

RAE-CEA-15P18

RELATÓRIO DE ANÁLISE ESTATÍSTICA SOBRE O PROJETO:

**“Análise da expressão do fator de transcrição TCF21/POD-1 e
dos genes PINK1 e BUB1B em tumores adrenocorticiais
humanos”**

**Marília Vieira Padúla
Júlia Maria Pavan Soler**

**São Paulo
2015**

CENTRO DE ESTATÍSTICA APLICADA – CEA – USP
RELATÓRIO DE ANÁLISE ESTATÍSTICA

TÍTULO: Relatório de Análise Estatística sobre o Projeto: “Análise da expressão do fator de transcrição TCF21/POD-1 e dos genes PINK1 e BUB1B em tumores adrenocorticais humanos”.

PESQUISADOR: Bárbara dos Santos Passaia

ORIENTADOR: Profa. Dra. Cláudimara Ferini Pacicco Lotfi

INSTITUIÇÃO: Instituto de Ciências Biomédicas III da Universidade de São Paulo

FINALIDADE DO PROJETO: Mestrado

ÁREA DE APLICAÇÃO: Ciências Médicas e Biológicas

RESPONSÁVEIS PELA ANÁLISE: Prof. Júlia Maria Pavan Soler
Marília Vieira Padúla

RELATÓRIO ELABORADO POR: Marília Vieira Padúla

REFERÊNCIA DESTE TRABALHO:

PADÚLA, M.V. e SOLER, J.M.P. (2015). Relatório de Análise Estatística sobre o Projeto: “Análise da expressão do fator de transcrição TCF21/POD-1 e dos genes PINK1 e BUB1B em tumores adrenocorticais humanos”. São Paulo, IME-USP, RAE-CEA-15P18

FICHA TÉCNICA

REFERÊNCIAS BIBLIOGRÁFICAS

1. BUSSAB, W. O. e MORETTIN, P. A. (2013). **Estatística Básica**. 8a.ed. São Paulo: Saraiva, 540p.
2. COLOSIMO, E.A., e GIOLO, S.R. (2006). **Análise de Sobrevivência Aplicada**. ABE – Projeto Fisher. 1a ed. São Paulo: Edgard Blucher, 392p.
3. FRANÇA, M.M., de SOUZA, B.F., SANTOS, M.G., LERARIO, A.M., FRAGOSO, C.B.V., LATRONICO, A.C., KUICK, R.D., HAMMER, G.D., LOTFI, C.F.P. (2013). POD-1 binding to the E-box sequence inhibits SF-1 and star expression in human adrenocortical tumor cells. **Molecular and Cellular Endocrinology**, 371 (2013): 140:147.
4. GOMES, S.M., AMARAL, C.M., ALBUQUERQUE, G.C.de, SILVA, F.F., e IBIAPINA, G.R. (2003). Carcinomas adrenocorticiais: divergências entre critérios prognósticos. **Rev. Ciênc. Saúde Nova Esperança**, 11(3): 72-8.
5. HOSMER, D.W. and LEMESHOW, S. (2000). **Applied Logistic Regression**. 2nd ed. New York: Wiley. 392 p.
6. KOSMIDIS, I. and FIRTH, D. (2009). Bias reduction in exponential family nonlinear models. **Biometrika**, 96(4): 793-804.
7. LIVAK, K.J., and SCHMITTGEN, T.D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta Ct}$ Method. **Methods**, 25(4): 402-8.
8. PAULA, G.A. (2013). **MODELOS DE REGRESSÃO com apoio computacional**. São Paulo: IME-USP, 428 p.

PROGRAMAS COMPUTACIONAIS UTILIZADOS

Microsoft Excel for Windows ®, versão 2013.

Microsoft Word for Windows ®, versão 2013.

MINITAB for Windows ®, versão 17

R for Windows ®, versão 3.1.3

SAS Enterprise Guide®, versão 5.1

TÉCNICAS ESTATÍSTICAS UTILIZADAS: (classificação ISI entre parênteses)

Análise Descritiva Unidimensional (03:010)

Análise Descritiva Multidimensional (03:020)

Análise de Sobrevivência (13:070)

Regressão Logística (07:090)

Resumo

O presente trabalho apresenta a análise estatística de dados referentes a tumores adrenocorticais em humanos. O principal objetivo é relacionar o diagnóstico desses tumores, em adenomas ou carcinomas, com informações genéticas, clínicas e demográficas dos pacientes. Resultados indicam grande heterogeneidade nas características genéticas e clínicas dos pacientes nos diferentes grupos de faixa etária, bem como a significância de variáveis genéticas na predição do diagnóstico dos pacientes.

Índice

1. Introdução	7
2. Objetivos	8
3. Descrição do Estudo	8
4. Descrição das Variáveis	10
4.1 Variáveis Clínicas e Demográficas	10
4.2 Variáveis Genéticas	12
5. Análise Descritiva	12
5.1 Variáveis Clínicas e Demográficas	13
5.2 Variáveis Genéticas	15
6. Análise Inferencial	18
7. Conclusões	22
APÊNDICE A - Tabelas	23
APÊNDICE B - Gráficos	34
APÊNDICE C – Modelos Finais	49

1. Introdução

As glândulas adrenais (ou suprarrenais) são glândulas endócrinas, localizadas nos polos superiores dos rins, que participam de todo metabolismo do corpo. Sua função principal está ligada à síntese de hormônios importantes no processo metabólico, e de alguns hormônios sexuais.

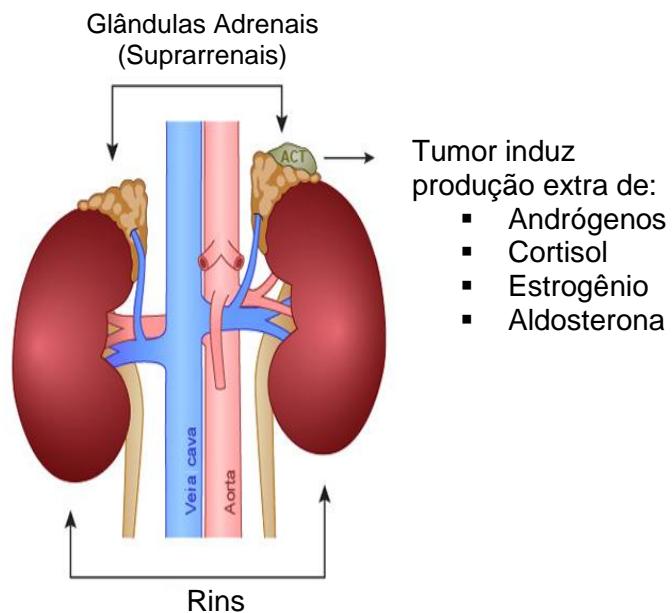


Figura 1: Localização das glândulas adrenais.

Tumores adrenocorticiais ocorrem nas células do córtex da glândula adrenal (capa externa da glândula adrenal, responsável pela produção de esteróides), podendo aparecer em qualquer idade da vida de um indivíduo.

O diagnóstico, em adenoma ou carcinoma, é definido pela combinação de critérios clínicos, hormonais e radiológicos do paciente. O que ocorre é que esses critérios, muitas vezes, não são suficientes para um diagnóstico correto do tumor (GOMES et al, 2003). Esse fato leva muitos pesquisadores a buscarem critérios adicionais para tornar esse diagnóstico mais assertivo e confiável, como é o caso desse estudo.

2. Objetivos

O principal objetivo deste estudo é determinar, a partir de características genéticas, um método mais assertivo no diagnóstico de pacientes com tumor adrenocortical. Mais especificamente, deseja-se verificar se as expressões dos genes POD1, BUB1B e PINK1 são diferenciais em adenomas e carcinomas. Ainda, pretende-se relacionar as expressões desses genes com dados clínicos e demográficos dos pacientes (Sexo, Faixa etária, Escore de Weiss, entre outros) a fim de determinar um limiar de separação entre adenomas e carcinomas baseado nas informações genéticas, além das clínicas e demográficas.

Adicionalmente, existe interesse em analisar se a expressão combinada dos genes alvo citados está associada ao diagnóstico dos pacientes, além da relação de associação entre cada par destes genes.

3. Descrição do Estudo

Os dados utilizados neste estudo referem-se a informações clínicas, demográficas e genéticas de 130 pacientes diagnosticados com tumor adrenocortical. Esses diagnósticos foram realizados entre janeiro de 1981 e outubro de 2014, no Hospital das Clínicas Campus de São Paulo - FMUSP (108 pacientes) e Campus de Ribeirão Preto - FMUSPPR (22 pacientes).

Dentre todos os pacientes, têm-se os seguintes Grupos de faixas etárias:

- Pediátricos (35 pacientes);
- Adolescentes (16 pacientes);
- Adultos (79 pacientes).

Essa categorização foi proposta pioneiramente neste estudo, já que a classificação etária comumente utilizada nas pesquisas tem sido em dois níveis: Pediátricos e Adultos (de 0 a 18 anos de idade, e acima de 18 anos de idade, respectivamente). Essa nova categorização foi proposta pois há indícios de que existam diferenças entre as características de um tumor adrenocortical nos diferentes grupos etários. As glândulas adrenais humanas ainda estão em formação até os cinco anos de idade, logo, acredita-se que considerar

pacientes com idade inferior e superior a cinco anos no mesmo grupo pode levar à perda de informações relevantes sobre o diagnóstico dos pacientes e características do tumor.

Neste estudo, a informação genética considerada foi obtida através do método de quantificação de expressão genética PCR *real-time*. Essa é uma técnica de laboratório baseada no princípio da reação em cadeia de polimerase (PCR) para multiplicar ácidos nucleicos e quantificar o DNA obtido. Resumidamente, o método quantifica determinado gene no material genético de um paciente.

A resposta obtida por esse método é o número de ciclos (C_t) necessários para amplificação total das amostras, usando *primers* que amplificam cada gene de interesse. A Figura 2 representa a placa comercial onde a amplificação do material genético dos pacientes é feita. Cada gene é amplificado em triplicata, ou seja, o *primer* para amplificação de cada gene alvo é inserido em três “caselas” com material genético do mesmo paciente. A triplicata é utilizada para controle da variabilidade. Para mais detalhes sobre o método, ver LIVAK and SCHMITTGEN (2001).

	1	2	3	4	5	6	7	8	9	10	11	12
Paciente 1	A	Norm.1	Norm.1	Norm.1	Norm.2	Norm.2	Norm.2	Gene1	Gene1	Gene1	Gene2	Gene2
	B	Gene3	Gene3	Gene3	Gene4	Gene4	Gene4					
	C											
	D	Calib.	Calib.	Calib.								
	E											
	F											
	G											
	H											

Figura 2: Esquema de placa comercial utilizada para amplificação de material genético via PCR *real-time*.

A partir do número de ciclos C_t (expressão), é feita a seguinte transformação:

- $\Delta C_{t,q} = [\text{Média dos ciclos do Gene Alvo} - \text{Média dos ciclos do Normalizador}]$ para a amostra q; (I)
- $\Delta\Delta C_t = \Delta C_t - \Delta C_{t,cb}$ é a diferença na leitura dos ciclos do gene alvo entre amostras q e de um calibrador; (II)
- Expressão Relativa = $2^{-\Delta\Delta C_t}$ (III)

Normalizador é um gene endógeno, cuja expressão é estável sob diversas situações. Neste estudo, foram utilizados dois normalizadores: GUSB e β -actina. Calibrador é um tecido normal (sem tumor) comercial de referência. Ambos, normalizador e calibrador, são utilizados para controle dos dados.

Em (I) e (II), a média se refere à média das triplicatas avaliadas em cada indivíduo. No caso do normalizador, se refere, ainda, à média entre os dois normalizadores utilizados.

O valor da expressão relativa, $2^{-\Delta\Delta C_t}$, indica a expressão normalizada de um gene alvo no tecido tumoral relativo a um tecido normal. Essa medida é usada para aferir a expressão de um gene em uma amostra de tecido, e sua natureza multiplicativa é decorrente do processo exponencial que descreve a reação PCR (LIVAK and SCHMITTGEN, 2001).

4. Descrição das Variáveis

4.1 Variáveis Clínicas e Demográficas

- **Grupo:** variável categórica que indica a faixa etária à qual o paciente pertence. Apresenta três níveis:
 - Pediátrico: 0 a 5 anos de idade
 - Adolescente: 5 a 18 anos de idade
 - Adulto: acima de 18 anos de idade
- **Idade:** refere-se à idade, em anos, em que o paciente foi diagnosticado com tumor adrenocortical.
- **Sexo:** variável indicadora para o sexo do paciente:
 - Masculino;
 - Feminino.

- **Diagnóstico:** variável categórica que traz informação sobre a gravidade do tumor. Apresenta dois níveis:
 - Adenoma, tumor classificado como benigno;
 - Carcinoma, tumor classificado como maligno.
- **Evolução:** variável indicadora para ocorrência de óbito do paciente devido ao tumor adrenocortical. Apresenta os níveis:
 - Não houve óbito decorrente do tumor até o final do acompanhamento médico;
 - Houve óbito decorrente do tumor durante o acompanhamento médico.
- **Escore de Weiss:** variável quantitativa discreta, que assume valores de 0 a 9, de acordo com os critérios de Weiss. Os nove critérios de Weiss são critérios anatopatológicos, que estão descritos na Tabela A.5. A cada critério apresentado pelo paciente, soma-se um ponto ao seu escore de Weiss. O diagnóstico do paciente pode ser feito com base no escore de Weiss: de 0 a 3 pontos, diagnostica-se o paciente com adenoma; acima de 3 pontos, diagnostica-se o paciente com carcinoma.
- **Data de Nascimento:** expresso em dia, mês e ano (dd/mm/aaaa).
- **Metástase:** variável indicadora para a ocorrência de metástase no paciente. Apresenta as categorias:
 - não houve metástase;
 - houve metástase.
- **DataDiag:** expresso em dia, mês e ano (dd/mm/aaaa). Indica a data em que o paciente foi diagnosticado com tumor adrenocortical.
- **DataFim:** expresso em dia, mês e ano (dd/mm/aaaa). Indica a data em que o paciente encerrou acompanhamento médico, seja por falecimento, pelo fim do tratamento do tumor, ou outro motivo qualquer.
- **Tempo:** indica o tempo, em anos, que o paciente esteve sob acompanhamento médico para tratamento do tumor, ou seja, é a diferença entre a data de diagnóstico do paciente e a data do fim do acompanhamento.

4.2 Variáveis Genéticas

- **SF1:** expressão relativa do gene SF1 no material genético de cada paciente.
- **POD1:** expressão relativa do gene POD1 no material genético de cada paciente.
- **BUB1B:** expressão relativa do gene BUB1B no material genético de cada paciente.
- **PINK1:** expressão relativa do gene PINK1 no material genético de cada paciente.
- **P53:** variável categórica que se refere à incidência de mutação no gene P53 em cada paciente (apenas para os grupos Pediátrico e Adolescente). Apresenta dois níveis:
 - Sim, caso haja ocorrência de mutação no gene P53;
 - Não, caso não haja mutação nesse gene.
- **IGF1R:** expressão relativa do gene IGF1R no material genético de cada paciente.
- **IGF2:** expressão relativa do gene IGF2 no material genético de cada paciente.
- **β-catenina anormal:** variável categórica que se refere à presença de anomalia na proteína β-catenina em cada paciente. Apresenta dois níveis:
 - Presente, caso haja anomalia na proteína β-catenina no paciente;
 - Ausente, caso não haja anomalia nessa proteína.

5. Análise Descritiva

Na análise descritiva preliminar dos dados, são apresentadas tabelas de medidas resumo (médias, desvios-padrão, máximos, mínimos e tamanhos amostrais) e gráficos de dispersão (*boxplots* e *dotplots*) e de perfis (médias e erros-padrão) para variáveis quantitativas. Tabelas de frequências e gráficos de barras são apresentados para descrever as variáveis categóricas. Para mais detalhes, ver BUSSAB e MORETTIN, 2013. São também apresentadas curvas

de Kaplan-Meier para descrever os tempos até a ocorrência de eventos de interesse (ver COLOSIMO e GIOLO, 2006).

5.1 Variáveis Clínicas e Demográficas

Analizando-se o diagnóstico dos pacientes (Tabela A.6), se adenomas ou carcinomas, divididos em cada grupo (Gráfico B.1), nota-se que entre os pacientes pediátricos e adultos, é maior a ocorrência de adenomas (77 e 56%, respectivamente). Já entre os adolescentes, a ocorrência de carcinomas (62%) é mais frequente. No contexto de razões de chances, tem-se que a chance de um adolescente ter diagnóstico de carcinoma é 5.63 vezes maior que a chance de um pediátrico. Já a chance de um adulto ter diagnóstico de carcinoma é 2.68 vezes maior que a chance de um pediátrico. Finalmente, a chance de um adolescente ter diagnóstico de carcinoma é 2.10 vezes maior que a chance de um adulto.

Quanto ao diagnóstico dos pacientes dentro de cada grupo divididos por sexo (Tabela A.10 e Gráfico B.2), observa-se que para pediátricos, a ocorrência de adenomas é maior que a de carcinomas em ambos os sexos (83 e 66%, respectivamente). Já para os adolescentes, a ocorrência de carcinomas é maior para o sexo feminino (66%), e não há diferenças de diagnóstico para o sexo masculino. No caso dos adultos, tem-se predominância de adenomas para o sexo feminino (60%) e de carcinomas para o sexo masculino (58%). No contexto de razões de chances, tem-se que para os pediátricos, a chance de um indivíduo do sexo masculino ter diagnóstico de carcinoma é 2.38 vezes maior que a chance de um paciente do sexo feminino ter o mesmo diagnóstico. No grupo dos adolescentes, a chance de uma paciente ter diagnóstico de carcinoma é o dobro da chance de um paciente do sexo masculino. Já para os adultos, a chance de um indivíduo do sexo masculino ter diagnóstico de carcinoma é 2.06 vezes maior que a chance de uma paciente ter o mesmo diagnóstico.

A análise do diagnóstico dos pacientes por grupo e Evolução (Tabela A.11 e Gráfico B.3) permite observar que, dentre os pacientes que não vieram a óbito em decorrência do tumor, há predominância de diagnósticos de adenoma para todos os grupos (75, 64 e 90%, respectivamente). Já entre os

casos em que houve morte decorrente do tumor, há totalidade de pacientes com carcinoma.

Já o estudo do diagnóstico dos pacientes em cada grupo, pela ocorrência de metástase (Tabela A.12 e Gráfico B.4), mostra que os pacientes que sofreram metástase são, em sua totalidade, pacientes com carcinoma. Nos pediátricos e adultos que não sofreram metástase há predominância de pacientes com adenoma (96 e 69%). Já para os adolescentes que não sofreram metástase, a predominância é de pacientes com carcinoma (60%).

O Gráfico B.5 apresenta a dispersão da idade de diagnóstico dos pacientes nos diferentes grupos. Para categorias combinadas (Diagnóstico-Grupo) com tamanho amostral menor que 10, optou-se por utilizar o *dotplot*, e para amostras maiores que 10, *boxplot*. Observa-se que, dentro de cada grupo, a distribuição da idade de diagnóstico não parece ter relação com o diagnóstico. Já o Gráfico B.6 permite observar que, dentro de cada grupo, a distribuição das idades de diagnóstico é homogênea.

Sobre a dispersão do escore de Weiss dos pacientes (Gráfico B.7), nota-se que, em todos os grupos, apesar das medianas do escore dos pacientes com carcinoma se mostrarem superiores às dos pacientes com adenomas, existem pacientes com escores acima de três que foram diagnosticados com adenoma (notadamente nos pediátricos). Este fato parece demonstrar inconsistência, visto que a classificação do tumor em adenoma utilizando o critério de Weiss é para escores menores ou iguais a três.

Quanto à dispersão do tempo de acompanhamento médico dos pacientes (Gráfico B.8), nota-se, em todos os grupos, menor mediana de tempo de acompanhamento para pacientes diagnosticados com carcinoma, sendo mais evidente entre os pediátricos e adolescentes. Além disso, os pacientes adultos diagnosticados com adenoma apresentam menor tempo de acompanhamento mediano em relação aos pediátricos e adolescentes com mesmo diagnóstico.

Ainda, no Gráfico B.9, vemos a curva de sobrevivência, com estimador de Kaplan Meier, para cada grupo (independente do diagnóstico). É possível observar que é menor a sobrevivência para o grupo de pediátricos, seguido dos adultos. Os adolescentes apresentam maiores probabilidades de

sobrevida. Nesse gráfico, o evento considerado é o óbito do paciente decorrente do tumor. Devido a isso, existem muitas censuras, já que a grande parte dos pacientes foi diagnosticada com adenoma, não resultando em óbito.

As funções de sobrevida para pacientes diagnosticados com carcinoma, segundo Grupo, são apresentadas no Gráfico B.10. Nota-se que a sobrevida é maior entre os pacientes adultos, seguido pelos pediátricos e por último, adolescentes. No entanto, deve-se considerar que a maior parte das observações é censurada, isto é, não há ocorrência de óbito até a data da última consulta.

5.2 Variáveis Genéticas

O comportamento dos genes de interesse, com base nas médias e desvios-padrão da expressão relativa, pode ser visto nos Gráficos B.11 a B.18. Devido à escala da medida de expressão genética, os desvios abaixo da média foram truncados em zero nos gráficos citados. Ainda, as Tabelas A.7 a A.9 apresentam as medidas resumo dessas variáveis.

O gene SF1 apresenta média de expressão relativa nos carcinomas maior que nos adenomas para os grupos Adolescentes (2.32 e 0.84) e Adultos (3.48 e 1.93). Já para os pediátricos, a média da expressão relativa dos pacientes diagnosticados com carcinomas é inferior à média de expressão relativa nos adenomas (0.68 e 2.23). Tal fato reflete o efeito da faixa etária no padrão da expressão deste gene. Contudo, destaca-se a elevada heterogeneidade da expressão gênica entre indivíduos do mesmo grupo, que pode ser ressaltada pelos valores dos coeficientes de variação (CV) apresentados na Tabela A.15. Destaca-se, por exemplo, o CV igual a 187.18 em adenomas pediátricos.

Nos casos dos genes POD1 e BUB1B, nota-se uma tendência que se mantém entre os grupos, já que para todas as faixas etárias, as médias de expressão relativa destes genes em pacientes diagnosticados com carcinoma (0.91, 1.12 e 1.78 para expressão relativa de POD1 em pediátricos, adolescentes e adultos, respectivamente, e 20.45, 8.89 e 8.2 para BUB1B) são superiores às médias de expressão relativa de pacientes com diagnóstico de adenoma (0.62, 0.52 e 1.28 para expressão relativa de POD1 em pediátricos,

adolescentes e adultos, respectivamente, e 3.03, 3.04 e 0.59 para BUB1B). Em algumas categorias combinadas (Diagnóstico-Grupo) essas diferenças de médias são pouco notáveis. Também há casos com grande variabilidade nos valores das expressões relativas, como em carcinomas adultos (CV igual a 458.17 e 140.87 para POD1 e BUB1B, respectivamente), por exemplo.

O gene PINK1 apresenta a maior heterogeneidade entre os grupos. Entre os pediátricos, a média e dispersão das expressões relativas entre os pacientes diagnosticados com adenoma e carcinoma é praticamente igual (médias iguais a 0.83 e 0.48, respectivamente). No caso dos adolescentes, a média e desvio-padrão para os pacientes com carcinoma são bastante superiores às dos pacientes com adenoma (médias iguais a 3.97 e 1.01, respectivamente). Finalmente, entre os adultos a situação anterior se inverte: a média e desvio-padrão para os pacientes com carcinoma são inferiores às dos pacientes com adenoma (médias iguais a 0.65 e 1.1, respectivamente). Os coeficientes de variação para adenomas e carcinomas nos diferentes grupos também são bastante elevados (Tabela A.15).

Nos *boxplots* e *dotplots* das expressões relativas dos genes alvo citados acima, pode-se verificar alguns valores atípicos (*outliers*) dentre as categorias combinadas. É importante salientar que estes *outliers* não configuram erros de medição, uma vez que o uso de normalizadores e calibradores no processo de PCR *real-time* visa justamente anular tais fontes de variabilidade.

Os Gráficos B.19 e B.20 representam as médias e desvios-padrão das expressões relativas dos genes IGF1R e IGF2 (normalizado e calibrado), respectivamente, dentro de cada grupo e para cada diagnóstico. No caso do gene IGF1R, nota-se que a média da expressão relativa desse gene é superior em pacientes com carcinomas, em relação àqueles com adenomas, para os grupos pediátrico e adolescente. Já para o grupo de adultos, diferenças entre a expressão relativa de IGF1R entre adenomas e carcinomas não são muito notáveis, ainda que a média da expressão relativa para carcinomas seja ligeiramente inferior à de adenomas.

Já para o gene IGF2, as diferenças entre pacientes diagnosticados com adenoma e carcinoma são mais evidentes para os grupos de adolescentes e adultos, com a média da expressão relativa de IGF2 maior nos pacientes com carcinoma destes grupos. O grupo de pediátricos não apresenta grandes

discrepâncias entre as médias da expressão relativa de IGF2, apesar dos pacientes com carcinoma apresentarem média ligeiramente menor.

Em ambos os gráficos, B.19 e B.20, temos categorias combinadas (Diagnóstico-Grupo) com número reduzido de observações, ou seja, o peso de cada observação é grande no cálculo da média e desvio padrão.

A análise do diagnóstico dos pacientes, em adenomas ou carcinomas, por grupo e ocorrência de mutação no gene P53 (Tabela A.13) é apresentada no Gráfico B.21. No grupo pediátrico, a predominância de pacientes com diagnóstico de adenoma é superior ao de carcinoma, tanto para pacientes que sofreram mutação no gene P53, quanto para os que não sofreram (74 e 83%, respectivamente). Entretanto, para os adolescentes, a predominância é de carcinomas para os pacientes com mutação no gene P53 (72%), e de adenomas para os pacientes sem a presença de tal mutação (75%). No contexto das razões de chances, tem-se que a chance de um paciente que apresenta mutação no gene P53 ser diagnosticado como maligno é 1.75 vezes maior que a chance de um paciente que não apresenta mutação nesse gene, no grupo pediátrico, e 7.5 vezes maior no grupo adolescentes.

O comportamento do diagnóstico dos pacientes em relação à expressão anômala da proteína β-catenina por grupo pode ser visto na Tabela A.14, e graficamente no Gráfico B.22. No grupo pediátrico, tanto na ausência como na presença de anomalia na proteína β-catenina, há predominância de pacientes diagnosticados com adenoma (80 e 100%, respectivamente). No grupo de adolescentes, nenhum paciente apresentou anomalia na β-catenina. Finalmente, entre os adultos, na presença de anomalia, prevalece o diagnóstico de adenoma (60%), e na ausência de anomalia, há igualdade na proporção de pacientes com cada diagnóstico.

O estudo das expressões combinadas dos pares de genes (Gráficos B.23 a B.28) foi feito sob duas abordagens: cálculo de correlações entre as expressões relativas de cada par de genes alvo (Tabelas A.01 a A.04), ou medindo as diferenças de ciclos $\Delta C_{tGene1} - \Delta C_{tGene2}$ (expressões normalizadas, mas não calibradas), como pode ser visto nas Tabelas A.15 a A.17.

O cálculo de correlações, métodos de Pearson e Spearman (mais robusto), dentro de cada grupo, apresentou somente um par de genes com correlação acima de 0.7 (0.737 pelo método de Pearson, entre BUB1B e SF1

no Grupo dos adolescentes) e poucos pares de genes com correlações moderadas (0.54 pelo método de Spearman, entre SF1 e POD1 no Grupo dos pediátricos, por exemplo), fato considerado inesperado com base em publicação de estudo anterior (FRANÇA et al, 2003). Nos gráficos das diferenças (Gráficos B.23 a B.28), mais uma vez, o que chama atenção é a heterogeneidade entre os grupos de Idade. Por exemplo, no Gráfico B.23, temos a medida $\Delta C_{tPOD1} - \Delta C_{tBUB1B}$ (expressão combinada de POD1 e BUB1B). Para os grupos Pediátrico e Adolescente, temos que, em média, as medidas de expressão de BUB1B são superiores às de POD1 (médias menores que zero). Já para o grupo Adulto, o contrário ocorre.

6. Análise Inferencial

Nesta seção serão apresentadas análises inferenciais buscando atender aos objetivos do estudo. Devido ao caráter binário da variável resposta de interesse (Diagnóstico: adenoma ou carcinoma), optou-se, primeiramente, por ajustar modelos de regressão logística aos dados (Hosmer and Lemeshow, 2000).

Uma pré-seleção de variáveis preditoras foi feita da seguinte maneira: ajustou-se um modelo de regressão logística para cada variável do estudo na presença do fator Grupo e da interação entre eles. Ressalta-se que o efeito de interação entre grupo e expressão gênica é o de maior interesse, pois permite analisar se o padrão de expressão é diferente relativamente à faixa etária. Resultados destes ajustes podem ser vistos na Tabela A.19. Utilizando um nível de significância (α) de 30%, os modelos incluindo as interações entre Grupo e Sexo, Grupo e SF1, Grupo e BUB1B e Grupo e PINK1 foram pré-selecionados.

Para as variáveis que não mostraram significância (a um nível α de 30%) da interação com Grupo no passo anterior, foram ajustados modelos reduzidos sem as interações (somente com efeitos principais). Mantido o mesmo α , nenhuma das variáveis consideradas neste passo foi selecionada (detalhes na Tabela A.20).

Como era de interesse observar a relação das expressões relativas entre os genes alvo, o terceiro passo de pré-seleção considerou o ajuste de

modelos logísticos para as interações, duas a duas entre estes genes (sempre juntamente com Grupo). Como pode ser visto na Tabela A.21, as interações entre SF1 e POD1 (valor-p igual a 0.08), POD1 e BUB1B (valor-p igual a 0.25) e entre POD1 e PINK1 (valor-p igual a 0.23) mostraram-se estatisticamente significantes (a um nível α de 30%). Como em presença de outros genes POD1 mostrou significância, decidiu-se avaliar a significância da interação de POD1 com Grupo na presença destes genes (Tabela A.22). Assim, na presença de SF1 e BUB1B, a interação de POD1 com Grupo mostrou-se significante.

Finalmente, nos modelos em que as interações entre os pares de genes não se mostraram significantes, estas foram removidas, e ajustaram-se novos modelos para observar a significância de um gene na presença do outro (Tabela A.23).

Neste processo de pré-seleção de variáveis, notou-se não fazer sentido considerar como variáveis preditoras Weiss (já que a informação da resposta Diagnóstico é determinada, em geral, pelo escore de Weiss) e Idade (pois a classificação de um paciente em cada Grupo de faixa etária é baseado na Idade de diagnóstico, ou seja, estar-se-ia considerando a mesma informação duas vezes.).

Após o procedimento de pré-seleção de variáveis citado acima, o modelo “completo” foi ajustado incluindo, simultaneamente, as seguintes variáveis explicativas: Grupo, Sexo (em interação com Grupo), SF1 (em interação com Grupo), BUB1B (em interação com Grupo), POD1(em interação com Grupo, na presença de SF1 ou BUB1B), PINK1 (em interação com Grupo), e as interações entre SF1 e POD1, POD1 e BUB1B e POD1 e PINK1.

A redução do modelo completo foi feita manualmente, com base nos valores-p de cada variável, e adotando um novo nível de significância de (α) 10%. Quatro modelos mostraram-se significantes.

No primeiro modelo, cujo ajuste pode ser visto em detalhe na Tabela C.1, o Diagnóstico do paciente é predito pelas variáveis Grupo, SF1, BUB1B e POD1. Aparecem no modelo interações significantes entre o Grupo dos adultos e BUB1B (valor-p 0.02), e o Grupo dos adultos e POD1 (valor-p 0.01). Isto é, em comparação com o Grupo de referência (pediátricos), a expressão relativa de BUB1B é significativamente diferente no Grupo dos adultos. Este comportamento se repete para a expressão relativa de POD1. É importante

notar que o sinal do efeito estimado de interação entre o Grupo dos adultos e BUB1B é negativo, e positivo para a interação entre Grupo adulto e POD1. Em outras palavras, ao aumentar em uma unidade a expressão relativa de BUB1B de um paciente do adulto, a probabilidade deste paciente ser diagnosticado com carcinoma diminui, mantendo fixas as demais variáveis do modelo. O oposto ocorre ao aumentar em uma unidade a expressão relativa de POD1 de um paciente adulto: a probabilidade deste paciente ser diagnosticado com carcinoma aumenta, mantendo fixas as demais variáveis do modelo.

Ainda no modelo 1, mostra-se significante a interação entre os genes SF1 e POD1. O sinal negativo desta interação, na presença de sinais positivos para ambos os efeitos principais correspondentes, indica que a cada mudança de uma unidade na expressão destes genes há uma diminuição esperada que se multiplica na chance do diagnóstico de carcinoma. Na análise de diagnóstico do ajuste do modelo 1 (Gráfico C.1), notamos pontos de influência e alavanca, que, apesar de bastante evidentes, não indicaram nenhum valor clinicamente atípico, após averiguação. O poder preditivo do modelo é de aproximadamente 85%, visto na Curva ROC, indicando bom ajuste.

No modelo 2 (detalhes na Tabela C.2), o Diagnóstico é predito pelas variáveis Grupo, BUB1B e PINK1. Nota-se que os efeitos desses genes são inversos, já que, a estimativa do efeito de BUB1B é positivo e a estimativa do efeito de PINK1 é negativo. De acordo com o ajuste, o aumento de uma unidade na expressão relativa de BUB1B conduz a um aumento esperado na chance do diagnóstico do paciente ser carcinoma de 25%, mantidas fixas as categorias de grupo e expressão relativa de PINK1. O oposto acontece com PINK1, isto é, a cada unidade acrescida na expressão relativa de PINK1, a chance do diagnóstico ser carcinoma diminui 17%, mantidas fixas as categorias de grupo e expressão relativa de BUB1B. Os gráficos de análise de diagnóstico do ajuste do modelo 1 (Gráfico C.2) indicam bom ajuste (destaque para curva ROC, com poder preditivo de 81%). Nota-se, contudo, a presença de pontos influentes e de alavanca, em particular os dados do paciente 867a. Uma revisão destes dados não indicou nenhum valor clinicamente atípico.

Considerando o terceiro modelo selecionado (ajuste apresentado na Tabela C.3), o Diagnóstico do paciente é predito em função das variáveis Grupo e da interação entre SF1 e POD1 (na presença de efeitos principais

desses genes). O sinal negativo desta interação, na presença de sinais positivos para ambos os efeitos principais correspondentes, indica que a cada mudança de uma unidade na expressão destes genes há uma diminuição esperada que se multiplica na chance do diagnóstico de carcinoma. Na análise de diagnóstico do ajuste do modelo 3 (Gráfico C.3), é possível notar o menor poder preditivo entre os quatro modelos apresentados (aproximadamente 66%, como se vê na Curva ROC). Ainda assim, apresenta resíduos bem comportados, e pontos de alavanca e influência que não indicaram nenhum valor clinicamente atípico.

Dado o reduzido tamanho amostral para algumas variáveis do estudo, modelos de regressão logística penalizados (KOSMIDIS and FIRTH, 2009) também foram ajustados para as variáveis p53, IGF1R, IGF2 e β-catenina, utilizando o pacote brglm do R (Tabela A.24). Vê-se que as variáveis IGF1R e IGF2 são significantes na predição do Diagnóstico. Dessa forma, ajustou-se um modelo logístico contendo essas duas variáveis genéticas, bem como Grupo, e as interações entre elas. Após a redução do modelo citado, o modelo 4 (detalhes na Tabela C.4) foi obtido, somente com os efeitos principais. Vê-se que ambos os efeitos de IGF1R e IGF2 são positivos no ajuste da probabilidade do Diagnóstico, ou seja, a cada aumento de uma unidade na expressão relativa de IGF1R, a chance esperada do diagnóstico ser carcinoma aumenta em 45%, mantidas fixas as categorias de Grupo e expressão relativa de IGF2. Da mesma forma, a cada aumento de uma unidade na expressão relativa de IGF2, a chance esperada do diagnóstico do paciente ser carcinoma cresce em 1.9%, mantendo fixas as demais variáveis do modelo. Na análise de diagnóstico do ajuste do modelo 4, Gráfico C.4, notamos um bom ajuste (80% de poder preditivo, segundo a curva ROC). Também se destaca que o paciente A10 aparece como ponto influente no ajuste, mas uma conferência de seus dados não evidencia nenhuma observação clinicamente atípica.

É apresentada na Tabela C.5 um comparativo entre os valores observados e ajustados segundo os quatro modelos finais para o Diagnóstico de pacientes. Com exceção do modelo 3, os demais apresentam bons ajustes relativamente aos valores observados. Na comparação dos modelos, que pode ser vista na Tabela C.6, nota-se que, pelo critério *Deviance*, o modelo 3 não está bem ajustado.

7. Conclusões

A análise descritiva dos dados indicou algumas diferenças entre os diagnósticos de adenoma e carcinoma relativamente ao grupo de faixa etária ao qual o paciente pertence, as quais se mostraram significantes na análise inferencial, uma vez que os quatro modelos ajustados apresentam efeito significante para os Grupos adolescente e adulto em comparação com o pediátrico, que é o Grupo de referência. Tal fato corrobora a classificação de faixa etária proposta pioneiramente neste estudo.

Por meio do ajuste de modelos de regressão logística pôde-se confirmar o preponderante poder preditivo do escore de Weiss relativamente às demais variáveis, o que já era esperado decorrente de como o diagnóstico é realizado. Na ausência desta informação, foi possível identificar variáveis genéticas com efeito preditivo significante, como SF1, BUB1B, POD1, PINK1 (modelos 1 e 2), IGF1R e IGF2 (modelo 4).

APÊNDICE A

TABELAS

Tabela A.1: Correlações entre os genes SF1, POD1, BUB1 e PINK1 (Pearson na diagonal superior e Spearman na diagonal inferior).

	Geral			
	SF1	POD1	BUB1B	PINK1
SF1	1	0.017	-0.039	-0.019
POD1	0.347	1	0.144	-0.014
BUB1B	0.269	0.059	1	0.160
PINK1	0.275	0.167	0.160	1

Tabela A.2: Correlações entre os genes SF1, POD1, BUB1 e PINK1 para o grupo Pediátrico (Pearson na diagonal superior e Spearman na diagonal inferior).

	Pediátrico			
	SF1	POD1	BUB1B	PINK1
SF1	1	0.247	-0.071	-0.026
POD1	0.537	1	0.019	-0.163
BUB1B	0.41	0.234	1	-0.037
PINK1	0.13	0.055	0.155	1

Tabela A.3: Correlações entre os genes SF1, POD1, BUB1 e PINK1 para o grupo Adolescente (Pearson na diagonal superior e Spearman na diagonal inferior).

	Adolescente			
	SF1	POD1	BUB1B	PINK1
SF1	1	0.019	0.737	0.084
POD1	0.486	1	0.033	-0.089
BUB1B	0.525	0.288	1	0.534
PINK1	0.179	-0.226	0.029	1

Tabela A.4: Correlações entre os genes SF1, POD1, BUB1 e PINK1 para o grupo Adulto (Pearson na diagonal superior e Spearman na diagonal inferior).

	Adulto			
	SF1	POD1	BUB1B	PINK1
SF1	1	-0.011	-0.039	0.009
POD1	0.295	1	0.217	0.019
BUB1B	0.209	0.065	1	-0.003
PINK1	0.357	0.325	0.272	1

Tabela A.5: Critérios histológicos de Weiss para Carcinoma Adrenocortical.

Critério	Achados
1	Grau nuclear III ou IV
2	Índice mitótico > 5/50 campo de grande aumento
3	Presença de mitoses atípicas
4	Células claras constituindo 25 % ou menos do tumor
5	Arquitetura difusa
6	Necrose celular no tumor
7	Invasão de estruturas venosas
8	Invasão sinusoidal
9	Invasão sinusoidal

Tabela A.6: Frequências de Diagnóstico

Diagnóstico	Frequência	%
Adenoma	77	59
Carcinoma	53	41

Tabela A.7: Medidas Resumo para Variáveis Quantitativas (Pediátricos).

Pediátrico											
Variável	Adenoma						Carcinoma				
	Média	DP	Mín	Máx	N	Média	DP	Mín	Máx	N	
Idade	1.84	0.77	0.43	3.33	27	2.44	0.93	0.90	3.83	8	
SF1	2.23	4.17	0.02	21.00	27	0.68	0.52	0.25	1.65	8	
POD1	0.62	1.61	0.01	8.45	27	0.91	1.03	0.14	2.85	6	
BUB1B	3.03	3.21	0.14	11.12	27	20.45	31.38	0.12	76.64	8	
PINK1	0.83	0.82	0.01	3.44	26	0.48	0.31	0.01	0.96	8	
IGF1R	2.94	1.35	1.16	5.27	10	8.05	3.52	4.04	10.62	3	
IGF2	34.17	9.98	22.89	50.40	10	31.64	12.59	18.60	43.72	3	
Weiss	4.31	2.28	0.00	8.00	26	6.38	1.30	4.00	8.00	8	
Tempo	8.35	5.26	0.08	19.17	27	4.81	5.80	0.45	16.75	8	

Tabela A.8: Medidas Resumo para Variáveis Quantitativas (Adolescentes).

Adolescente											
Variável	Adenoma						Carcinoma				
	Media	DP	Mín	Max	N	Media	DP	Mín	Max	N	
Idade	9.50	4.59	5.04	16.90	6	13.25	4.42	5.54	17.70	10	
SF1	0.84	0.48	0.27	1.54	6	2.32	4.15	0.05	13.59	10	
POD1	0.52	0.52	0.06	1.47	6	1.12	2.73	0.03	8.37	9	

BUB1B	3.04	6.15	0.17	15.56	6	8.89	13.63	0.06	35.49	9
PINK1	1.01	1.12	0.05	3.16	6	3.97	8.80	0.23	27.34	9
IGF1R	1.81	1.55	0.46	3.51	3	4.62	4.66	1.32	9.95	3
IGF2	33.43	41.41	0.27	79.85	3	106.22	94.27	4.66	190.92	3
Weiss	2.17	1.47	1.00	5.00	6	6.70	1.64	4.00	9.00	10
Tempo	7.39	5.53	0.00	15.16	6	2.54	3.04	0.67	10.42	10

Tabela A.9: Medidas Resumo para Variáveis Quantitativas (Adultos).

Adulto											
Variável	Adenoma						Carcinoma				
	Média	DP	Mín	Máx	N	Média	DP	Mín	Máx	N	
Idade	40.59	13.95	18.00	74.70	44	42.23	16.77	19.00	83.00	35	
SF1	1.93	3.11	0.04	10.67	44	3.48	5.96	0.01	20.46	35	
POD1	1.28	2.19	0.04	11.10	44	1.78	8.14	0.00	48.47	35	
BUB1B	0.59	1.07	0.04	6.18	42	8.20	11.56	0.10	43.48	34	
PINK1	1.10	1.45	0.04	7.88	42	0.65	0.69	0.01	3.05	34	
IGF1R	1.39	1.02	0.00	3.34	16	1.17	0.88	0.00	2.94	14	
IGF2	10.96	28.55	0.00	112.49	16	73.20	73.13	0.00	228.49	14	
Weiss	0.72	0.93	0.00	4.00	43	5.74	1.94	3.00	9.00	34	
Tempo	4.95	5.26	0.00	19.09	44	4.18	5.89	0.00	31.35	35	

Tabela A.10: Frequências da variável Sexo por Grupo e Diagnóstico.

Sexo	Pediátrico		Adolescente		Adulto	
	Adenoma	Carcinoma	Adenoma	Carcinoma	Adenoma	Carcinoma
Feminino	19	4	4	8	36	24
Masculino	8	4	2	2	8	11

Tabela A.11: Frequências da variável Evolução por Grupo e Diagnóstico.

Evolução	Pediátrico		Adolescente		Adulto	
	Adenoma	Carcinoma	Adenoma	Carcinoma	Adenoma	Carcinoma
Não houve óbito decorrente do tumor	27	3	5	2	32	18
Houve óbito decorrente do tumor	0	5	0	7	0	12

Tabela A.12: Frequências de Metástase por Grupo e Diagnóstico.

Metástase	Pediátrico		Adolescente		Adulto	
	Adenoma	Carcinoma	Adenoma	Carcinoma	Adenoma	Carcinoma
Não	26	1	6	3	32	14

Ocorreu	0	7	0	5	0	16
---------	---	---	---	---	---	----

Tabela A.13: Frequências de Mutação no gene P53 por Grupo e Diagnóstico.

Mutação P53	Pediátrico		Adolescente	
	Adenoma	Carcinoma	Adenoma	Carcinoma
Não	5	1	3	1
Sim	20	7	2	5

Tabela A.14: Frequências de proteína β-catenina anômala por Grupo e Diagnóstico.

β- Catenina anômala	Pediátrico		Adolescente		Adulto	
	Adenoma	Carcinoma	Adenoma	Carcinoma	Adenoma	Carcinoma
Ausente	8	2	2	2	5	5
Presente	1	0	0	0	3	2

Tabela A.15: Coeficientes de Variação.

Variável	Pediátrico		Adolescente		Adulto	
	Adenoma	Carcinoma	Adenoma	Carcinoma	Adenoma	Carcinoma
Coeficiente de variação						
Idade	42.04	38.07	48.27	33.37	34.38	39.72
SF1	187.18	76.15	57.54	178.76	161.53	171.48
POD1	261.20	112.51	100.38	243.20	170.96	458.17
BUB1B	105.87	153.48	202.49	153.31	180.14	140.87
PINK1	99.46	64.58	111.02	221.43	131.25	106.51
IGF1R	45.82	43.71	85.68	100.97	73.28	75.15
IGF2	29.22	39.79	123.88	88.75	260.60	99.91
Weiss	52.84	20.43	67.94	24.42	129.58	33.88
Tempo	62.96	120.58	74.73	119.56	106.30	140.99

Tabela A.16: Medidas Resumo para Expressões Combinadas (Pediátrico).

Pediátrico										
Variável	Adenoma					Carcinoma				
	Média	DP	Mín	Máx	N	Média	DP	Mín	Máx	N
BUB-PINK	-1.22	2.34	-6.56	4.07	26	-2.43	3.74	-6.86	3.01	8
POD-BUB	2.60	2.60	-4.66	6.68	27	1.10	4.24	-4.75	5.75	8

POD-PINK	1.43	3.35	-6.88	6.24	27	-1.33	2.79	-5.57	2.87	8
SF1-POD	-3.93	2.07	-7.82	0.16	27	-1.20	2.15	-4.25	2.15	8
SF1-BUB	-1.33	2.29	-5.64	3.19	27	-0.10	3.12	-4.19	4.12	8
SF1-PINK	-2.50	2.17	-7.33	1.89	27	-2.53	1.83	-6.23	-1.10	8

Tabela A.17: Medidas Resumo para Expressões Combinadas (Adolescentes).

Adolescente											
Variável	Adenoma					Carcinoma					N
	Média	DP	Mín	Máx	N	Média	DP	Mín	Máx	N	
BUB-PINK	2.03	6.63	-2.82	15.30	6	4.43	10.89	-1.45	34.82	10	
POD-BUB	0.84	6.43	-11.93	5.63	6	-3.46	14.46	-33.55	7.50	10	
POD-PINK	1.28	3.36	-3.68	4.69	6	2.38	3.40	-4.57	6.85	10	
SF1-POD	0.32	0.85	-1.20	1.10	6	1.31	4.88	-6.78	12.99	10	
SF1-BUB	-2.20	6.36	-15.15	1.06	6	-5.68	13.60	-31.15	13.59	10	
SF1-PINK	-0.17	1.23	-2.29	1.49	6	-1.25	9.81	-26.01	13.59	10	

Tabela A.18: Medidas Resumo para Expressões Combinadas (Adultos).

Adulto											
Variável	Adenoma					Carcinoma					N
	Média	DP	Mín	Máx	N	Média	DP	Mín	Máx	N	
BUB-PINK	2.97	2.43	-0.14	12.09	43	-0.52	3.64	-7.53	10.30	35	
POD-BUB	-3.29	2.62	-10.70	1.85	44	1.33	3.98	-9.08	9.20	35	
POD-PINK	-0.39	2.30	-6.54	5.72	44	0.81	3.09	-7.34	5.57	35	
SF1-POD				0					0		
SF1-BUB				0					0		
SF1-PINK				0					0		

Tabela A.19: Modelos de regressão logística para pré-seleção de variáveis com interações.

Variável Resposta: Adenoma (0) e Carcinoma (1)								
Variável Preditoras no Modelo	Parâmetros Ajustados			Coef. Estimado	DP	Estat. Z	p-valor	n
Grupo	Intercepto			-1.22	0.40	-3.02	<0.01	
	GrupoAdolescente			1.73	0.65	2.64	0.01	130
	GrupoAdulto			0.99	0.46	2.14	0.03	
Grupo, Sexo	Intercepto			-1.56	0.55	-2.83	<0.01	
	GrupoAdolescente			2.25	0.82	2.74	0.01	130
	GrupoAdulto			1.15	0.61	1.89	0.06	

	Smasc	0.87	0.82	1.05	0.29	
	GrupoAdolescente*Smasc	-1.56	1.43	-1.09	0.28	
	GrupoAdulto*Smasc	-0.14	0.98	-0.14	0.89	
Grupo, SF1	Intercepto	-0.63	0.56	-1.14	0.25	
	GrupoAdolescente	0.76	0.88	0.86	0.39	
	GrupoAdulto	0.21	0.62	0.34	0.73	
	SF1	-0.56	0.50	-1.12	0.26	126
	GrupoAdolescente*SF1	0.86	0.67	1.28	0.20	
	GrupoAdulto*SF1	0.63	0.50	1.26	0.21	
Grupo, POD1	Intercepto	-1.59	0.50	-3.18	<0.01	
	GrupoAdolescente	1.85	0.76	2.43	0.01	
	GrupoAdulto	1.33	0.55	2.42	0.02	
	POD1	0.11	0.25	0.43	0.66	127
	GrupoAdolescente*POD1	0.07	0.44	0.17	0.86	
	GrupoAdulto*POD1	-0.09	0.26	-0.36	0.71	
Grupo, BUB1B	Intercepto	-1.80	0.54	-3.31	<0.01	
	GrupoAdolescente	1.86	0.82	2.27	0.02	
	GrupoAdulto	0.73	0.63	1.16	0.25	
	BUB1B	0.09	0.07	1.22	0.22	126
	GrupoAdolescente*BUB1B	-0.02	0.10	-0.22	0.83	
	GrupoAdulto*BUB1B	0.38	0.20	1.95	0.05	
Grupo, PINK1	Intercepto	-0.61	0.59	-1.04	0.30	
	GrupoAdolescente	0.74	0.87	0.85	0.40	
	GrupoAdulto	0.77	0.67	1.15	0.25	
	PINK1	-0.90	0.80	-1.13	0.26	
	GrupoAdolescente*PINK1	1.06	0.86	1.24	0.22	
	GrupoAdulto*PINK1	0.46	0.84	0.55	0.58	
Grupo, Weiss	Intercepto	-4.39	1.75	-2.50	0.01	
	GrupoAdolescente	-1.00	3.43	-0.29	0.77	
	GrupoAdulto	-4.25	3.56	-1.20	0.23	
	Weiss	0.59	0.28	2.08	0.04	127
	GrupoAdolescente*Weiss	0.79	0.75	1.05	0.29	
	GrupoAdulto*Weiss	2.65	1.18	2.25	0.02	
Grupo, Evolução	Intercepto	-2.20	0.61	-3.61	<0.01	
	GrupoAdolescente	1.28	1.03	1.24	0.22	
	GrupoAdulto	1.62	0.68	2.40	0.02	
	Evolução	20.76	2917.01	0.01	0.99	111
	GrupoAdolescente*Evolução	-1.28	3819.27	0.00	1.00	
	GrupoAdulto*Evolução	-1.62	3471.94	0.00	1.00	
Grupo, Metástase	Intercepto	-3.26	1.02	-3.20	<0.01	
	GrupoAdolescente	2.57	1.24	2.07	0.04	
	GrupoAdulto	2.43	1.07	2.28	0.02	
	Metástase	22.82	4064.64	0.01	1.00	
	GrupoAdolescente*Metástase	-2.57	5748.26	0.00	1.00	
	GrupoAdulto*Metástase	-2.43	4873.33	0.00	1.00	

Tabela A.20: Modelos de regressão logística para pré-seleção de variáveis sem interações.

Variável Resposta: Adenoma (0) e Carcinoma (1)						
Variáveis Predictoras	Parâmetros Ajustados	Coef. Estimado	DP	Estat. Z	p-valor	n
Grupo, POD1	Intercepto	-1.52	0.45	-3.36	<0.01	127
	GrupoAdolescente	1.91	0.69	2.75	0.01	
	GrupoAdulto	1.26	0.51	2.49	0.01	
	POD1	0.02	0.04	0.51	0.61	
Grupo, Evolução	Intercepto	-2.20	0.61	-3.61	<0.01	111
	GrupoAdolescente	1.28	1.03	1.24	0.22	
	GrupoAdulto	1.62	0.68	2.40	0.02	
	Evolução	19.74	1278.86	0.02	0.99	
Grupo, Metástase	Intercepto	-3.26	1.02	-3.20	<0.01	112
	GrupoAdolescente	2.57	1.24	2.07	0.04	
	GrupoAdulto	2.43	1.07	2.28	0.02	
	Metástase	21.45	1806.51	0.01	0.99	

Tabela A.21: Modelos de regressão logística relacionando os genes alvo e considerando interações.

Variável Resposta: Adenoma (0) e Carcinoma (1)						
Variáveis Predictoras	Parâmetros Ajustados	Coef. Estimado	DP	Estat. Z	p-valor	n
Grupo, SF1*POD1	Intercepto	-1.71	0.48	-3.57	<0.01	127
	GrupoAdolescente	1.98	0.71	2.79	0.01	
	GrupoAdulto	1.30	0.52	2.50	0.01	
	SF1	0.16	0.07	2.21	0.03	
	POD1	0.08	0.09	0.81	0.42	
Grupo, SF1*BUB1B	SF1*POD1	-0.12	0.07	-1.78	0.08	126
	Intercepto	-2.32	0.56	-4.13	<0.01	
	GrupoAdolescente	2.05	0.77	2.66	0.01	
	GrupoAdulto	1.46	0.57	2.58	0.01	
	SF1	0.04	0.06	0.69	0.49	
Grupo, SF1*PINK1	BUB1B	0.18	0.07	2.65	0.01	125
	SF1*BUB1B	0.00	0.02	0.13	0.90	
	Intercepto	-1.33	0.43	-3.10	<0.01	
	GrupoAdolescente	1.62	0.68	2.37	0.02	
Grupo, SF1*PINK1	GrupoAdulto	0.96	0.47	2.05	0.04	125
	SF1	0.09	0.07	1.20	0.23	

	PINK1	0.08	0.13	0.66	0.51	
	SF1*PINK1	-0.06	0.08	-0.81	0.42	
Grupo, POD1*BUB1B	Intercepto	-2.28	0.63	-3.60	<0.01	
	GrupoAdolescente	2.04	0.85	2.40	0.02	
	GrupoAdulto	1.93	0.63	3.06	<0.01	123
	POD1	0.17	0.08	2.21	0.03	
	BUB1B	-0.65	0.46	-1.42	0.16	
	POD1*BUB1B	0.08	0.07	1.15	0.25	
Grupo, POD1*PINK1	Intercepto	-1.53	0.46	-3.31	<0.01	
	GrupoAdolescente	1.82	0.73	2.49	0.01	
	GrupoAdulto	1.33	0.51	2.58	0.01	122
	POD1	0.11	0.09	1.23	0.22	
	PINK1	0.04	0.08	0.53	0.60	
	POD1*PINK1	-0.17	0.14	-1.21	0.23	
Grupo, BUB1B*PINK1	Intercepto	-1.78	0.59	-3.01	<0.01	
	GrupoAdolescente	2.19	0.80	2.72	0.01	
	GrupoAdulto	1.51	0.58	2.63	0.01	125
	BUB1B	0.16	0.09	1.89	0.06	
	PINK1	-0.67	0.35	-1.89	0.06	
	BUB1B*PINK1	0.06	0.07	0.78	0.43	

Tabela A.22: Modelos de regressão logística relacionando os genes alvo, com interações com Grupo.

Variável Resposta: Adenoma (0) e Carcinoma (1)						
Variáveis Preditoras	Parâmetros Ajustados	Coef. Estimado	DP	Estat. Z	p-valor	n
Grupo, SF1, POD1	Intercepto	-1.57	0.74	-2.12	0.03	
	GrupoAdolescente	1.29	1.01	1.28	0.20	
	GrupoAdulto	1.07	0.78	1.38	0.17	
	SF1	-0.80	0.85	-0.94	0.35	
	POD1	1.96	0.97	2.01	0.04	127
	GrupoAdolescente*SF1	1.23	0.95	1.30	0.19	
	GrupoAdulto*SF1	1.11	0.84	1.32	0.19	
	GrupoAdolescente*POD1	-1.47	0.98	-1.50	0.13	
	GrupoAdulto*POD1	-1.90	0.97	-1.96	0.05	
	SF1*POD1	-0.22	0.11	-2.09	0.04	
Grupo, BUB1B, POD1	Intercepto	-2.12	0.77	-2.75	<0.01	
	GrupoAdolescente	2.03	1.02	1.99	0.04	
	GrupoAdulto	1.80	0.86	2.09	0.03	123
	BUB1B	0.07	0.13	0.52	0.60	
	POD1	0.08	0.29	0.30	0.76	
	GrupoAdolescente*BUB1B	-0.02	0.14	-0.16	0.87	

	GrupoAdulto*BUB1B	0.69	0.35	1.98	0.04
	GrupoAdolescente*POD1	-0.41	0.51	-0.81	0.42
	GrupoAdulto*POD1	-1.84	0.75	-2.47	0.01
	BUB1B*POD1	0.06	0.03	2.02	0.04
Grupo, PINK1, POD1	Intercepto	-0.95	0.69	-1.37	0.17
	GrupoAdolescente	0.54	1.09	0.49	0.62
	GrupoAdulto	1.00	0.76	1.32	0.19
	PINK1	-0.79	0.89	-0.89	0.37
	POD1	0.09	0.25	0.365	0.73
	GrupoAdolescente*PINK1	1.09	1.04	1.05	0.29
	GrupoAdulto*PINK1	0.61	0.93	0.66	0.51
	GrupoAdolescente*POD1	0.62	0.58	1.06	0.29
	GrupoAdulto*POD1	0.07	0.27	0.27	0.79
	PINK1*POD1	-0.32	0.26	-1.19	0.23

Tabela A.23: Modelos de regressão logística relacionando os genes alvo, sem interações.

Variável Resposta: Adenoma (0) e Carcinoma (1)						
Variáveis Predictoras	Parâmetros Ajustados	Coef. Estimado	DP	Estat. Z	p-valor	n
Grupo, SF1*BUB1B	Intercepto	-2.34	0.55	-4.21	<0.01	
	GrupoAdolescente	2.06	0.77	2.67	0.01	
	GrupoAdulto	1.47	0.57	2.60	0.01	124
	SF1	0.04	0.05	0.93	0.35	
	BUB1B	0.19	0.06	3.17	<0.01	
Grupo, SF1*PINK1	Intercepto	-1.26	0.42	-3.01	<0.01	
	GrupoAdolescente	1.63	0.68	2.38	0.02	
	GrupoAdulto	0.94	0.47	2.01	0.04	124
	SF1	0.04	0.04	0.93	0.35	
	PINK1	0.00	0.07	-0.01	0.99	
Grupo, BUB1B*PINK1	Intercepto	-2.19	0.55	-3.96	<0.01	
	GrupoAdolescente	2.17	0.80	2.72	0.01	
	GrupoAdulto	1.56	0.59	2.66	0.01	124
	BUB1B	0.22	0.07	3.39	<0.01	
	PINK1	-0.19	0.11	-1.78	0.08	

Tabela A.24: Modelos de regressão logística para variáveis com poucas observações

Variável Resposta: Adenoma (0) e Carcinoma (1)						
Variáveis Preditoras	Parâmetros Ajustados	Coef. Estimado	DP	Estat. Z	p-valor	n
Grupo, P53	Intercepto	-1.95	0.91	-2.13	0.03	
	GrupoAdolescente	1.47	0.78	1.89	0.06	44
	P53	1.02	0.92	1.11	0.27	
Grupo, IGF1R	Intercepto	-2.62	1.14	-2.31	0.02	
	GrupoAdolescente	1.66	1.22	1.36	0.17	49
	GrupoAdulto	2.07	1.03	2.00	0.05	
	IGF1R	0.33	0.18	1.80	0.07	
Grupo, IGF2	Intercepto	-1.75	0.70	-2.51	0.01	
	GrupoAdolescente	0.51	1.17	0.44	0.66	49
	GrupoAdulto	0.93	0.76	1.21	0.23	
	IGF2	0.02	0.01	2.43	0.02	
Grupo, β-catenina	Intercepto	-1.28	0.74	-1.72	0.09	
	GrupoAdolescente	1.28	1.25	1.03	0.31	30
	GrupoAdulto	1.31	0.94	1.39	0.16	
	β-catenina	-0.43	1.04	-0.42	0.68	

APÊNDICE B**GRÁFICOS**

Gráfico B.1: Diagnóstico por Grupo.

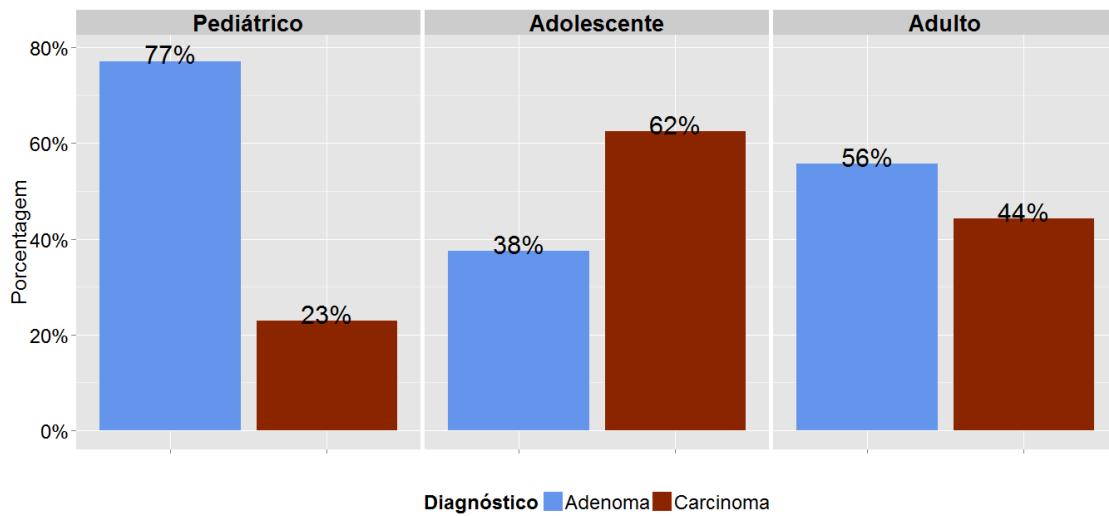


Gráfico B.2: Diagnóstico por Grupo e Sexo.

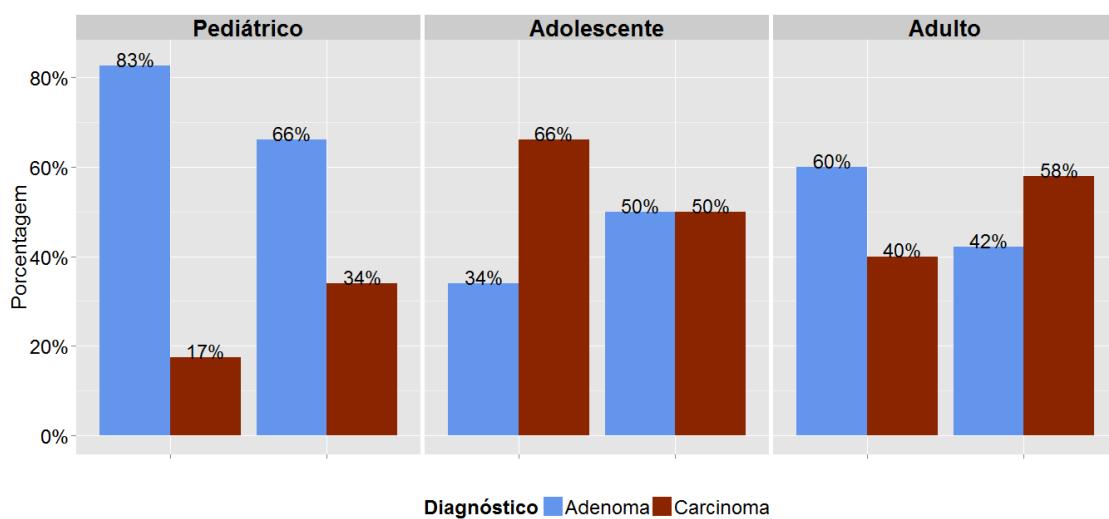


Gráfico B.3: Diagnóstico por Grupo e Evolução.

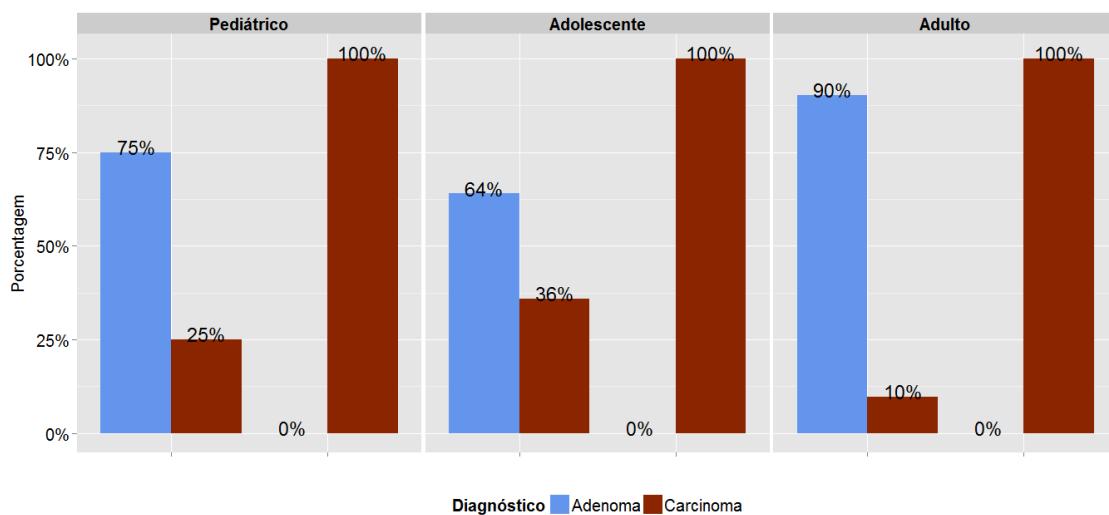


Gráfico B.4: Diagnóstico por Grupo e Ocorrência de Metástase.

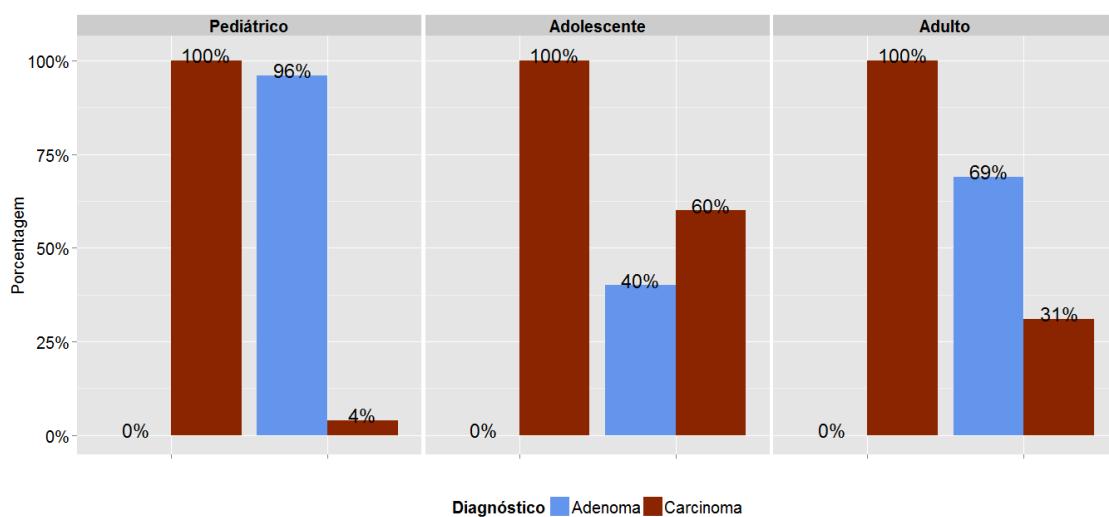


Gráfico B.5: *Boxplot e dotplot da Idade de Diagnóstico por Grupo e Diagnóstico.*

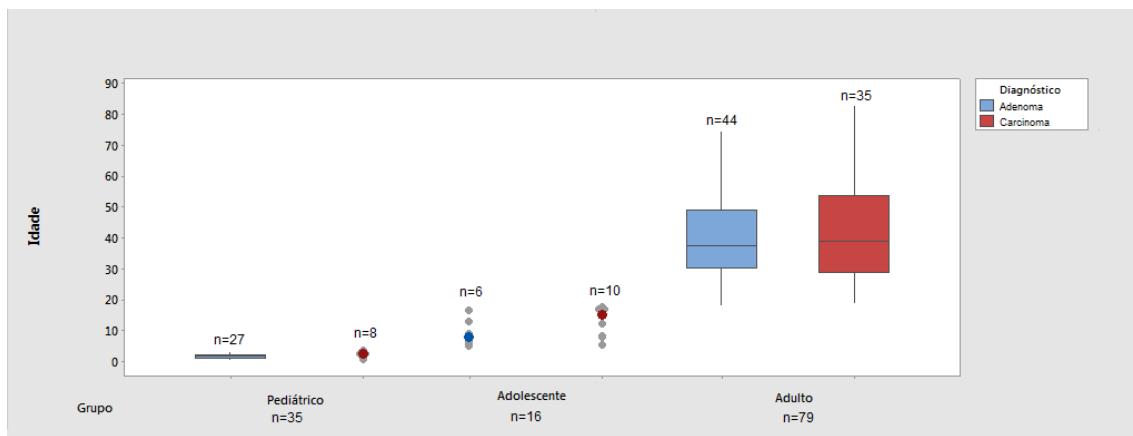


Gráfico B.6: Distribuição da Idade de Diagnóstico do paciente, por Grupo.

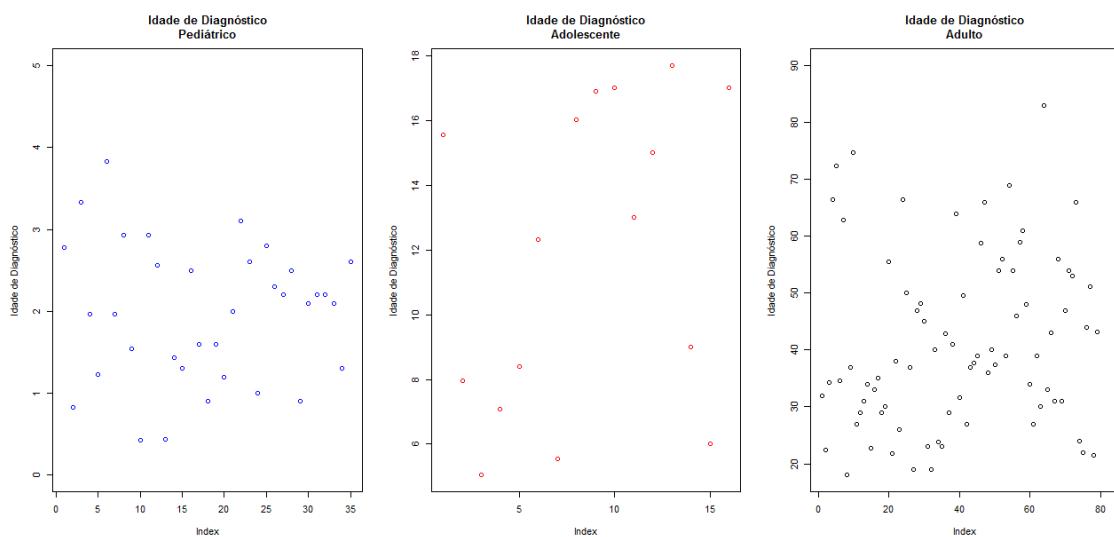


Gráfico B.7: Boxplot e dotplot do Escore de Weiss por Grupo e Diagnóstico.

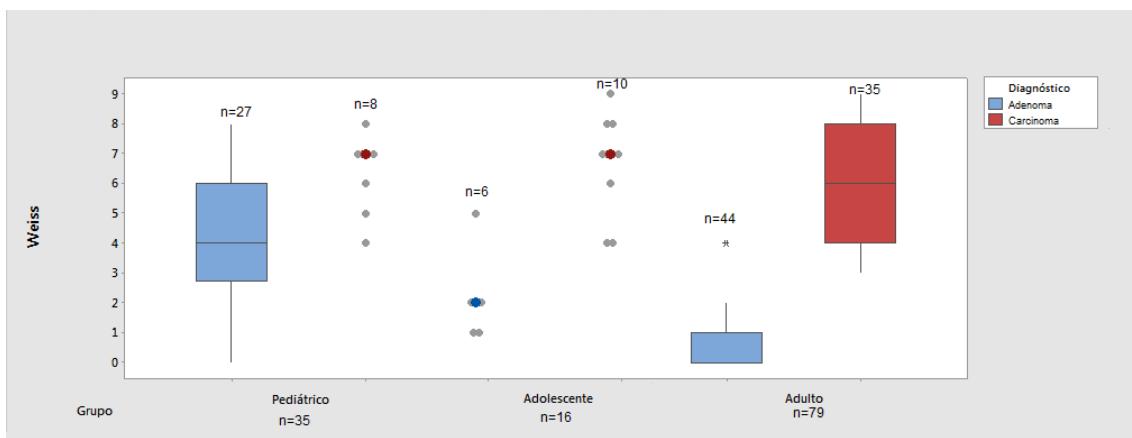


Gráfico B.8: Boxplot e dotplot do Tempo de Acompanhamento Médico (em anos) por Grupo e Diagnóstico.

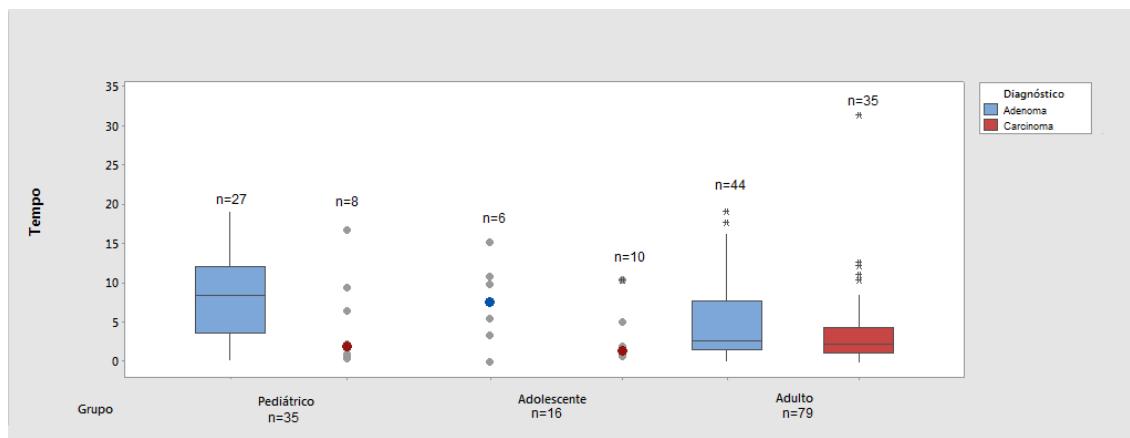


Gráfico B.9: Curva de Sobrevida, estimador de Kaplan-Meier, para o Tempo de Acompanhamento Médico (em anos) por Grupo.

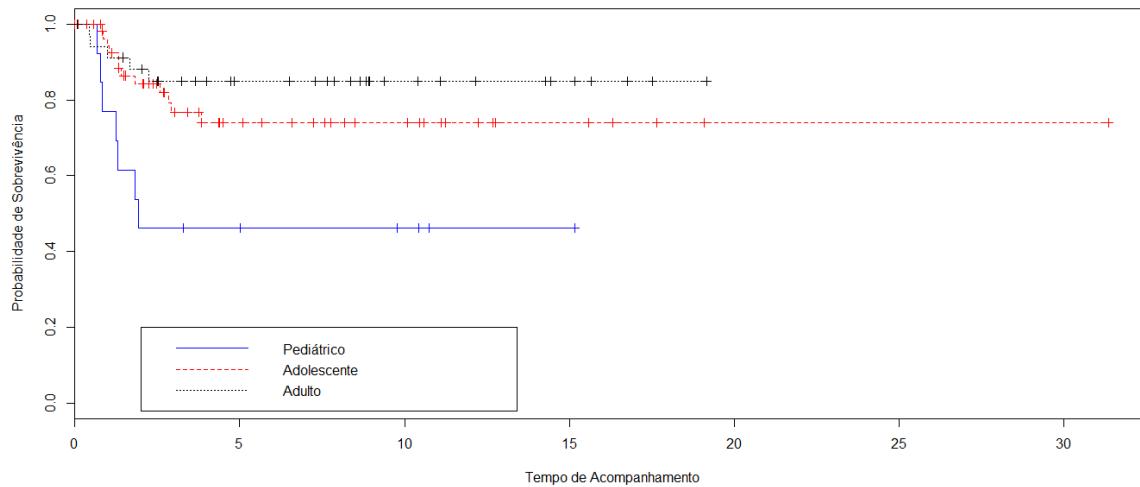


Gráfico B.10: Curva de Sobrevida, estimador de Kaplan-Meier, para o Tempo de Acompanhamento Médico (em anos) por Grupo e Diagnóstico Carcinoma.

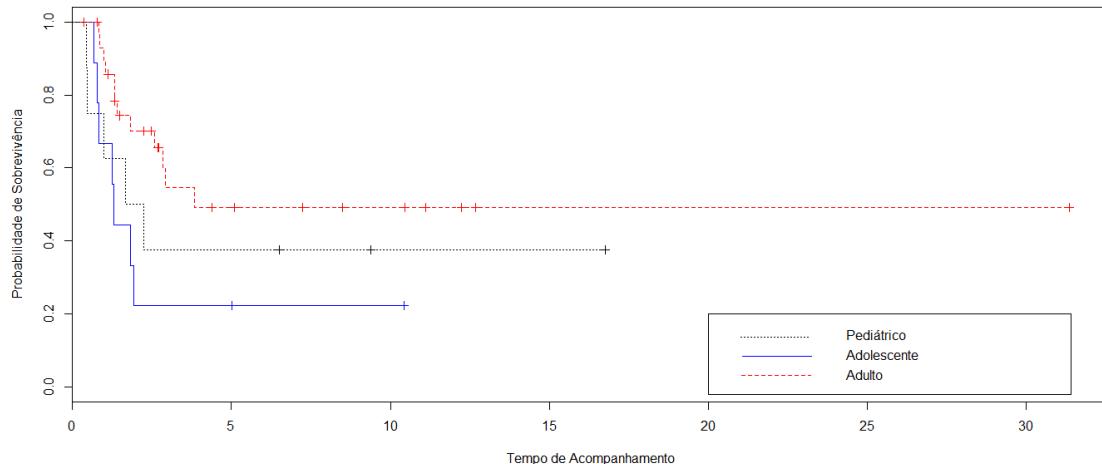


Gráfico B.11: Expressão Relativa do gene SF1 (Médias e desvios-padrão).

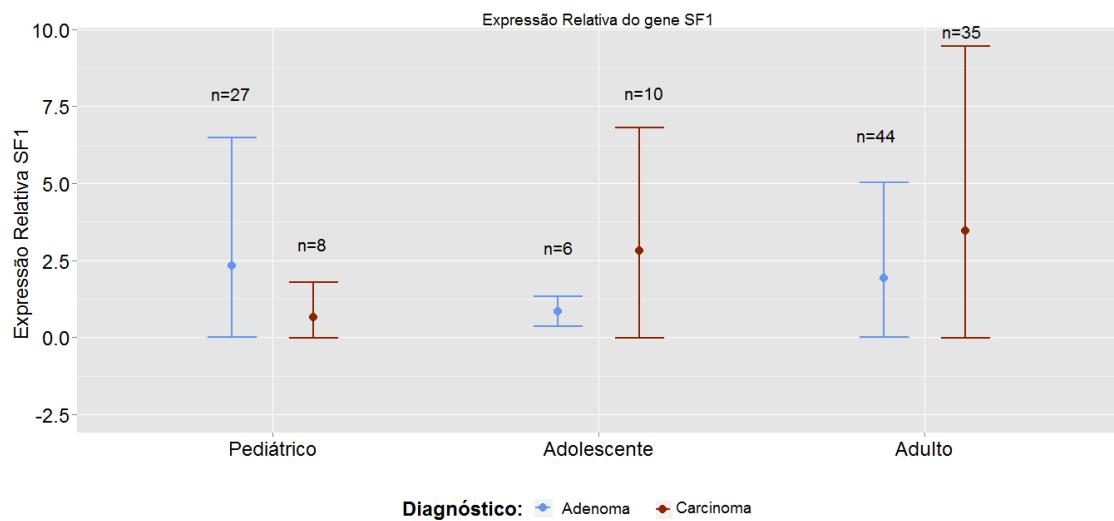


Gráfico B.12: Boxplot e dotplot para a Expressão Relativa do gene SF1

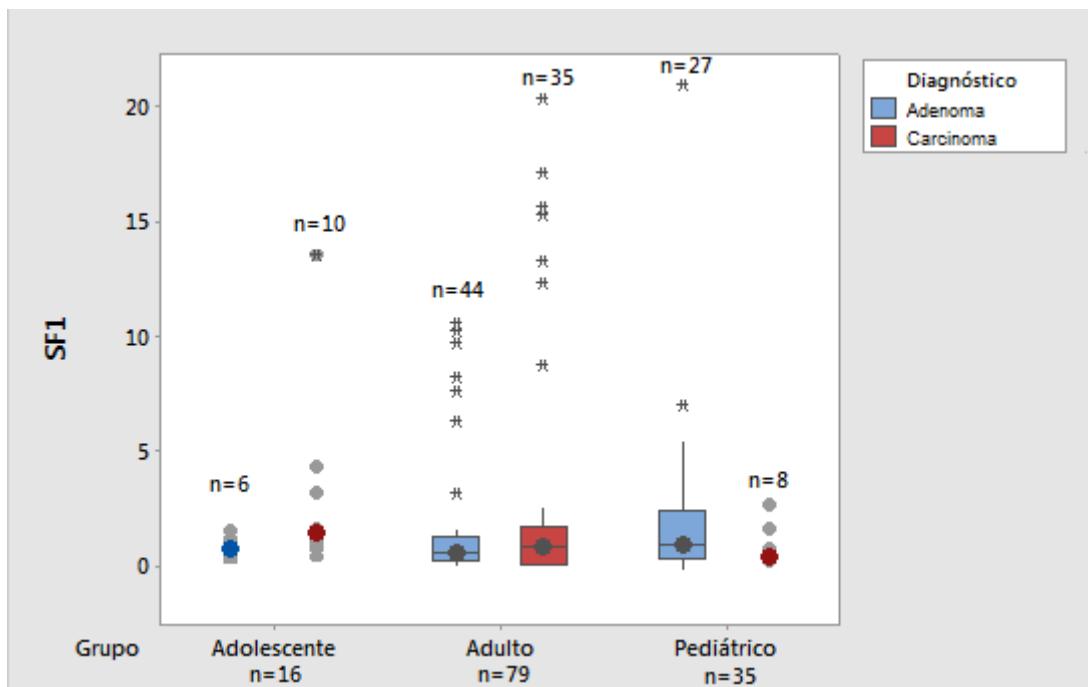


Gráfico B.13: Expressão Relativa do gene POD1 (Médias e desvios-padrão).

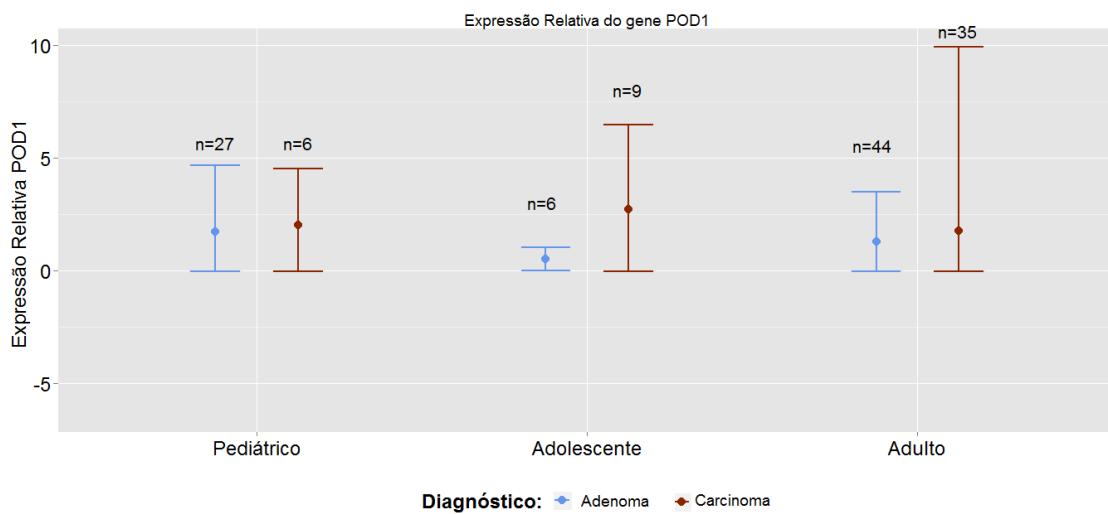


Gráfico B.14: Boxplot e dotplot para a Expressão Relativa do gene POD1.

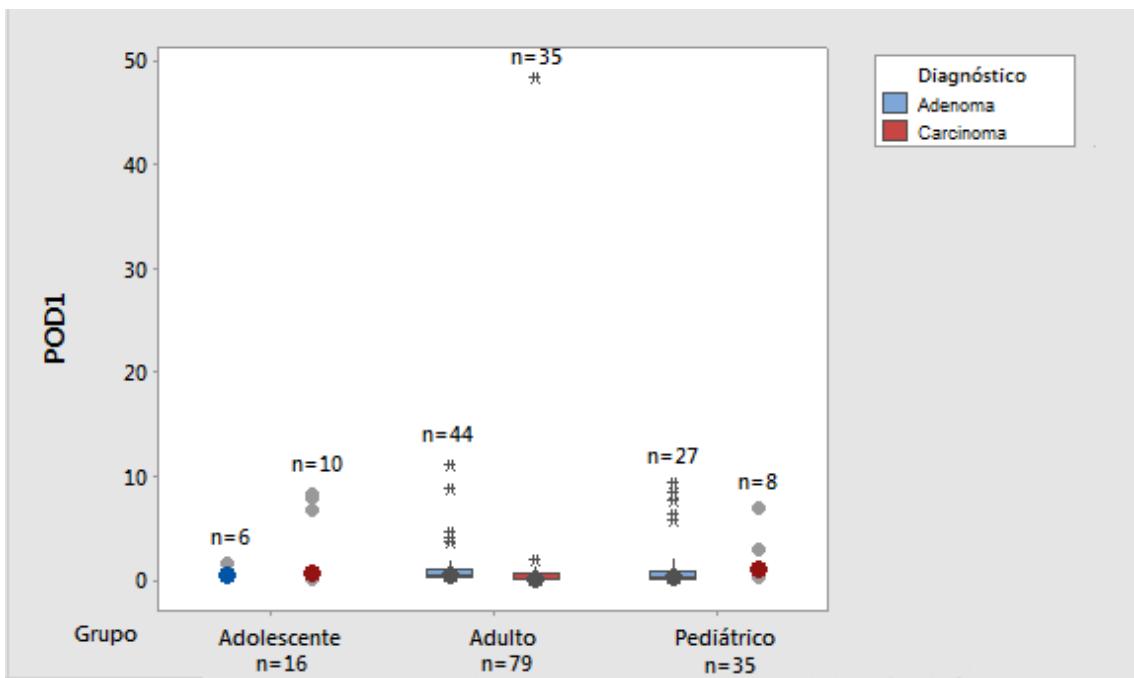


Gráfico B.15: Expressão Relativa do gene BUB1B (Médias e desvios-padrão).

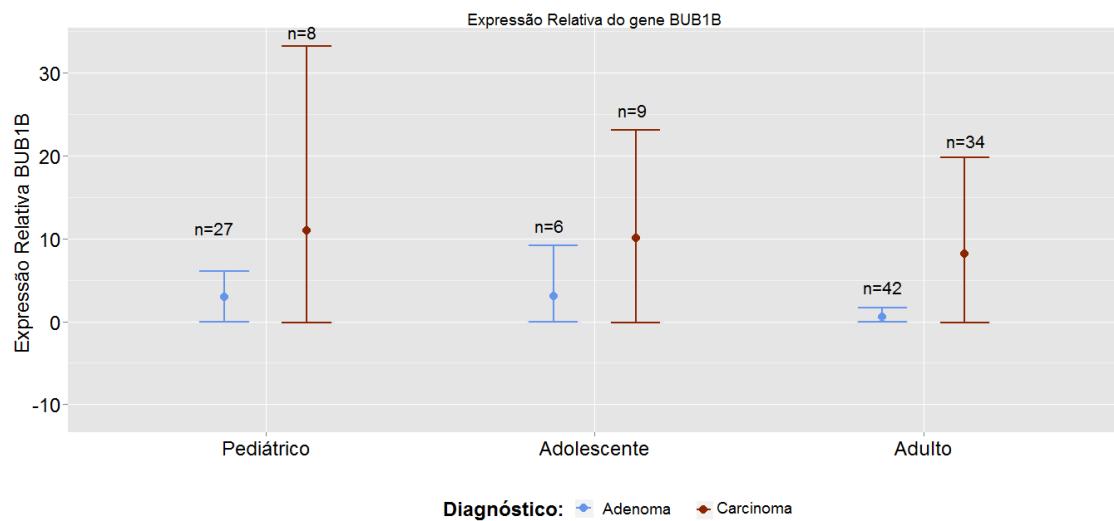


Gráfico B.16: Boxplot e dotplot para a Expressão Relativa do gene BUB1B.

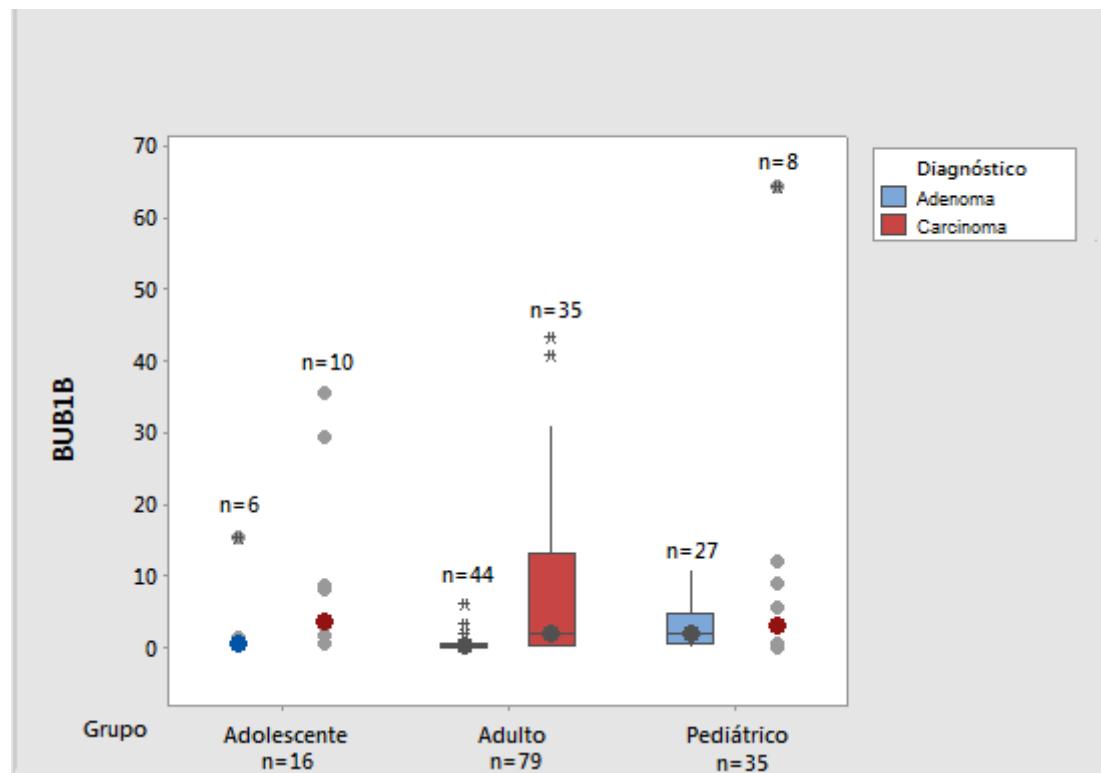


Gráfico B.17: Expressão Relativa do gene PINK1 (Médias e desvios-padrão).

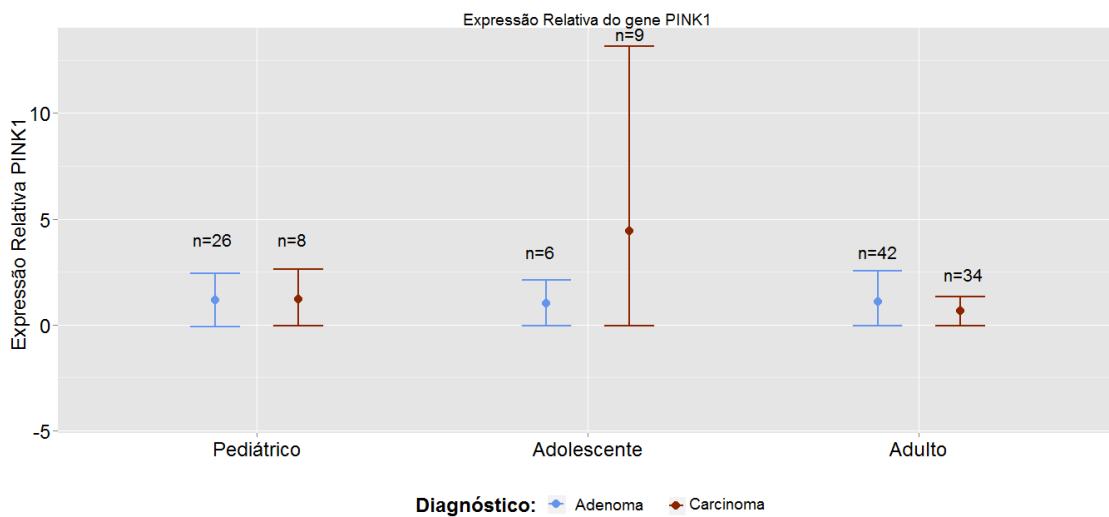


Gráfico B.18: Boxplot e dotplot para a Expressão Relativa do gene PINK1.

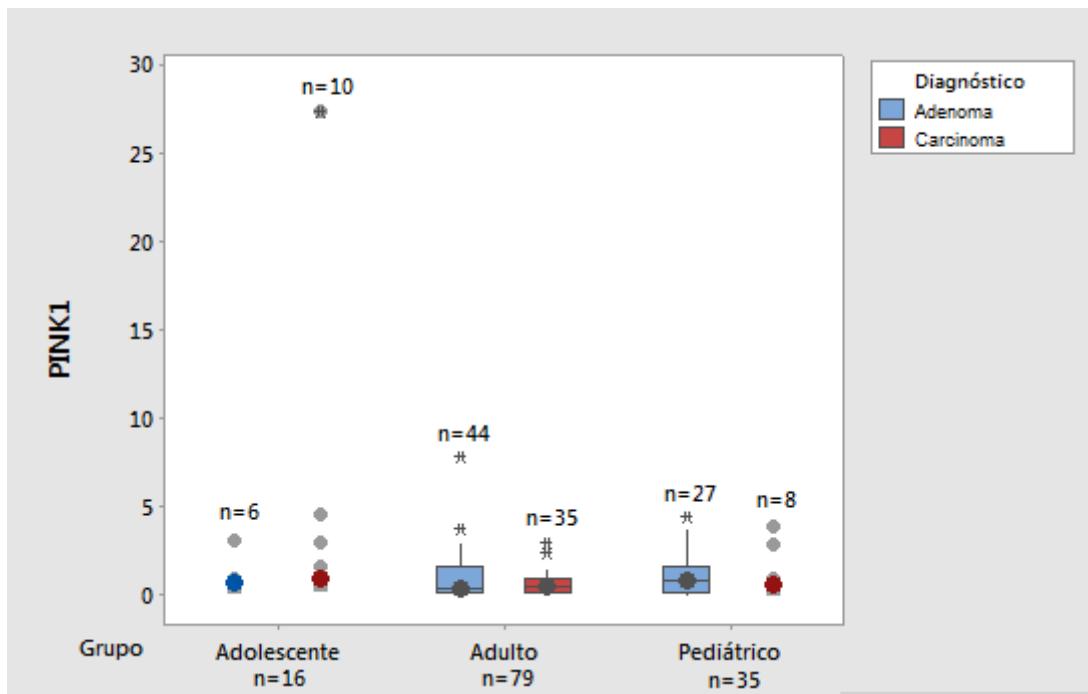


Gráfico B.19: Expressão Relativa do gene IGF1R (Médias e desvios-padrão).

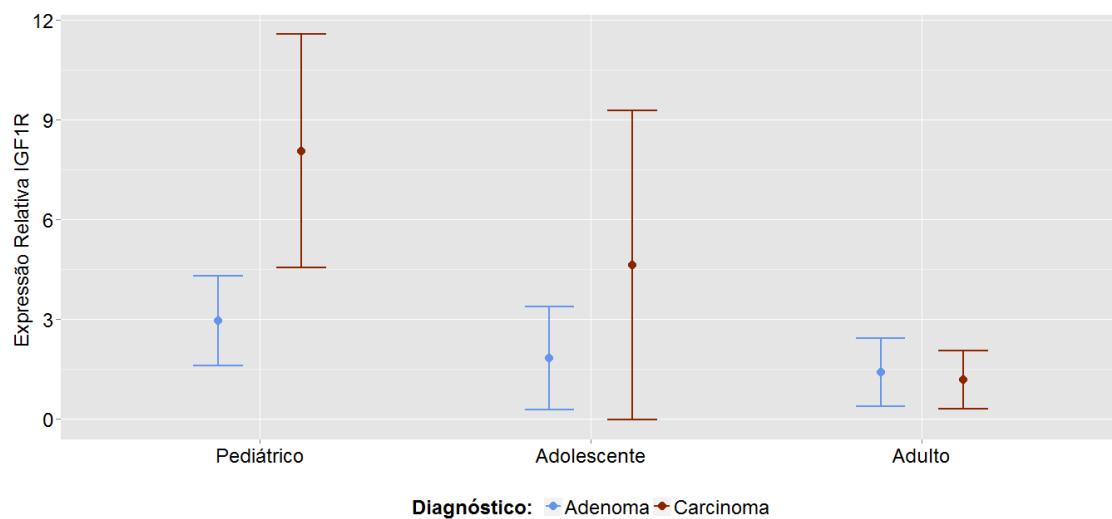


Gráfico B.20: Expressão Relativa do gene IGF2 (Médias e desvios-padrão).

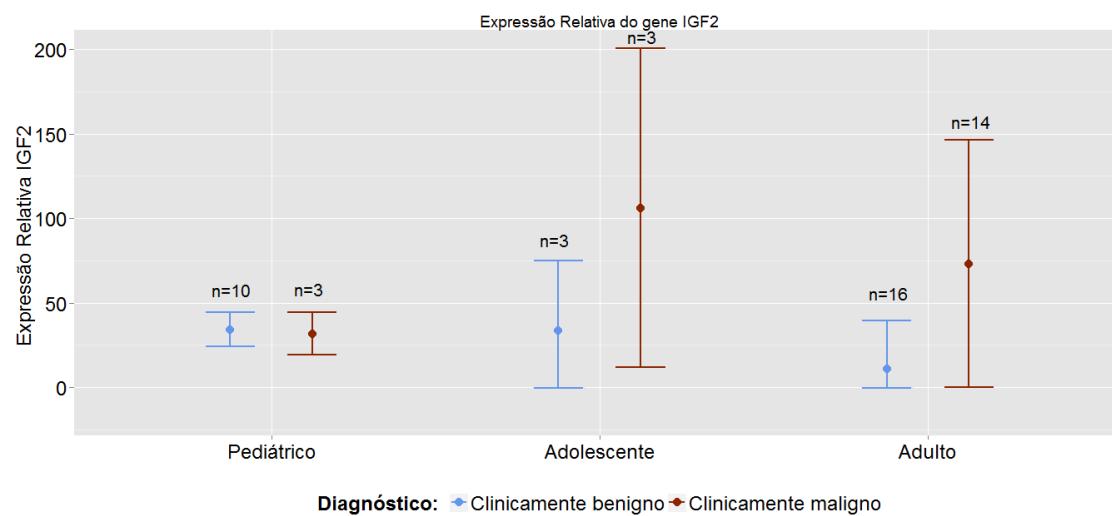


Gráfico B.21: Mutação no gene P53 por Grupo e Diagnóstico.

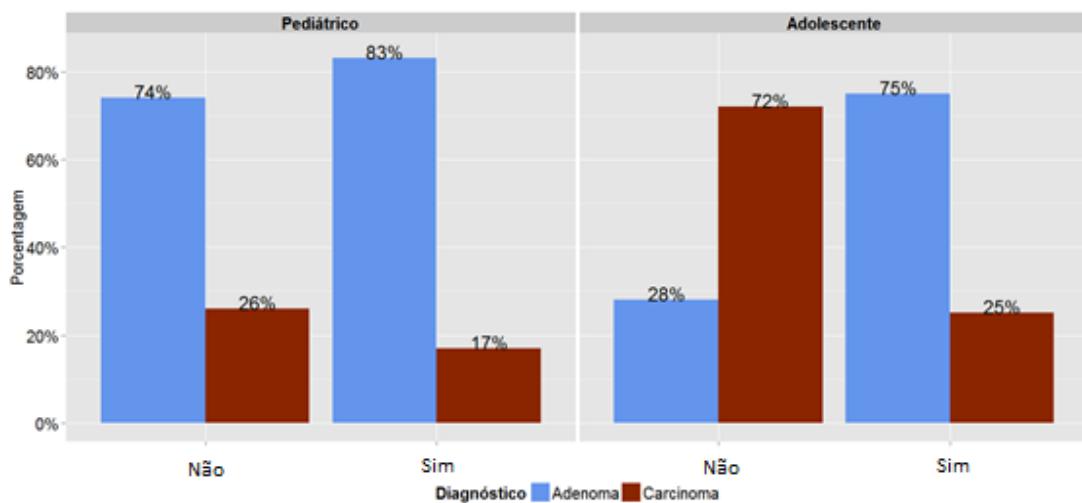


Gráfico B.22: Expressão Anômala da β-catenina por Grupo e Diagnóstico.

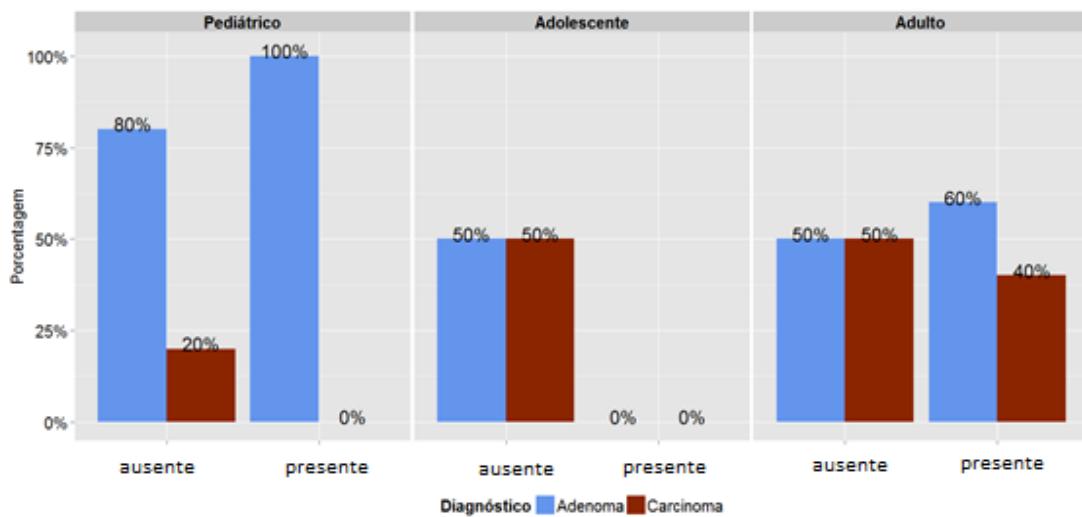


Gráfico B.23: Expressão Combinada de POD1 e BUB1B (Médias e desvios-padrão).

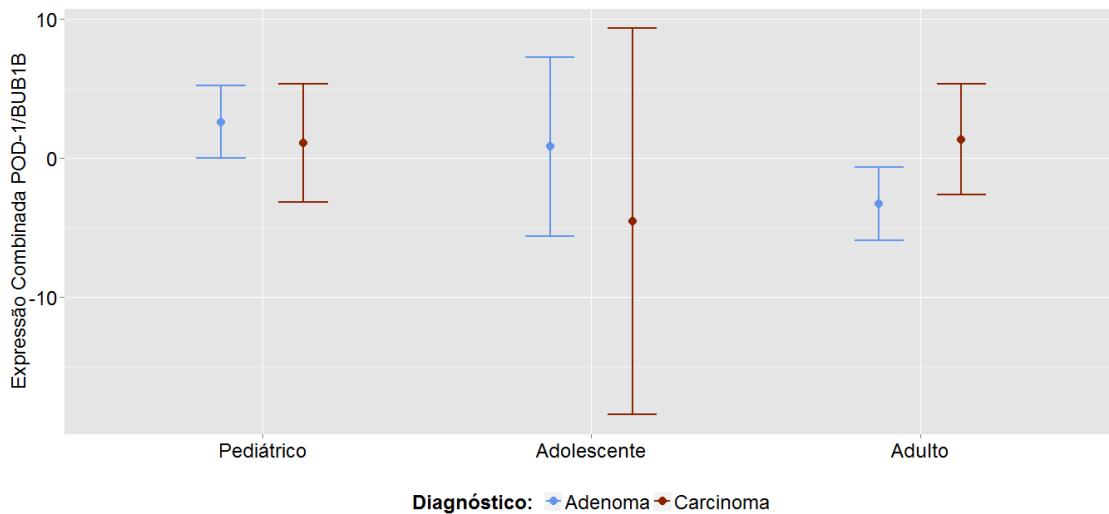


Gráfico B.24: Expressão Combinada de SF1 e BUB1B (Médias e desvios-padrão).

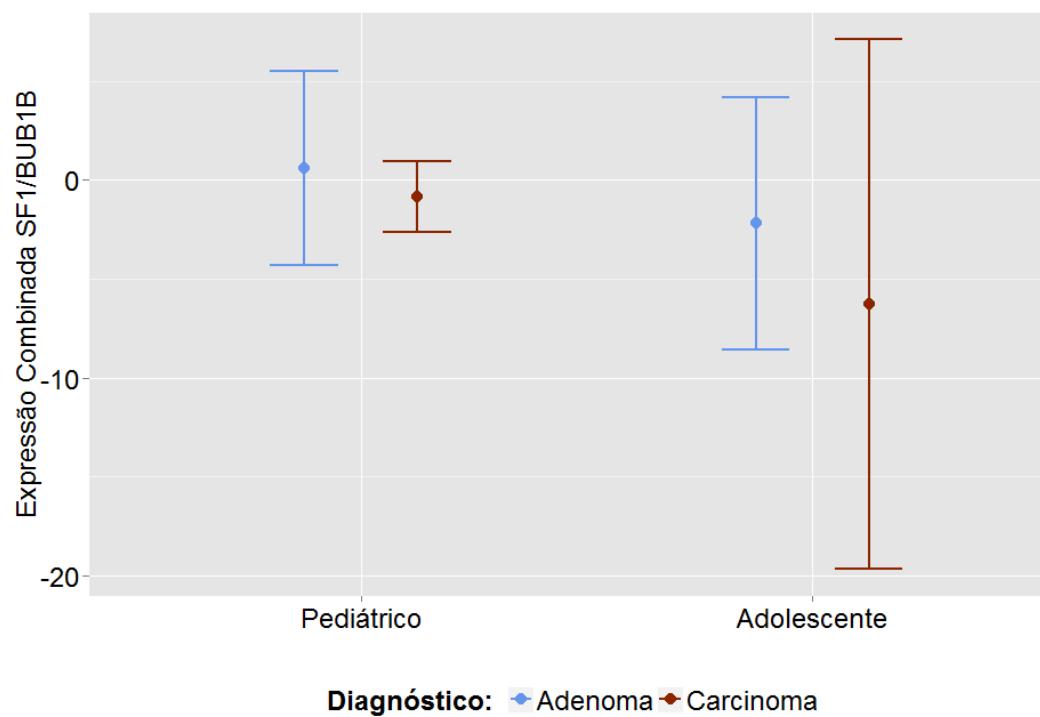


Gráfico B.25: Expressão Combinada de SF1 e POD1 (Médias e desvios-padrão).

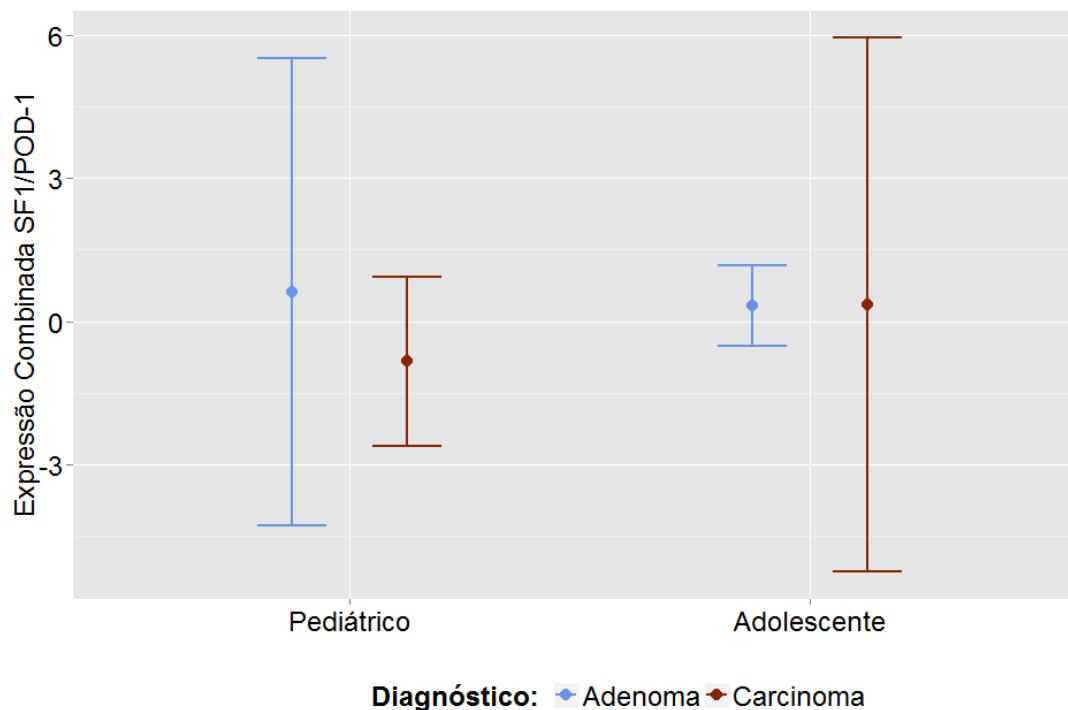


Gráfico B.26: Expressão Combinada de SF1 e PINK1 (Médias e desvios-padrão).

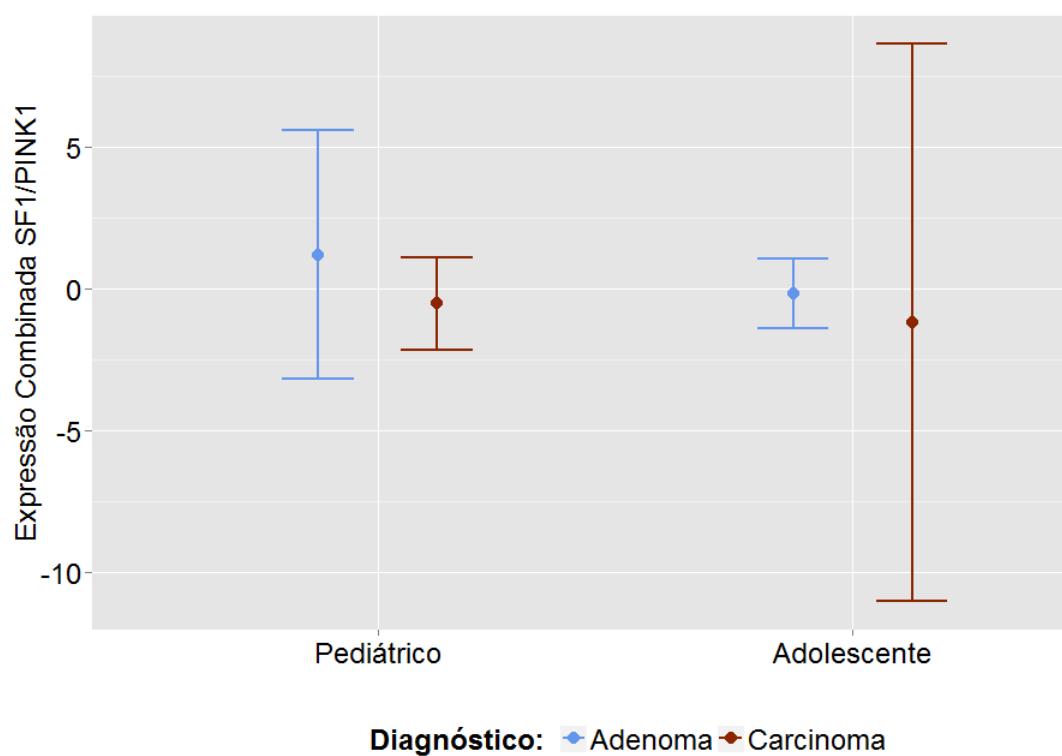


Gráfico B.27: Expressão Combinada de BUB1B e PINK1 (Médias e desvios-padrão).

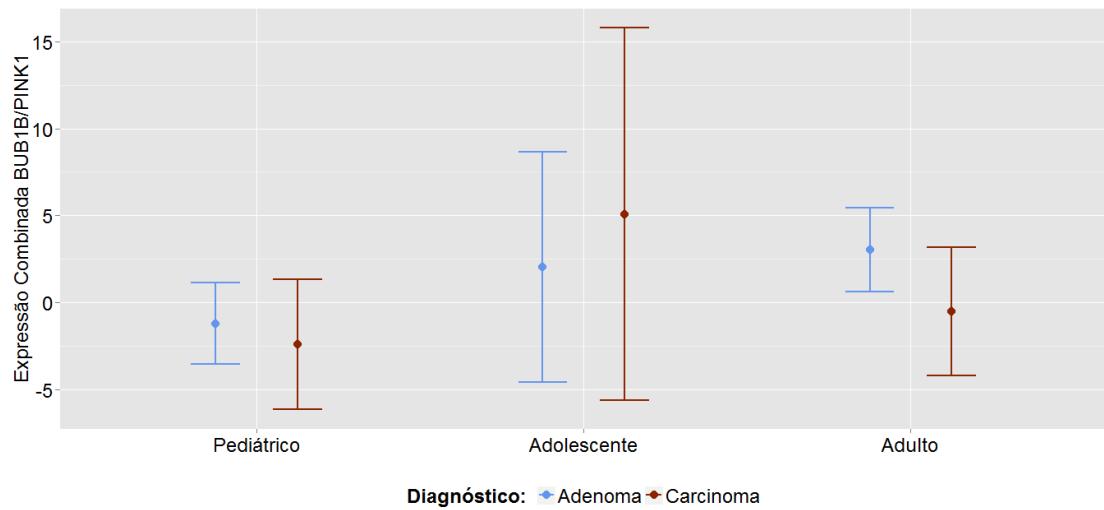
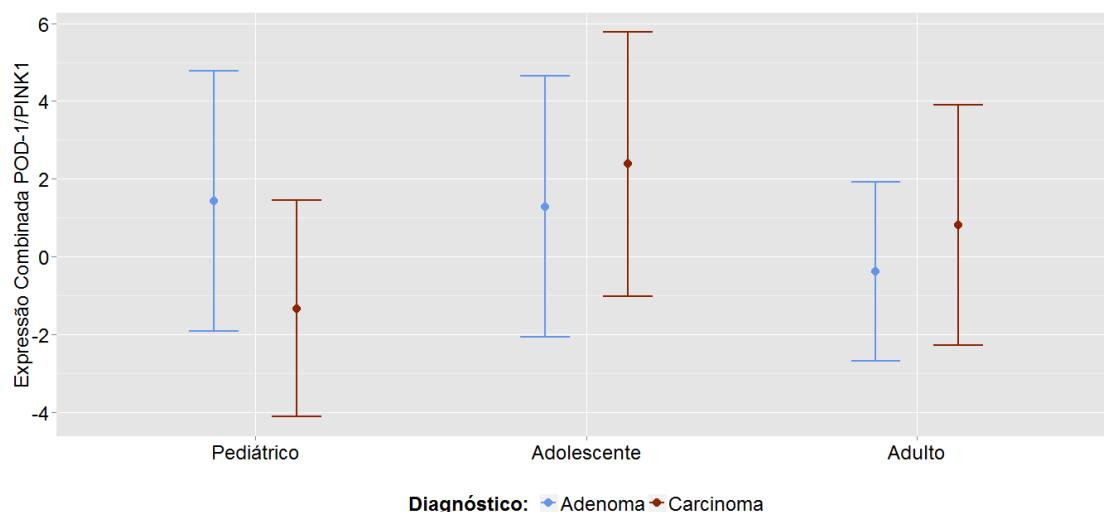


Gráfico B.28: Expressão Combinada de POD1 e PINK1 (Médias e desvios-padrão).



APÊNDICE C

MODELOS FINAIS

Modelo 1

Seja p a probabilidade de um paciente com tumor adrenocortical ser diagnosticado com carcinoma. De acordo com o modelo logístico:

$$p = \frac{1}{1 + e^{x\beta}}$$

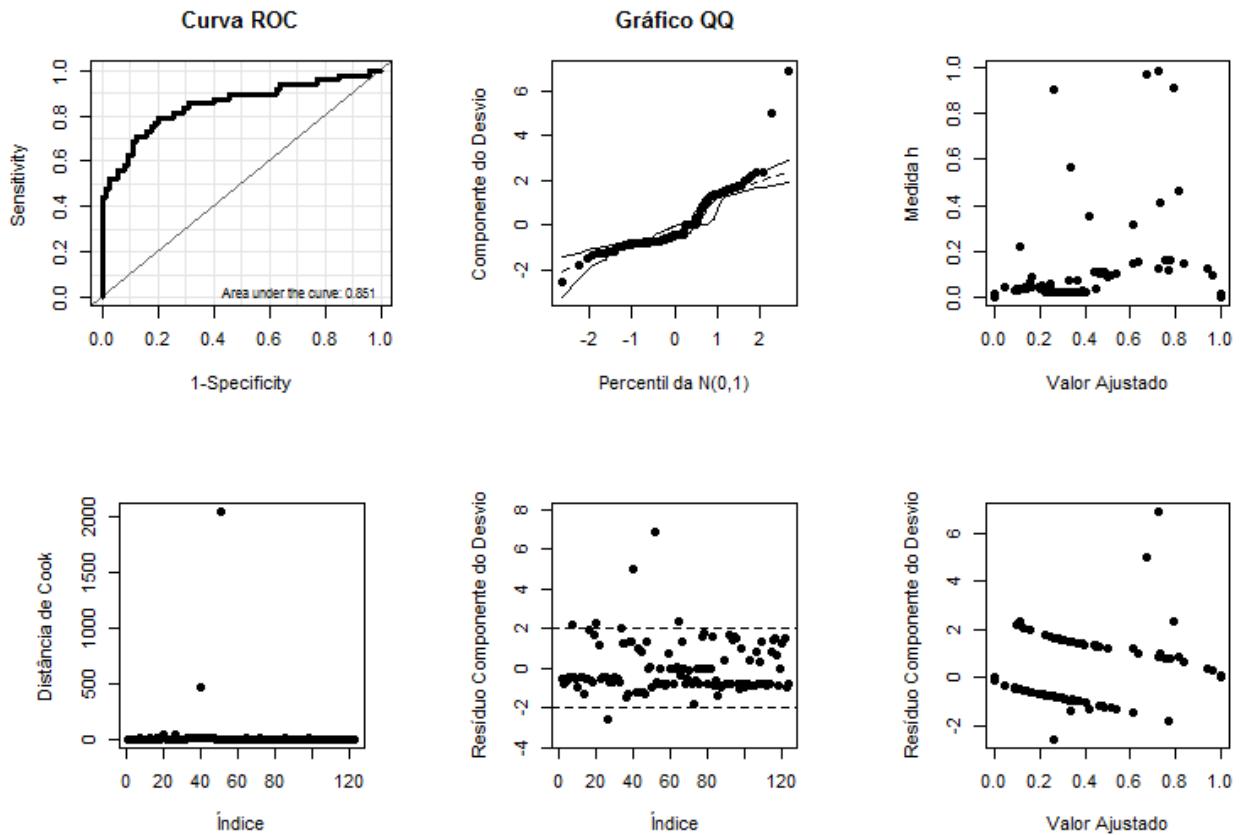
$$\ln\left(\frac{p}{1-p}\right) = X\beta = \beta_0 + \beta_1 GrupoAdolescente + \beta_2 GrupoAdulto + \\ + \beta_3 SF1 + \beta_4 BUB1B + \beta_5 POD1 + \beta_6 GrupoAdolescente * BUB1B + \\ + \beta_7 GrupoAdulto * BUB1B + \beta_8 GrupoAdolescente * POD1 + \beta_9 GrupoAdulto * POD1 + \\ + \beta_{10} SF1 * POD1$$

Após ajuste, obteve-se:

Tabela C.1: Ajuste do modelo 1.

n=123	Estimativa (β_i)	Erro Padrão	Estat. Z	Valor-p	$RC (e^{\beta_i})$	IC(RC,90%)
Intercepto	-2.41	0.65	-3.68	<0.001	0.09	(0.03 ; 0.26)
GrupoAdolescente	2.14	0.92	2.32	0.02	8.48	(1.87 ; 38.45)
GrupoAdulto	1.36	0.71	1.91	0.05	3.88	(1.21 ; 12.36)
SF1	0.15	0.08	1.99	0.04	1.16	(1.03 ; 1.32)
BUB1B	0.04	0.04	1.00	0.32	1.04	(0.98 ; 1.11)
POD1	1.03	0.48	2.14	0.03	2.79	(1.27 ; 6.14)
GrupoAdolescente*BUB1B	-0.004	0.07	-0.06	0.95	0.99	(0.89 ; 1.11)
GrupoAdulto*BUB1B	0.68	0.29	2.28	0.02	1.97	(1.21 ; 3.19)
GrupoAdolescente*POD1	-0.75	0.48	-1.55	0.12	0.47	(0.21 ; 1.05)
GrupoAdulto*POD1	-1.32	0.53	-2.48	0.01	0.27	(0.11 ; 0.64)
SF1*POD1	-0.14	0.06	-2.16	0.03	0.87	(0.78 ; 0.97)

Gráfico C.1: Diagnóstico do modelo 1.



Modelo 2

Seja p a probabilidade de um paciente com tumor adrenocortical ser diagnosticado com carcinoma. De acordo com o modelo logístico:

$$p = \frac{1}{1 + e^{X\beta}}$$

$$\ln\left(\frac{p}{1-p}\right) = X\beta = \beta_0 + \beta_1 GrupoAdolescente + \beta_2 GrupoAdulto + \beta_3 BUB1B + \beta_4 PINK1$$

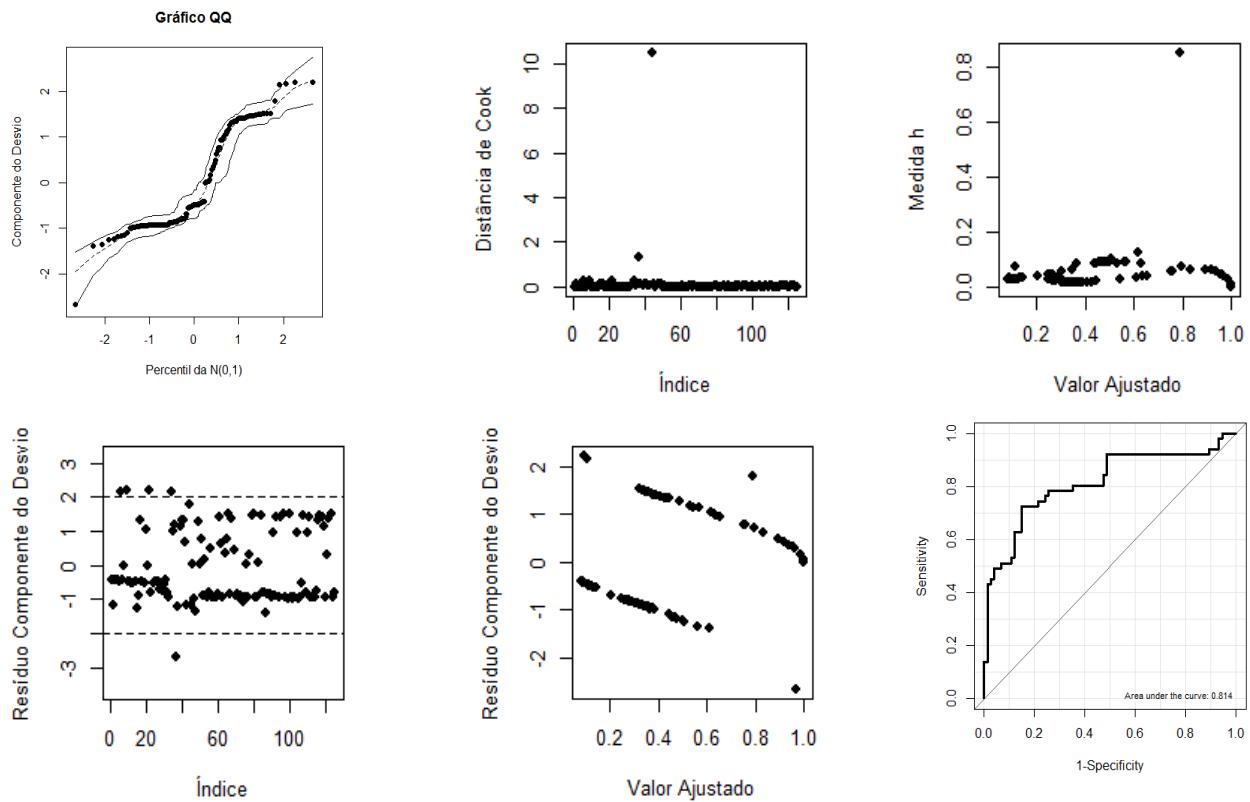
Após ajuste, obteve-se:

Tabela C.2: Ajuste do modelo 2.

n=125	Estimativa (β_i)	Erro Padrão	Estat.Z	Valor-p	$RC (e^{\hat{\beta}_i})$	IC(RC,90%)
Intercepto	-2.19	0.55	-3.96	<0.001	0.11	(0.05 ; 0.28)
GrupoAdolescente	2.17	0.8	2.72	0.01	8.76	(2.36 ; 32.52)
GrupoAdulto	1.56	0.59	2.66	0.01	4.78	(1.80 ; 12.52)

BUB1B	0.22	0.07	3.39	<0.001	1.25	(1.11 ; 1.40)
PINK1	-0.19	0.11	-1.78	0.07	0.83	(0.69 ; 0.99)

Gráfico C.2: Diagnóstico do modelo 2.



Modelo 3

Seja p a probabilidade de um paciente com tumor adrenocortical ser diagnosticado com carcinoma. De acordo com o modelo logístico:

$$p = \frac{1}{1 + e^{x\beta}}$$

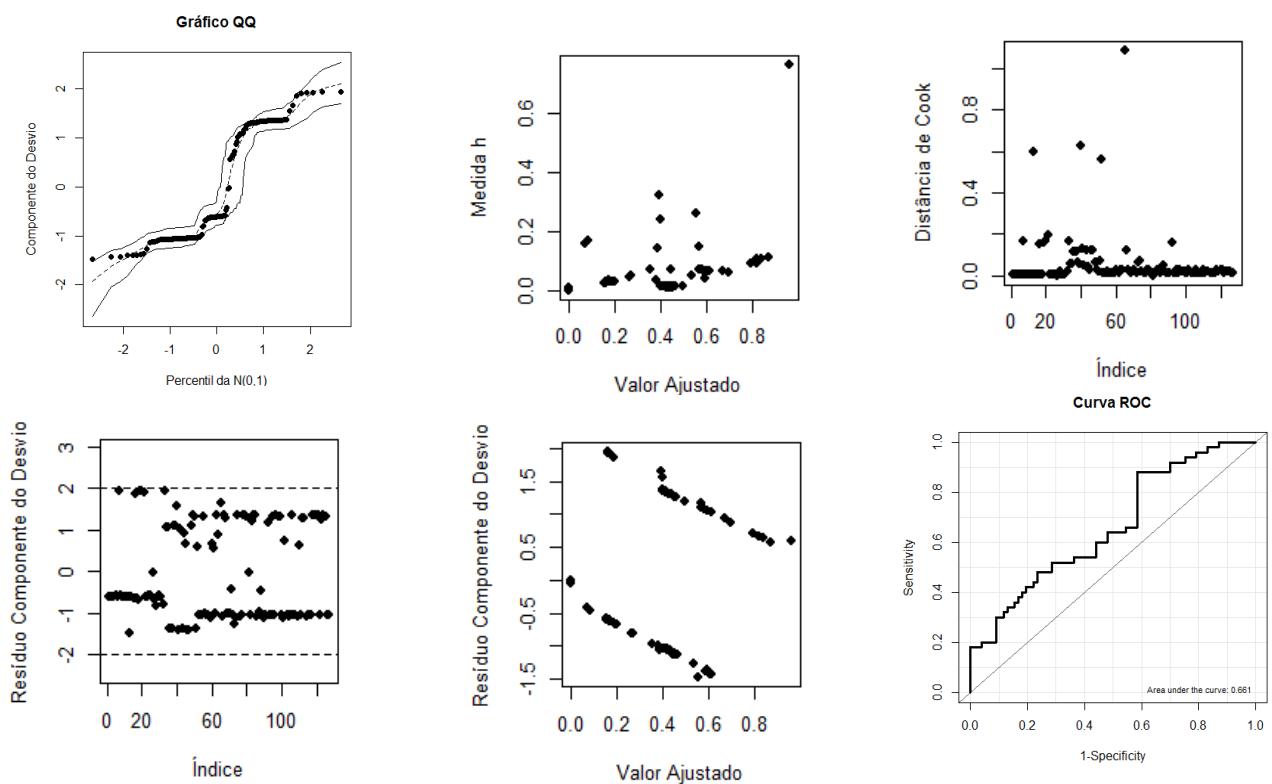
$$\ln\left(\frac{p}{1-p}\right) = X\beta = \beta_0 + \beta_1 GrupoAdolescente + \beta_2 GrupoAdulto + \beta_3 SF1 + \\ + \beta_4 POD1 + \beta_5 SF1 * POD1$$

Após ajuste, obteve-se:

Tabela C.3: Ajuste do modelo 3.

n=127	Estimativa (β_i)	Erro Padrão	Estat.Z	Valor-p	$RC (e^{\hat{\beta}_1})$	IC(RC,90%)
Intercepto	-1.71	0.48	-3.57	<0.001	0.18	(0.08 ; 0.39)
GrupoAdolescente	1.98	0.71	2.79	0.01	7.23	(2.26 ; 23.21)
GrupoAdulto	1.3	0.52	2.5	0.01	3.68	(1.56 ; 8.61)
SF1	0.16	0.07	2.21	0.03	1.17	(1.05 ; 1.32)
POD1	0.08	0.09	0.81	0.42	1.08	(0.93 ; 1.26)
SF1*POD1	-0.12	0.07	-1.78	0.08	0.89	(0.79 ; 0.99)

Gráfico C.3: Diagnóstico do modelo 3.



Modelo 4

Seja p a probabilidade de um paciente com tumor adrenocortical ser diagnosticado com carcinoma. De acordo com o modelo logístico:

$$p = \frac{1}{1 + e^{x\beta}}$$

$$\ln\left(\frac{p}{1-p}\right) = X\beta = \beta_0 + \beta_1 GrupoAdolescente + \beta_2 GrupoAdulto + \beta_3 IGF1R +$$

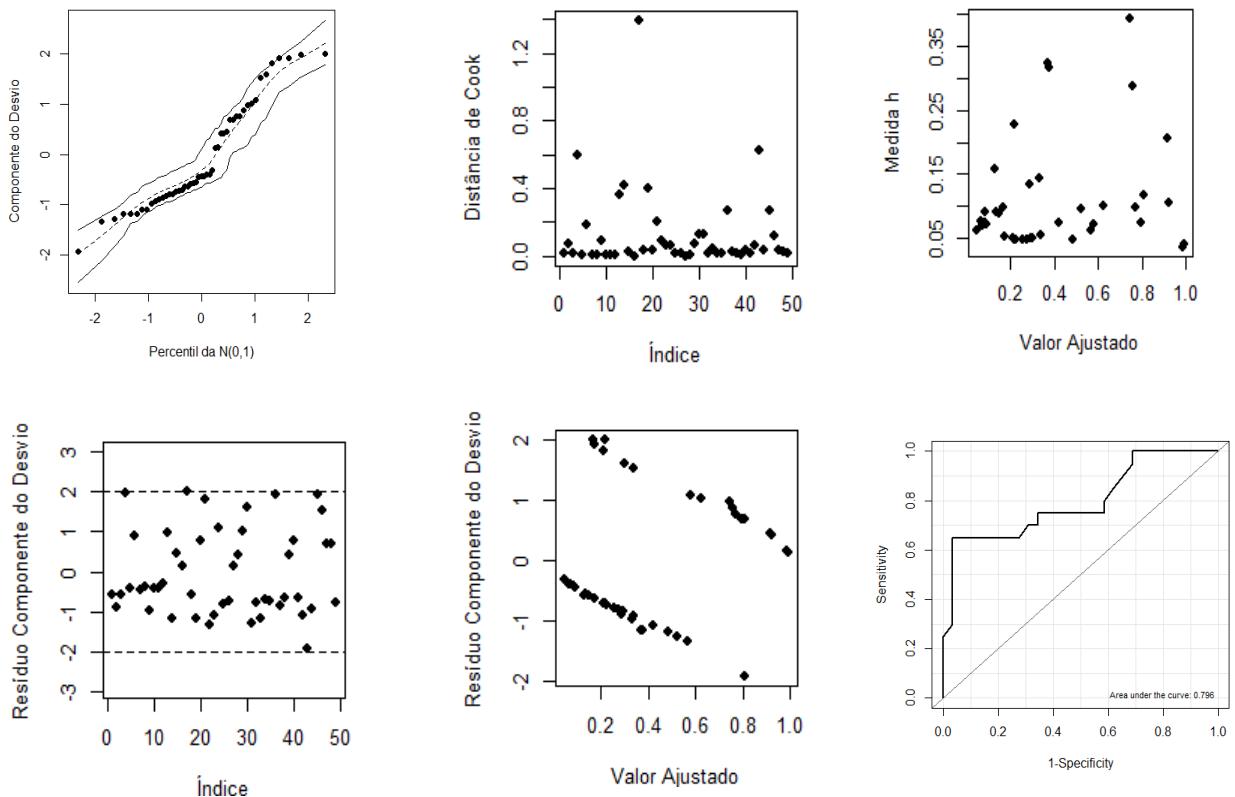
$$+ \beta_4 IGF2$$

Após ajuste, obteve-se:

Tabela C.4: Ajuste do modelo 4.

n=49	Estimativa (β_i)	Erro Padrão	Estat.Z	Valor-p	RC	IC(,90%)
Intercepto	-3.52	1.3	-2.71	0.01	0.03	(0.003 ; 0.25)
GrupoAdolescente	1.37	1.4	0.98	0.33	3.93	(0.39 ; 39.10)
GrupoAdulto	2.2	1.12	1.96	0.05	8.99	(1.44 ; 56.64)
IGF1R	0.37	0.2	1.84	0.07	1.45	(1.04 ; 2.01)
IGF2	0.02	0.01	2.44	0.01	1.02	(1.00 ; 1.04)

Gráfico C.4: Diagnóstico do modelo 4.



Comparação de modelos

Tabela C.5: Diagnóstico de pacientes: valores observados e ajustados.

Paciente	Grupo	SF1	POD1	BUB1B	IGF1R	IGF2	PINK1	Diagn.	P(Y=1)	P(Y=1)	P(Y=1)	P(Y=1)
									Modelo 1	Modelo 2	Modelo 3	Modelo 4
A1	Pediátrico	21.00	0.59	0.89	3.84	29.00	0.85	0	0.42	0.10	0.55	0.18
A12	Pediátrico	1.65	0.14	9.05	4.04	32.59	0.28	1	0.16	0.45	0.19	0.20
C6	Adolescente	0.87	0.28	0.34	1.47	0.27	3.16	0	0.48	0.37	0.60	0.17
C4	Adolescente	4.33	0.57	35.49	9.95	123.08	0.67	1	0.81	1.00	0.67	0.98
B8	Adulto	0.08	0.20	0.04	1.21	0.58	0.04	0	0.25	0.35	0.41	0.30
B1	Adulto	0.03	0.06	40.80	1.14	96.31	1.26	1	1.00	1.00	0.40	0.72

Tabela C.6: Comparação dos modelos ajustados.

Modelo	AIC	Null Deviance	Deviance Residual
1	132.39	142.80 (122 g.l.)	110.39 (112 g.l.)
2	141.62	169.03 (124 g.l.)	131.62 (120 g.l.)
3	163.29	170.28 (126 g.l.)	151.29 (121 g.l.)
4	57.97	53.99 (48 g.l.)	47.97 (44 g.l.)