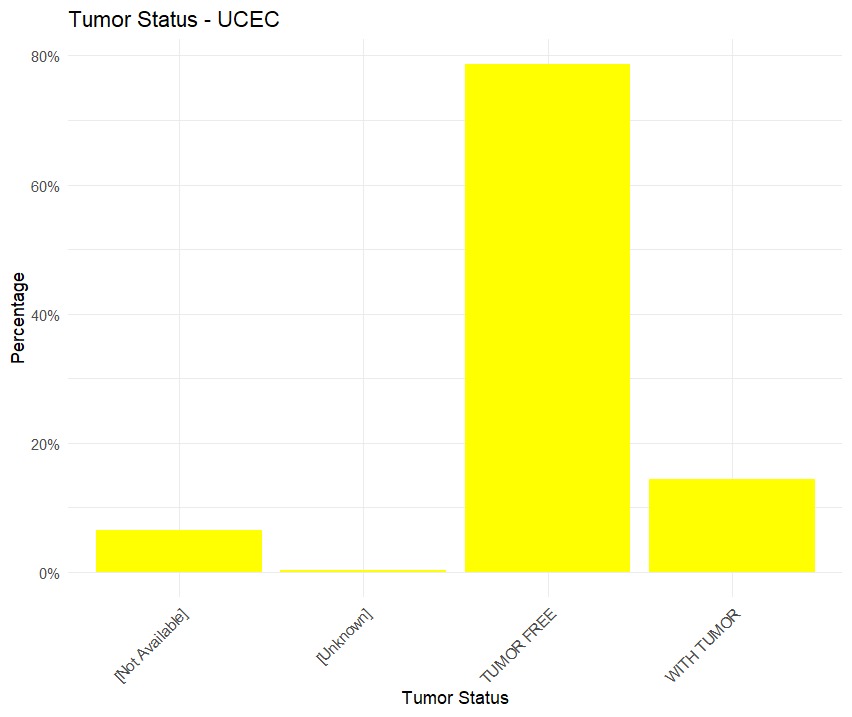


Fonte: De autoria própria.

Entre as principais informações, observa-se que algumas combinações de genes apresentam altos valores de coocorrência, sugerindo que essas mutações tendem a ocorrer juntas no câncer endometrial. A região com tons mais intensos ao longo da diagonal principal indica genes frequentemente mutados simultaneamente. Por outro lado, algumas combinações destacadas em marrom indicam que mutações nesses genes ocorrem de forma excludente, sugerindo que mutações em um gene tornam desnecessárias mutações em outro no desenvolvimento do câncer. Os asteriscos (\*) em muitos pares indicam que a interação entre os genes não é aleatória e possui relevância biológica ou clínica.

Análises Clínicas:

Figura 10- Valores da Tumor Status para UCEC.

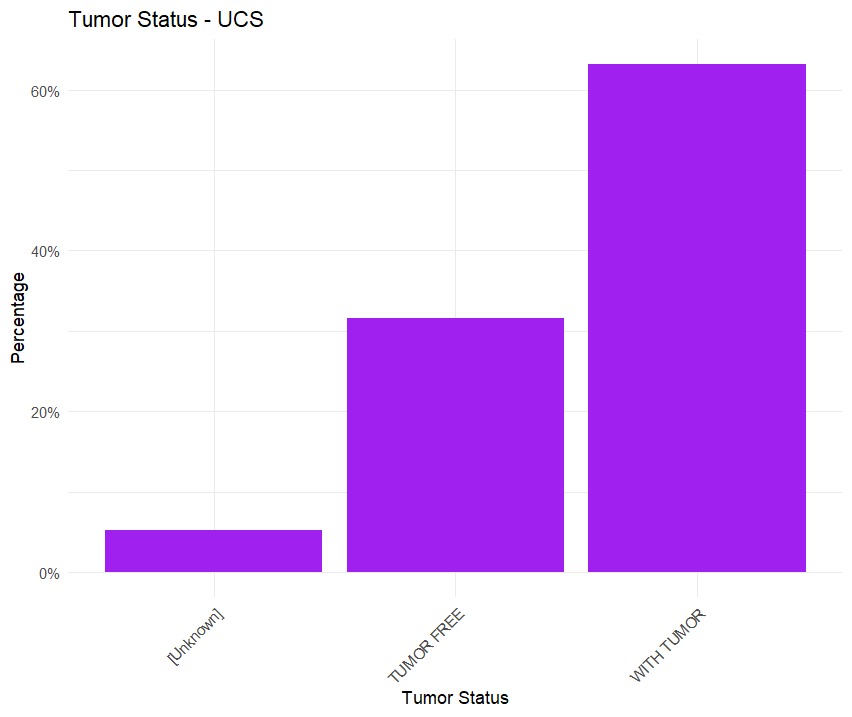


Fonte: De autoria própria.

O eixo X representa diferentes categorias de status tumoral, como [Not Available], [Unknown], Tumor-Free e With Tumor, enquanto o eixo Y exibe a percentagem de indivíduos em cada categoria. A maior parte dos pacientes (cerca de 80%) encontra-se na categoria Tumor-Free, indicando ausência de tumor no momento do estudo. Menos de 20% estão na categoria With Tumor, enquanto dados classificados como [Not Available] ou [Unknown] são mínimos.

A alta proporção de pacientes Tumor-Free sugere uma população majoritariamente em remissão, o que pode impactar análises de características associadas ao estado tumoral.

Figura 11- Valores da Tumor Status para UCS



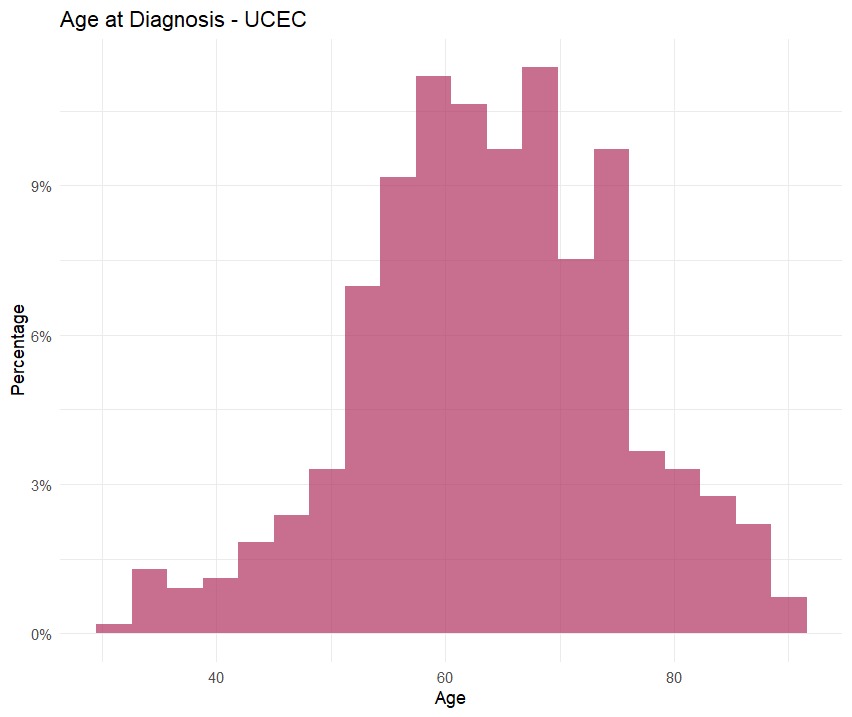
Fonte: De autoria própria.

O gráfico intitulado "Tumor Status - UCS" apresenta a distribuição do status tumoral em uma população estudada, utilizando um formato de barras para facilitar a visualização dos dados. No eixo horizontal (X), estão representadas três categorias de status do tumor: "Unknown" (Desconhecido), "TUMOR FREE" (Sem Tumor) e "WITH TUMOR" (Com Tumor). No eixo vertical (Y), é indicada a porcentagem de indivíduos pertencentes a cada uma dessas categorias.

A análise do gráfico mostra que a categoria "Unknown" possui uma porcentagem muito baixa, enquanto a categoria "TUMOR FREE" exibe uma quantidade moderada de indivíduos. A maior representatividade está na categoria "WITH TUMOR", que ultrapassa 60% do total.

Esse gráfico é útil para compreender a distribuição do status tumoral entre os participantes do estudo. Ele revela que mais de 60% dos indivíduos têm tumores, uma parcela moderada está livre de tumores, e uma pequena fração tem o status de tumor classificado como desconhecido.

Figura 12- Valores da Age at Diag. para UCS.



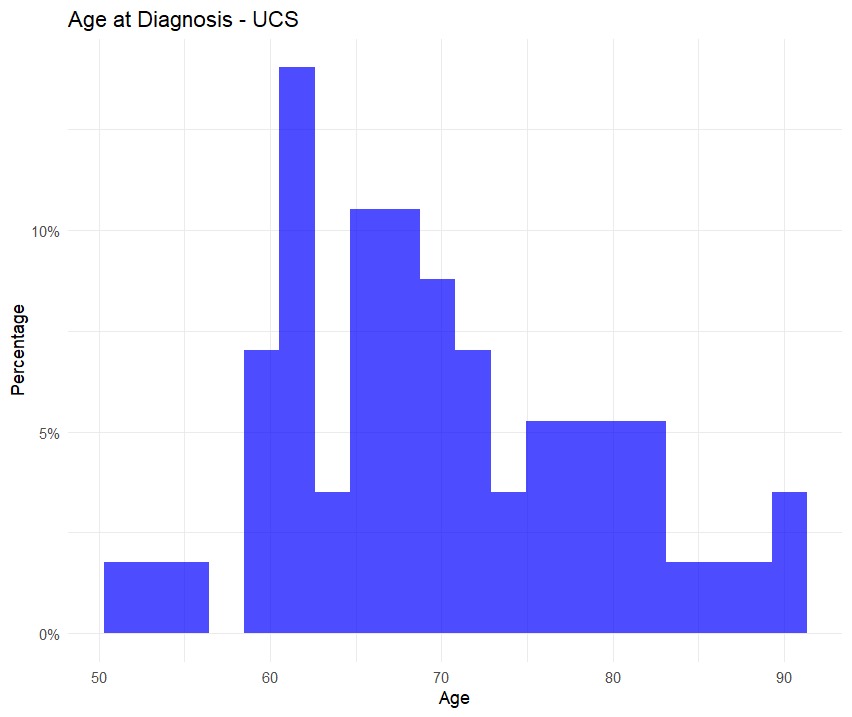
Fonte: De autoria própria.

O gráfico fornecido é um histograma intitulado "Age at Diagnosis - UCEC" (Idade ao Diagnóstico - Carcinoma Endometrial do Corpo Uterino). Ele apresenta a relação entre a idade dos indivíduos no momento do diagnóstico e a porcentagem de ocorrências em cada faixa etária.

No eixo horizontal, estão representadas as idades dos indivíduos ao serem diagnosticados, enquanto o eixo vertical mostra a porcentagem de diagnósticos em cada intervalo de idade. A análise do histograma revela que a faixa etária com maior porcentagem de diagnósticos está em torno dos 60 anos. Além disso, o gráfico apresenta uma distribuição aproximadamente em forma de sino (bell-shaped), indicando que a maioria dos diagnósticos ocorre entre as idades de 50 e 70 anos. Diagnósticos em idades mais jovens e em idades mais avançadas são consideravelmente menos frequentes.

Essa distribuição etária é fundamental para compreender o perfil demográfico do Carcinoma Endometrial do Corpo Uterino (UCEC). Os dados são particularmente relevantes para pesquisas médicas e para o planejamento de políticas de saúde pública, ajudando a identificar as faixas etárias de maior risco e possibilitando intervenções mais direcionadas e eficazes.

Figura 13- Valores da Age at Diag. para UCS.

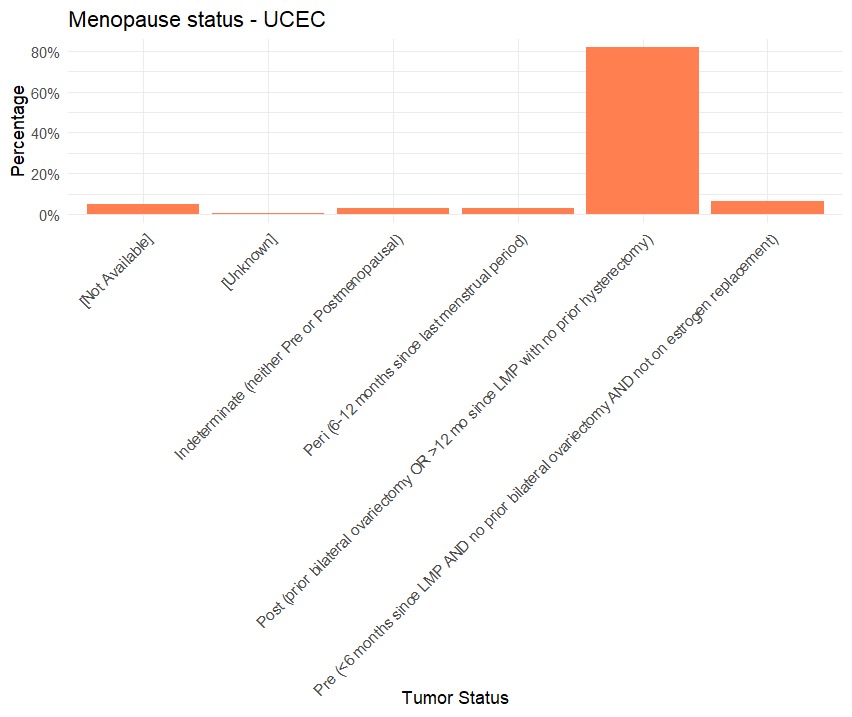


Fonte: De autoria própria

O gráfico da figura 13 mostra a idade em que os pacientes foram diagnosticados com carcinossarcoma uterino. É possível identificar um pico nos 60 anos, com as incidências ocorrendo em sua maioria depois dessa idade.

Quando comparado ao carcinoma endometrial, é possível observar similaridades nas idades de diagnóstico, apesar da diferença de tamanho de amostra tornar os gráficos bem diferentes.

Figura 14- Valores da Menopause Status para UCS.



Fonte: De autoria própria.

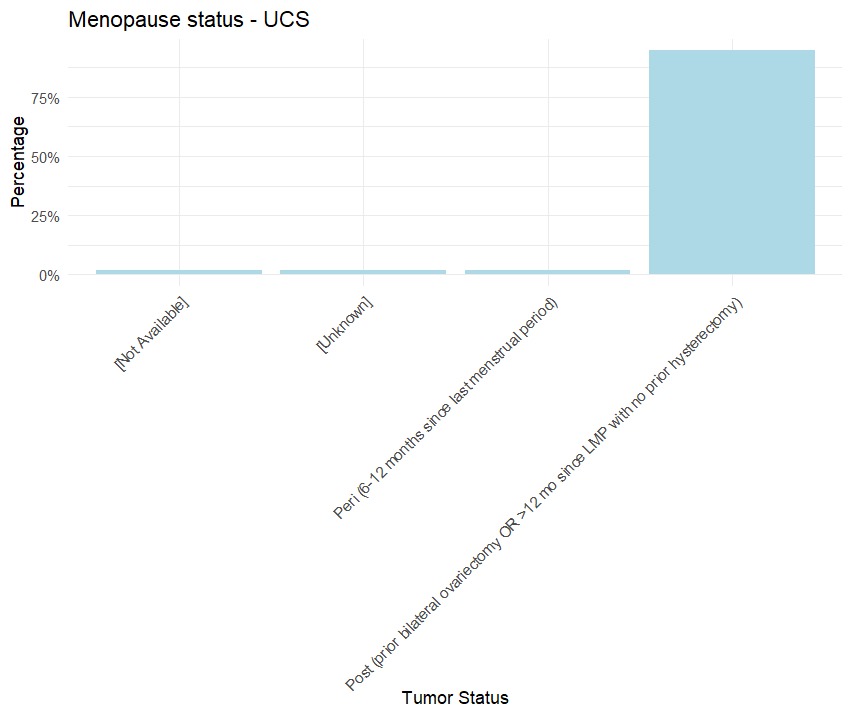
O gráfico intitulado "Menopause status - UCEC" (Status de Menopausa - Carcinoma Endometrial do Corpo Uterino) apresenta a distribuição dos diferentes status de menopausa entre indivíduos diagnosticados com UCEC. O eixo horizontal representa as categorias de status de menopausa, enquanto o eixo vertical indica a porcentagem correspondente de indivíduos em cada categoria.

As categorias incluem dados classificados como [Not Available] (Não Disponível), [Unknown] (Desconhecido), Indeterminate (Indeterminado, indicando que o indivíduo não é claramente pré nem pós-menopausa), Peri (<12 meses desde a última menstruação), Peri (<12 meses ou >12 meses desde a última menstruação sem histerectomia prévia), Post (>6 meses desde a última menstruação, sem ovariectomia bilateral prévia e sem reposição de estrogênio), Post (ovariectomia bilateral prévia e sem reposição de estrogênio), e Pre (<6 meses desde a última menstruação, sem ovariectomia bilateral prévia).

A análise revela que a categoria "Post (>6 meses desde a última menstruação, sem ovariectomia bilateral prévia e sem reposição de estrogênio)" apresenta a maior porcentagem, em torno de 60%. Todas as outras categorias possuem porcentagens significativamente menores.

Esse gráfico é importante para entender a distribuição dos status de menopausa entre os indivíduos diagnosticados com UCEC. Ele fornece insights valiosos sobre os fatores de risco e o perfil demográfico da doença, que podem contribuir para o desenvolvimento de estratégias de prevenção e tratamento mais eficazes.

Figura 15- Valores da Menopause Status para UCEC.



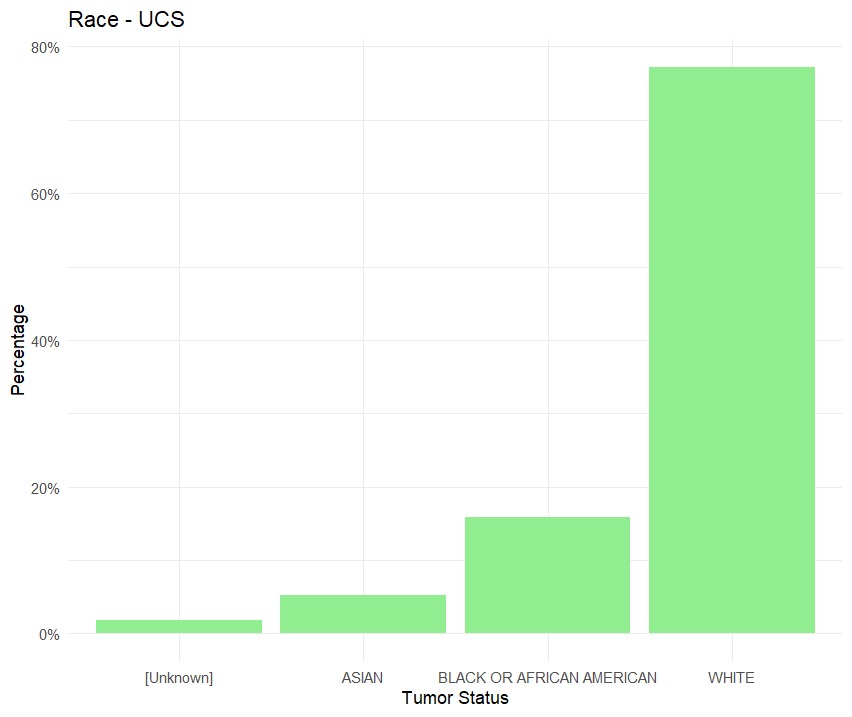
Fonte: De autoria própria.

O gráfico intitulado "Menopause status - UCS" apresenta a distribuição dos diferentes status de menopausa entre indivíduos diagnosticados com UCS (Uterine Carcinosarcoma). No eixo horizontal (X), são representadas as categorias de status de menopausa, que incluem: [Not Available] (Não Disponível), [Unknown] (Desconhecido), Peri (<12 meses desde a última menstruação) e Post (ovariectomia bilateral prévia ou >12 meses desde a última menstruação sem histerectomia prévia). O eixo vertical (Y) mostra a porcentagem correspondente a cada categoria, variando de 0% a 100%.

A análise do gráfico revela que a categoria "Post (ovariectomia bilateral prévia ou >12 meses desde a última menstruação sem histerectomia prévia)" apresenta a maior porcentagem, próxima de 100%. Em contrapartida, as demais categorias, como "Not Available", "Unknown" e "Peri", têm porcentagens muito baixas, praticamente inexistentes.

Esses dados indicam que quase todos os indivíduos estudados estão na categoria "Post", sugerindo uma predominância de mulheres na pós-menopausa entre os casos de UCS. Essa informação é relevante para entender os fatores de risco associados a essa condição e para o planejamento de estratégias preventivas e terapêuticas voltadas para essa população.

Figura 16- Valores da Race para UCS.

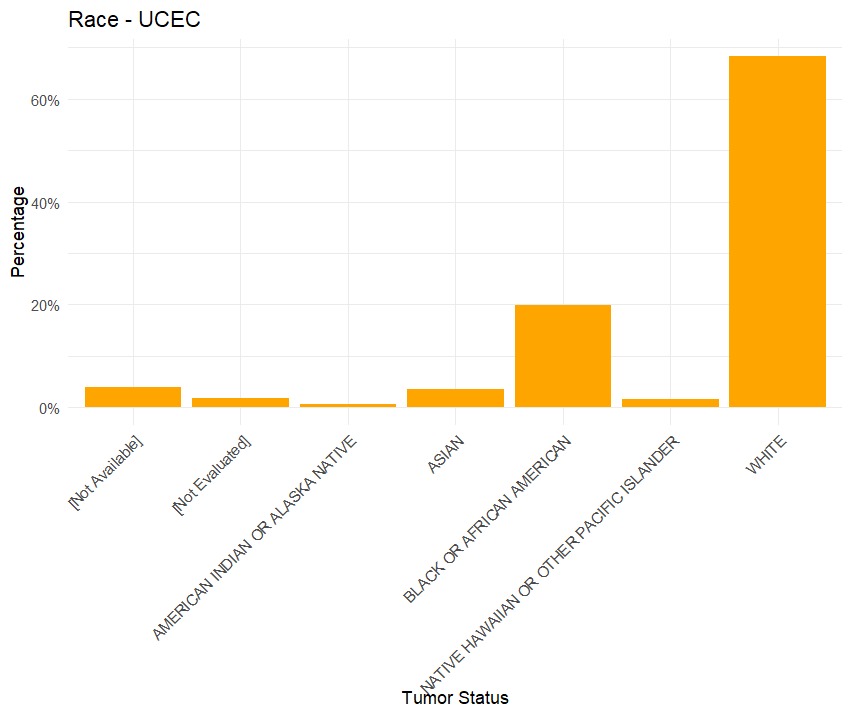


Fonte: De autoria própria.

O eixo X apresenta as diferentes categorias raciais, incluindo [Unknown], Asian, Black or African American e White, enquanto o eixo Y indica a percentagem de cada grupo racial. Observa-se um predomínio significativo da população branca, representando cerca de 80% do total. Outros grupos raciais, como Black or African American e Asian, contribuem com menos de 20%. Há uma pequena fração de dados classificados como [Unknown].

A predominância da população branca reflete possivelmente a composição demográfica da base de dados ou dos participantes do estudo. Essa sub-representação de outros grupos pode indicar a necessidade de maior atenção para evitar vieses nas análises relacionadas à etnia e status tumoral.

Figura 17- Valores da Race para UCEC.



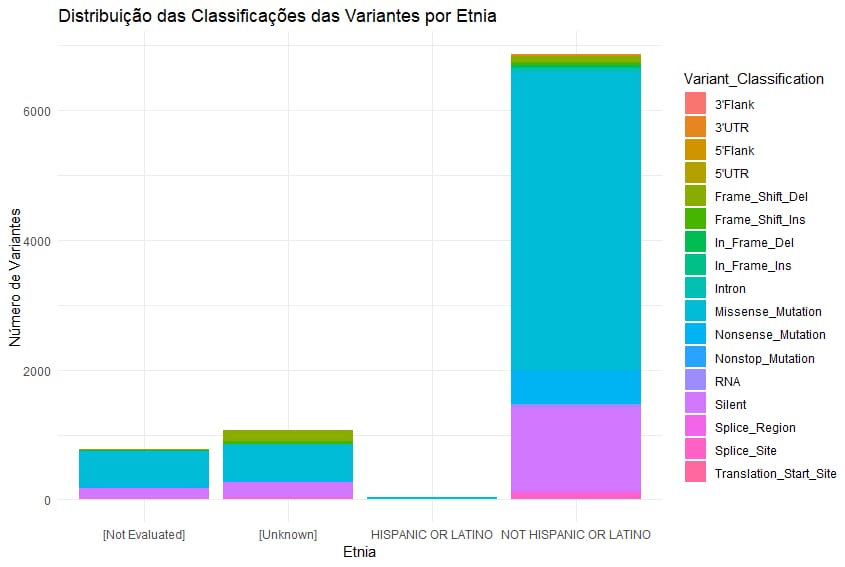
Fonte: De autoria própria.

O eixo X apresenta categorias raciais como American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White e outras classificações. O eixo Y mostra a percentagem relativa de cada grupo. Assim como em UCS, a população branca é predominante, representando mais de 60%. Black or African American e Asian aparecem como minorias, mas são mais representativos que outros grupos, como American Indian or Alaska Native, Native Hawaiian e Pacific Islander, cujas contribuições são insignificantes (próximas a 0%).

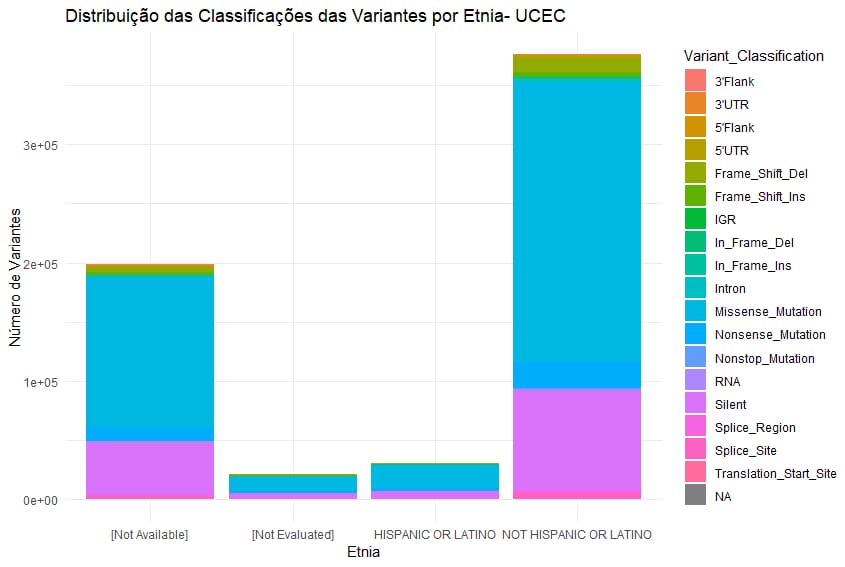
Esse padrão sugere um viés demográfico que favorece a população branca, aspecto relevante ao considerar a generalização dos resultados do estudo. A representatividade limitada de minorias pode afetar a aplicabilidade dos achados para essas populações.

Análises que combinam informações clínicas e dados de mutações:

Figura 18 e 19 - Comparação da Distribuição das Classificações das Variantes por Etnia para UCS e UCEC.



Fonte: De autoria própria.



Fonte: De autoria própria.

A Figura 18, intitulada "Distribuição das Classificações das Variantes por Etnia", mostra a distribuição das variantes genéticas em diferentes grupos étnicos. No eixo horizontal (X), estão representados os grupos étnicos: *[Not Evaluated]* (Não Avaliado), *[Unknown]* (Desconhecido), *HISPANIC OR LATINO* (Hispânico ou Latino), e *NOT HISPANIC OR LATINO* (Não Hispânico ou Latino). O eixo vertical (Y) indica o número de variantes genéticas, variando de 0 a 7000.

As variantes genéticas são classificadas em várias categorias, que são representadas por diferentes cores no gráfico. Essas categorias incluem: *3'Flank*, *3'UTR*, *5'Flank*, *5'UTR*, *Frame\_Shift\_Del*, *Frame\_Shift\_Ins*, *In\_Frame\_Del*, *In\_Frame\_Ins*, *Intron*, *Missense\_Mutation*, *Nonsense\_Mutation*, *Nonstop\_Mutation*, *RNA*, *Silent*, *Splice\_Region*, *Splice\_Site*, e *Translation\_Start\_Site*.

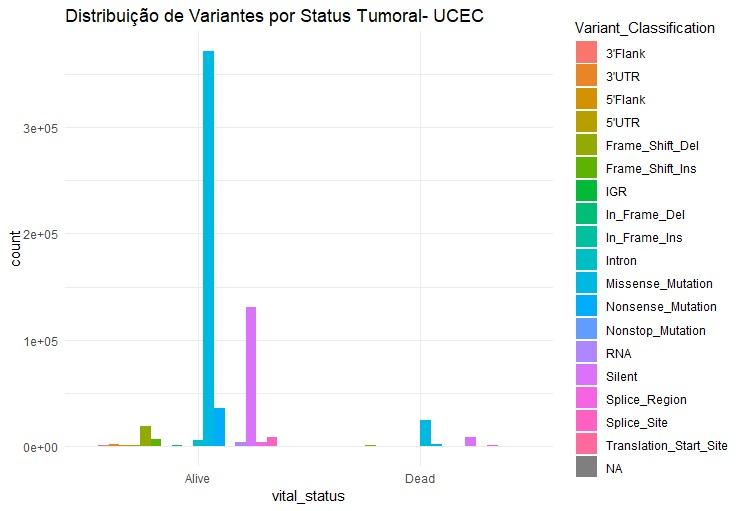
A análise do gráfico revela que o grupo *NOT HISPANIC OR LATINO* apresenta o maior número de variantes genéticas, indicando uma maior diversidade ou quantidade de variantes nesta população. O grupo *HISPANIC OR LATINO* tem um número considerável de variantes, mas significativamente menor em comparação ao grupo *NOT HISPANIC OR LATINO*. Já os grupos *[Not Evaluated]* e *[Unknown]* apresentam menos variantes genéticas, o que pode refletir uma menor quantidade de dados ou estudos disponíveis para essas populações.

O gráfico, identificado como Figura 19, apresenta a distribuição das classificações de variantes genéticas por etnia no contexto do UCEC (Carcinoma Endometrial do Corpo Uterino). No eixo vertical, são representados os números de variantes genéticas, enquanto o eixo horizontal exibe diferentes categorias étnicas: *Not Available* (Não Disponível), *Not Evaluated* (Não Avaliado), *HISPANIC OR LATINO* (Hispânico ou Latino) e *NOT HISPANIC OR LATINO* (Não Hispânico ou Latino).

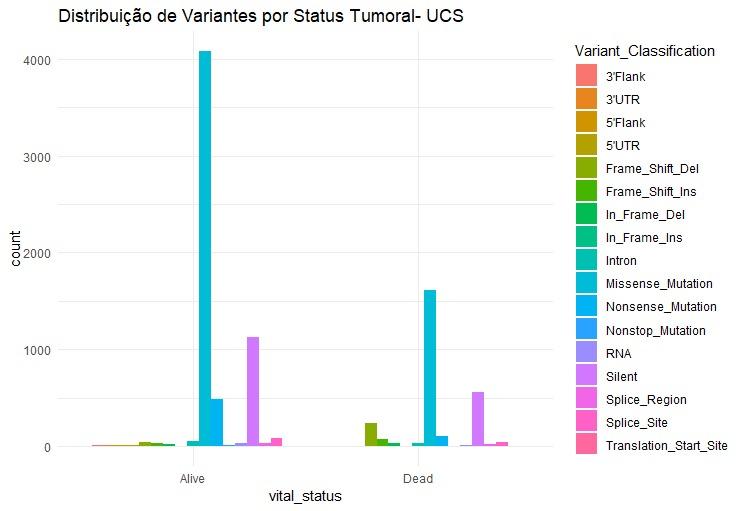
Cada variante genética é classificada em categorias como "3'Flank", "3'UTR", "5'Flank", "5'UTR", "Frame\_Shift\_Del", "Frame\_Shift\_Ins", "IGR", "In\_Frame\_Del", "In\_Frame\_Ins", "Intron", "Missense\_Mutation", "Nonsense\_Mutation", "Nonstop\_Mutation", "RNA", "Silent", "Splice\_Region", "Splice\_Site", "Translation\_Start\_Site" e "NA". A legenda, localizada à direita do gráfico, associa cada categoria a uma cor distinta, facilitando a interpretação visual.

A análise do gráfico revela que a maior concentração de variantes está nas categorias étnicas *Not Available* e *NOT HISPANIC OR LATINO*. Nesses grupos, as variantes "Missense\_Mutation" e "Silent" são predominantes. Já a categoria *HISPANIC OR LATINO* apresenta um número significativamente menor de variantes em comparação aos demais grupos. A categoria *Not Evaluated* também tem um número reduzido de variantes, mas ainda é relevante.

Figura 20 e 21 - Comparação da Distribuição das Classificações das Variantes por Status Tumoral para UCS e UCEC.



Fonte: De autoria própria.



Fonte: De autoria própria.

Analisando os dados mostrados no gráfico que relaciona o status vital dos pacientes com o tipo de classificação da variante, é possível notar que a maior parte das mutações ocorridas nos pacientes vivos de UCS é do tipo missense e silenciosa, enquanto nos pacientes que vieram à óbito há presença de uma parcela de mutação do tipo *Frame Shift.* De forma similar, os dados de UCEC mostram uma porcentagem visível de mutações missense e silenciosa, porém, diferente do grupo observado em UCS, há ocorrência de mutação *Frame Shift Ins* no grupo de pacientes vivos.

**6. Conclusão**

A análise genômica comparativa entre o Carcinossarcoma Uterino (UCS) e o Carcinoma Endometrial Uterino (UCEC) revelou diferenças significativas nos perfis moleculares e clínicos dos dois tipos de câncer. O UCEC apresentou uma maior diversidade mutacional, com uma carga mutacional variada entre as amostras, sugerindo uma maior complexidade genética e múltiplos caminhos de evolução tumoral. As mutações mais prevalentes envolveram genes como PTEN e PIK3CA, que desempenham papéis cruciais na sinalização celular e no crescimento tumoral. Em contraste, o UCS mostrou uma predominância de mutações no gene TP53, com 86% de mutações neste gene, indicando sua importância central na patogênese desse câncer mais raro, mas agressivo.

Essas descobertas fornecem insights valiosos para o desenvolvimento de terapias direcionadas, com o UCS sugerindo que terapias focadas em TP53 podem ser mais eficazes, enquanto o UCEC pode se beneficiar de abordagens mais amplas devido à sua maior heterogeneidade genética. A integração dos dados genômicos com informações clínicas poderá orientar estratégias de tratamento mais personalizadas, aprimorando o prognóstico e a qualidade de vida das pacientes com esses tipos de câncer.

Além disso, as análises clínicas revelam variações nos status tumoral, idade ao diagnóstico e menopausa entre os grupos, sugerindo padrões demográficos distintos. Por exemplo, a maioria dos pacientes com UCS apresenta tumores ativos (60%), enquanto o UCEC tem maior proporção de pacientes em remissão. Em termos de idade ao diagnóstico, ambos os grupos convergem em torno de 60 anos, com prevalência de mulheres na pós-menopausa.

Por fim, a distribuição de variantes genéticas por etnia e status tumoral evidência disparidades na representatividade populacional, destacando a predominância de variantes no grupo "Not Hispanic or Latino" e diferenças nas classificações de variantes entre os estados "Tumor-Free" e "With Tumor".

Esses resultados fornecem uma base sólida para o aprofundamento das investigações e desenvolvimento de abordagens terapêuticas específicas, considerando a heterogeneidade genética e as características clínicas dos subtipos UCEC e UCS.

**7. 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.

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.