Understanding the Mutations, Signaling, and genetic basis behind the progression of Hepatocellular carcinoma (HCC)

Understanding the Mutations, Signaling, and genetic basis behind the progression of Hepatocellular carcinoma (HCC)

Abstract

Hepatocellular carcinoma (HCC) represents a significant health burden as the most common primary liver cancer worldwide. This project delves into the genetic mutations, signaling pathways, and transcriptional mechanisms underlying HCC progression. Through the analysis of key driver mutations in genes such as TP53, CTNNB1, ARID1A, AXIN1, and ARID2, as well as transcriptional profiling using advanced computational tools, this study identifies critical biomarkers and disrupted pathways associated with the disease. Integrative bioinformatics approaches, including differential expression analysis and visualization techniques, unveil significant transcriptional alterations that highlight potential therapeutic targets and biomarkers for diagnosis. The findings aim to contribute to personalized treatment strategies and a deeper understanding of HCC pathogenesis.

Introduction: Hepatocellular carcinoma (HCC)

Hepatocellular carcinoma (HCC) accounts for 85-90% of all primary liver cancers and is a leading cause of cancer-related deaths worldwide. Its development is a multifaceted process influenced by genetic mutations, chronic liver disease, and environmental factors. Key driver mutations in tumor suppressor and oncogenes disrupt critical cellular processes such as apoptosis, DNA repair, and signaling pathways. This project investigates these genetic alterations and their implications on transcriptional dynamics in HCC. Utilizing datasets from platforms like IntOGen and transcriptional analysis pipelines, this study aims to map the molecular landscape of HCC, providing insights into its progression and identifying actionable biomarkers for early detection and targeted therapies.

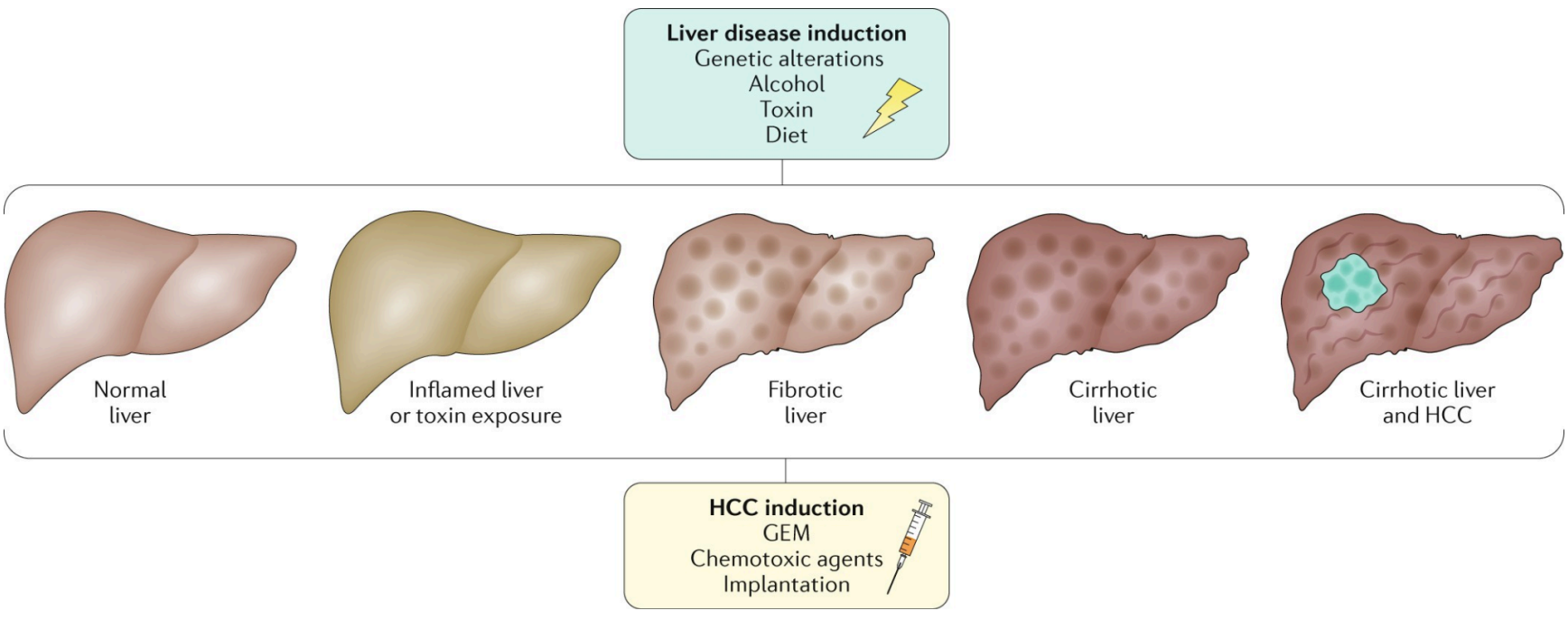


Image source: https://www.nature.com/articles/s41575-018-0033-6

Causes and Risk Factors

HCC often develops in the context of chronic liver disease or cirrhosis. Major risk factors include:

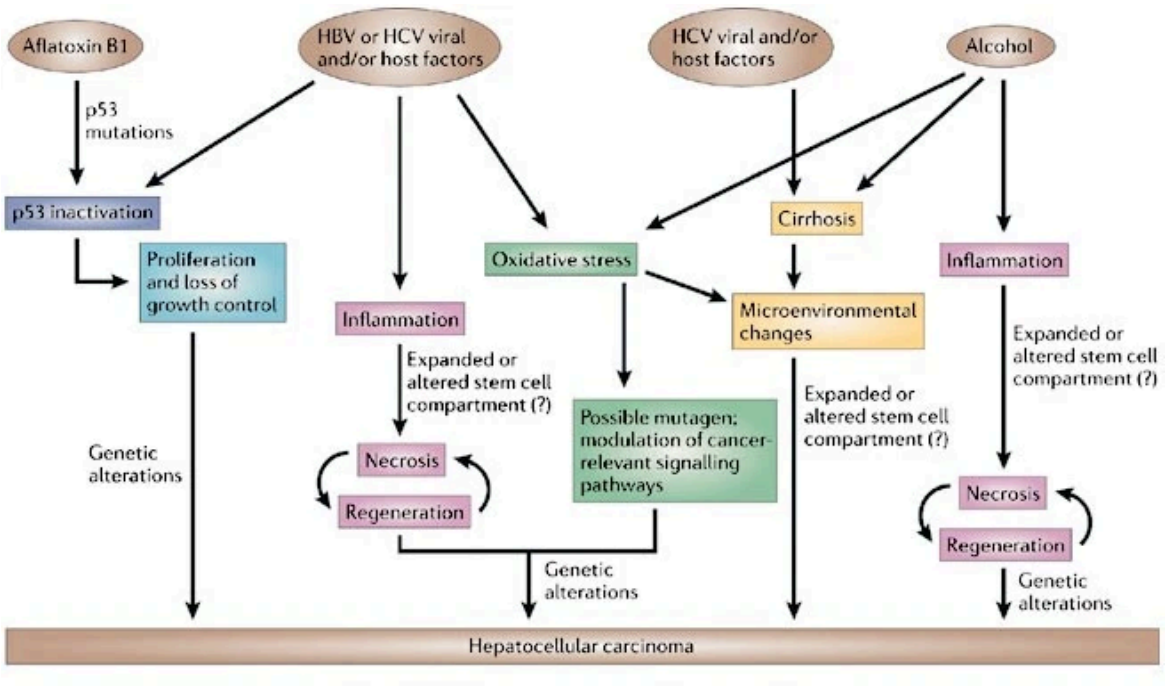
1. **Chronic Hepatitis B (HBV) or Hepatitis C (HCV) Infections**: These viral infections can lead to long-term liver damage and are common in HCC cases.
2. **Cirrhosis**: Scarring of the liver, typically from chronic hepatitis or heavy alcohol consumption, increases the risk of HCC.
3. **Non-Alcoholic Fatty Liver Disease (NAFLD)**: Associated with obesity, type 2 diabetes, and metabolic syndrome, this condition can lead to non-alcoholic steatohepatitis (NASH), which is linked to HCC.
4. **Alcohol Consumption**: Chronic heavy drinking contributes to liver damage and cirrhosis, raising HCC risk.
5. **Aflatoxin Exposure**: Consumption of foods contaminated with aflatoxin, a toxin from certain molds, is a risk factor, particularly in parts of Asia and Africa.
6. **Genetic and Metabolic Disorders**: Conditions such as hereditary hemochromatosis (iron overload) and alpha-1 antitrypsin deficiency increase HCC risk.

The development of HCC is a complex process influenced by several factors:

1. **Chronic Liver Injury:** Persistent damage to liver cells from various causes leads to inflammation and fibrosis, ultimately resulting in cirrhosis. This scarring disrupts normal liver architecture and function, creating an environment conducive to cancer development.
2. **Viral Infections:** Chronic infections with hepatitis B virus (HBV) and hepatitis C virus (HCV) are major contributors to HCC. For HBV, the **integration of viral DNA into the host genome** can lead to mutations in critical genes involved in cell cycle regulation and apoptosis. This integration often occurs at **telomerase reverse transcriptase (TERT) promoter sites**, leading to increased telomerase activity and cellular immortality, which are pivotal in oncogenesis.
3. **Alcohol Consumption:** Chronic excessive alcohol intake causes alcoholic liver disease, which can progress to cirrhosis. Alcohol metabolism **generates reactive oxygen species** that induce oxidative stress, leading to further

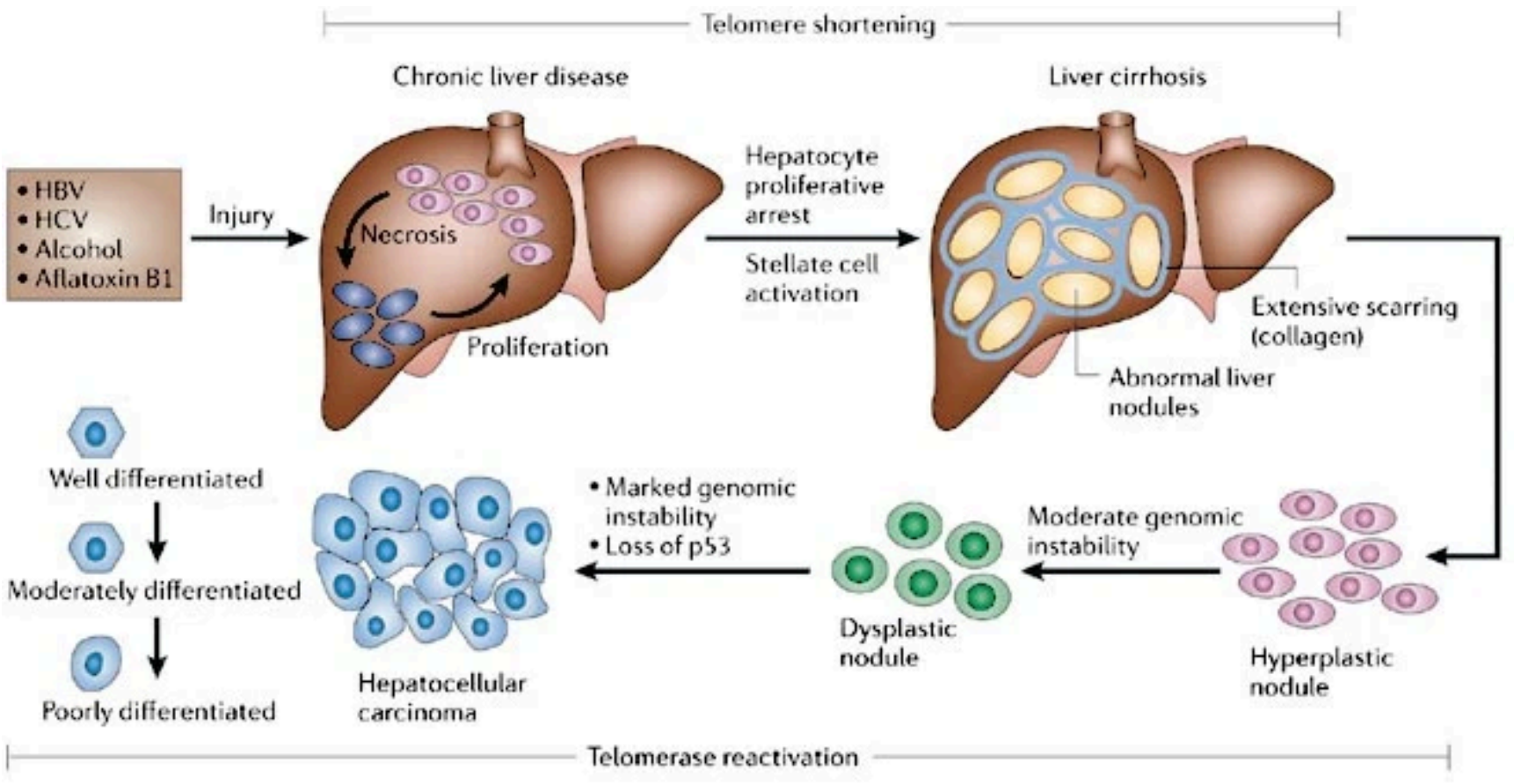
liver damage and inflammation. This chronic inflammatory state can promote genetic mutations and cellular proliferation, increasing the risk of HCC development.

4. **Non-Alcoholic Fatty Liver Disease (NAFLD):** Conditions like non-alcoholic steatohepatitis (NASH), associated with obesity and metabolic syndrome, are increasingly recognized as significant risk factors for HCC.



Histopathological progression and molecular features of HCC:

Liver damage caused by factors like HBV, HCV, alcohol, or aflatoxin B1 leads to repeated cycles of cell death and regeneration, resulting in chronic liver disease. Over time, this persistent damage triggers liver scarring and the development of cirrhosis, where abnormal liver nodules form amidst collagen deposits. These nodules can progress from hyperplastic (rapidly growing) to dysplastic (abnormally shaped and potentially pre-cancerous), eventually developing into hepatocellular carcinoma (HCC). In HCC, telomerase reactivation allows liver cells with damaged DNA and shortened telomeres to bypass cell death and continue dividing, further driving cancer progression (Hepatocellular carcinoma pathogenesis: from genes to environment, Nature).



Symptoms of HCC

Hepatocellular carcinoma (HCC), the most common type of primary liver cancer, often presents with vague symptoms that can be easily overlooked or mistaken for other liver diseases. Key symptoms include:

- **Unexplained weight loss** and **loss of appetite**
- **Upper abdominal pain**, which may be accompanied by a swollen or tender abdomen

- **Nausea and vomiting**
- **Fatigue and weakness**
- **Jaundice** (yellowing of the skin and eyes)
- **Swelling** in the abdomen or legs due to fluid accumulation These symptoms often appear in later stages of HCC, making early detection challenging.

Diagnosis of HCC

Diagnosing HCC involves a combination of blood tests, imaging techniques, and sometimes biopsy. Common diagnostic steps include:

1. **Blood tests**: Alpha-fetoprotein (AFP) levels, which can be elevated in HCC, although this test alone isn't definitive.
2. **Imaging**: Techniques like ultrasound, CT scans, and MRIs are used to identify liver masses, determine tumor size, and assess spread.
3. **Liver biopsy**: In cases where the diagnosis is uncertain, a biopsy may be done to confirm HCC by examining liver tissue under a microscope. Regular screening in high-risk individuals (such as those with chronic hepatitis or cirrhosis) is crucial for early detection.

Treatments for HCC

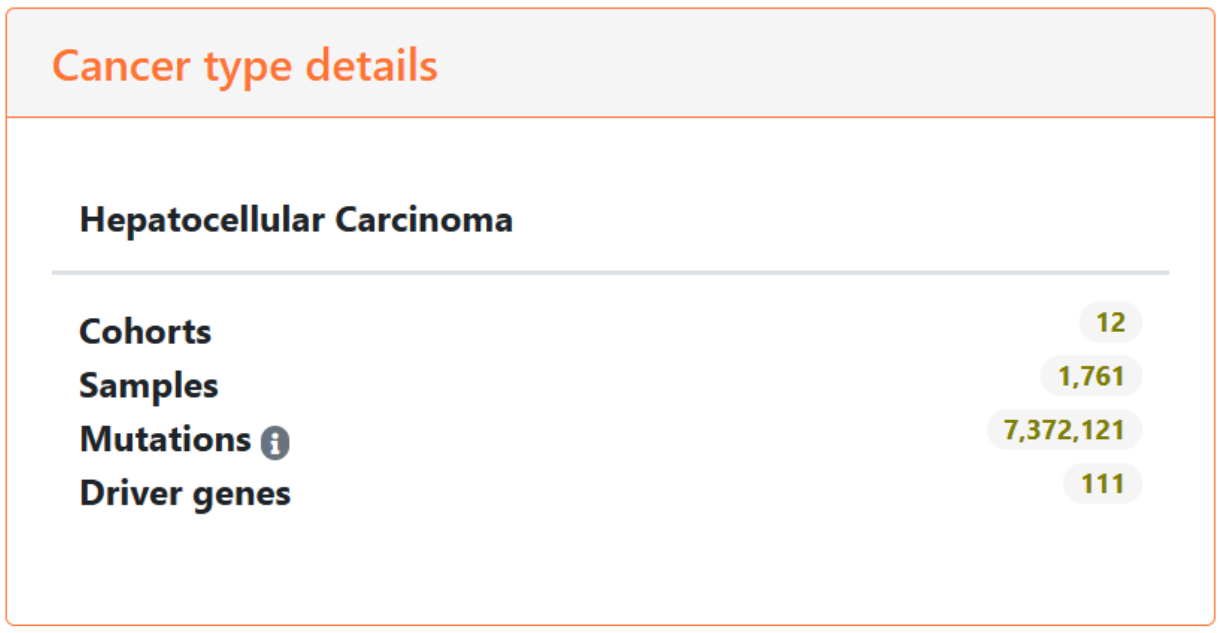
Treatment for HCC depends on factors like cancer stage, liver function, and overall patient health. Main treatment options include:

- **Surgery**: Surgical resection or liver transplant can offer a potential cure, especially in early-stage HCC.
- **Ablation therapy**: Techniques like radiofrequency ablation (RFA) and microwave ablation are used to destroy cancer cells directly within the liver.
- **Transarterial therapies**: Embolization and chemoembolization target the blood supply to the tumor, starving it of nutrients and slowing growth.
- **Systemic therapies**: Targeted drugs (like sorafenib or lenvatinib) and immunotherapies (like nivolumab) are used to treat advanced HCC.
- **Radiation therapy**: Selective internal radiation therapy (SIRT) may be used for some patients to shrink tumors. New therapies and combinations, especially involving immunotherapy, are being researched to improve survival outcomes for advanced HCC. Early detection is key to maximizing treatment effectiveness.

Task 1: Analyzing Mutations and Signaling Defects of a Cancer

Objective: To analyze mutation signatures of prominent genes associated with HCC, identify mutation types and their respective genomic domains, and assess the implications of these mutations on disease progression.

Summary of data collected from Intogen for HCC:



Intogen: https://www.intogen.org/search?cancer=HCC

Cohorts: There are 12 cohorts listed, each representing a different study or dataset for HCC. A cohort generally consists of samples grouped based on specific parameters like source, methodology, or patient characteristics.
Samples: There are 1,761 samples across all cohorts. These samples represent individual HCC cases or patient data used in the studies.
Mutations: This field shows a large number (7,372,121 mutations), indicating the total number of genetic mutations detected across all samples in the listed HCC studies.

Driver Genes: A total of 111 driver genes have been identified. Driver genes are those whose mutations contribute to the initiation or progression of HCC. Identifying these genes is crucial for understanding the mechanisms underlying cancer and for developing targeted therapies.

Cohort	Source	Age	Type	Sample size
TCGA_WXS_HCC	TCGA/PanCatAtlas, phs000178	A	P	360
PCAWG_WGS_LIVER_HCC	PCAWG	A	P	314
CBIOP_WXS_HCC_AMC_PRV	Asian Medical Center (Korea)	A	P	231
ICGC_WXS_HCC_LINC_JP_2019	ICGC Data Portal (LINC-JP)	A	P	213
ICGC_WXS_HCC_LICA_FR_2019	ICGC Data Portal (LICA-FR)	A	P	211
ICGC_WXS_HCC_LICA_CN_VARSCAN_2019	ICGC Data Portal (CN-VARSCAN)	A	P	175
ICGC_WXS_HCC_LICA_CN_STRELKA_2019	ICGC Data Portal (LICA-CN)	A	P	115
HARTWIG_WGS_HCC_2023	Hartwig Medical Foundation	A	M	52
ICGC_WGS_HCC_LICA_FR_2019	ICGC Data Portal (LICA-FR)	A	P	32
ICGC_WXS_HCC_LIAD_FR_2019	ICGC Data Portal (LIAD-FR)	A	P	30
ICGC_WGS_HCC_LICA_CN_VARSCAN_2019	ICGC Data Portal (CN-VARSCAN)	A	P	24
ICGC_WXS_HCC_LIHM_FR_2019	ICGC Data Portal (LIHM-FR)	A	P	4

Showing 1 to 12 of 12 entries

Previous 1 Next

Intogen: https://www.intogen.org/search?cancer=HCC

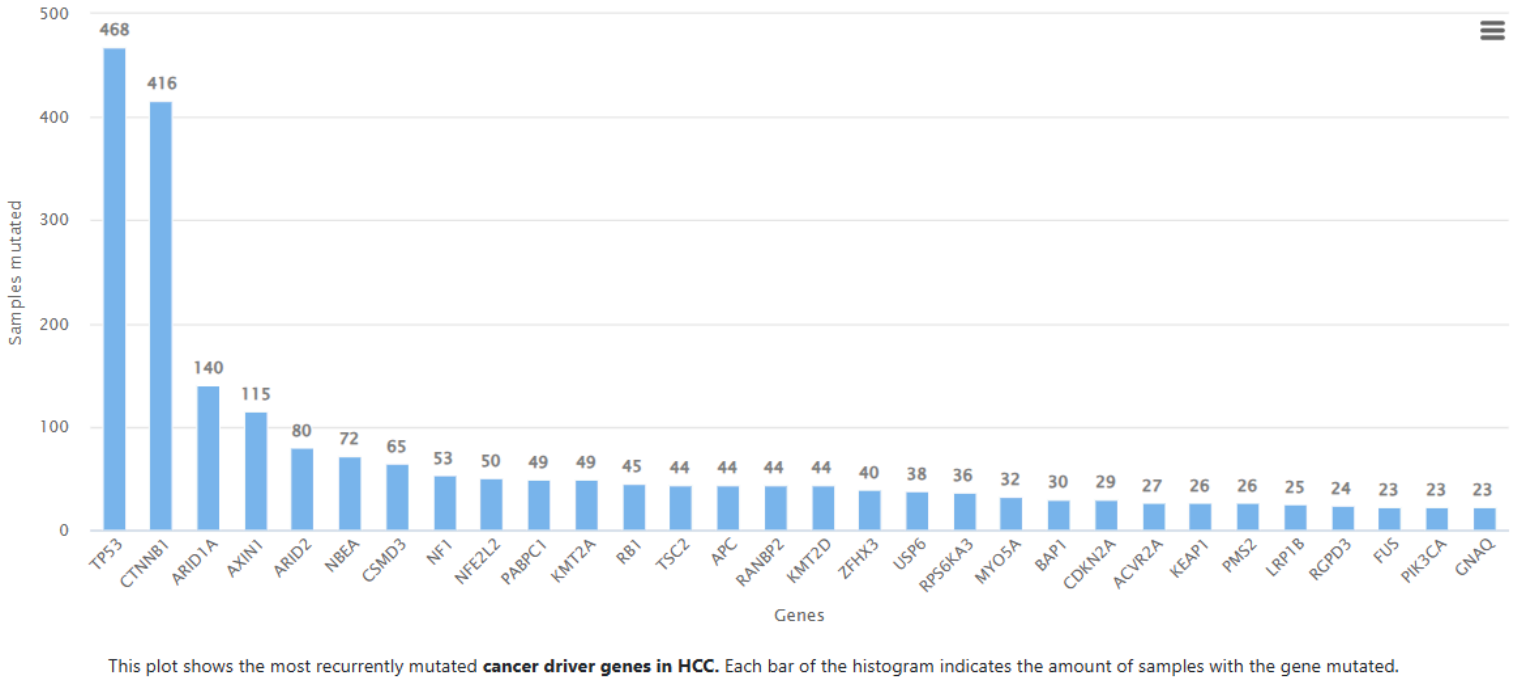
This table lists individual cohorts, each providing more specific information:

- **Cohort Name**: Identifies the cohort or study. For example, "TCGA_WXS_HCC" represents data from The Cancer Genome Atlas for Whole Exome Sequencing of HCC.
- **Source**: Indicates the organization or data portal that provided the data, such as TCGA, PCAWG, ICGC, or specific medical institutions.
- **Age and Type**: Age (A) represents adult samples and age (P) represents pediatric samples. While the type (P for primary tumor) and type (M for metastatic tumor) specifies the nature of the samples.
- **Sample Size**: Shows the number of samples in each cohort, with some cohorts having a larger sample size (e.g., 360 for TCGA_WXS_HCC) and others with smaller numbers (e.g., 4 for ICGC_WXS_HCC_LIHM_FR).

Key Interpretations on the samples and data collected from patients:

- The data is drawn from a variety of sources, including **TCGA**, **PCAWG**, and several cohorts within the **ICGC Data Portal**, indicating a well-rounded, international dataset. This variety is crucial for studying HCC across different populations and genetic backgrounds.
- Most cohorts list the **Type as "P" (Primary tumor).** Only one cohort from the Hartwig Medical Foundation (HARTWIG_WGS_HCC_2023) includes samples marked as "M" (Metastatic), which could provide insights into how HCC spreads or differs at the metastatic stage.
- All samples are from **adult (A) cases**, which aligns with HCC's prevalence, as it's more common in adults, often due to underlying risk factors like viral hepatitis and cirrhosis.
- The presence of recent datasets, such as **HARTWIG_WGS_HCC_2023**, suggests that newer studies continue to be added, which can help track evolving trends and potentially discover new therapeutic targets or diagnostic markers for HCC.

Summary of key genes mutated in HCC from Intogen:



This plot shows the most recurrently mutated **cancer driver genes in HCC.** Each bar of the histogram indicates the amount of samples with the gene mutated.

Intogen: https://www.intogen.org/search?cancer=HCC

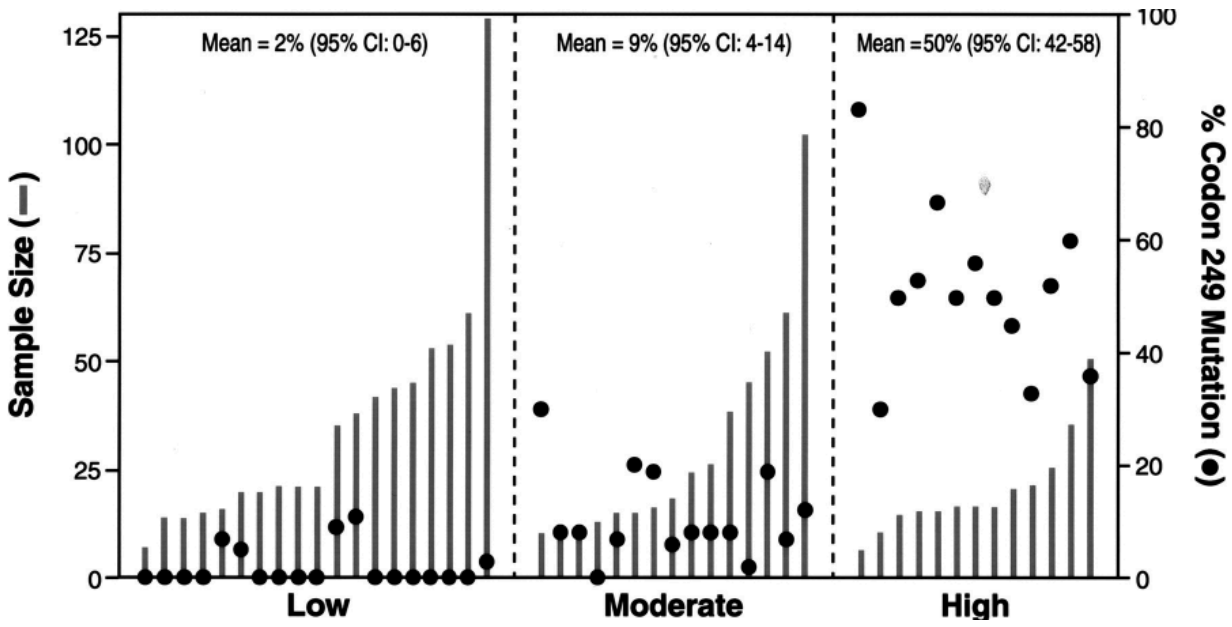
The graph shows the frequency of mutations in various genes associated with Hepatocellular Carcinoma (HCC), measured by the number of samples in which each gene is mutated. From this chart, we observe that the mutation frequency decreases significantly as we move from left to right.

Key Genes Mutated in HCC

These genes were selected based on mutation frequency and relevance to pathways commonly altered in HCC. The following genes were identified as highly mutated in HCC:

1. TP53 – Mutated in 468 samples:

The p53 protein functions as a guardian of the genome, responding to cellular stress by inducing cell cycle arrest or apoptosis when DNA damage occurs. When TP53 is mutated, its ability to perform these functions is compromised, leading to uncontrolled cell proliferation and increased tumorigenesis. Specifically, **mutations at codon 249 of TP53** are frequently observed in HCC cases associated with **aflatoxin exposure**, particularly the **R249S mutation**, which results from a **G:C to T:A transversion induced by aflatoxin B1 (AFB1) exposure** (TP53 mutations and hepatocellular carcinoma: insights into the etiology and pathogenesis of liver cancer, PubMed).



Frequency of p53 249ser mutation and sample size by aflatoxin level. Bars, sample size of each study referred to the left vertical axis. •, observed prevalence of each mutation in each study referred to the right vertical axis. The horizontal axis orders each study contributing information on p53 249ser mutations from smallest to largest in sample size within each aflatoxin level. The mean frequency (and CI) were derived from our meta-analysis.

Source: https://aacrjournals.org/cebp/article/10/6/617/164873/Hepatitis-B-Aflatoxin-B1-and-p53-Codon-249

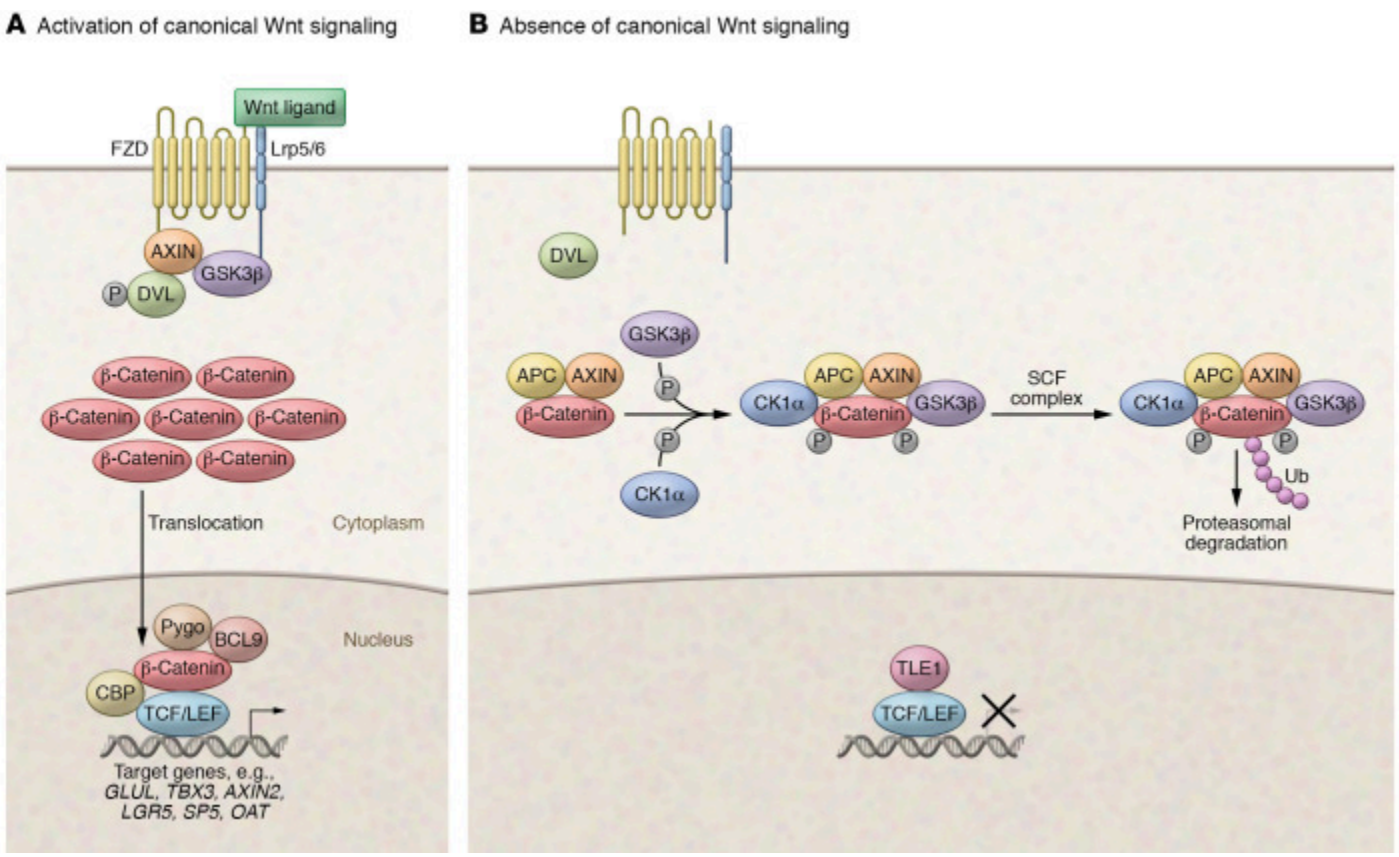
Aflatoxins are toxic compounds produced by certain molds found in food, and they are known carcinogens. The presence of aflatoxin in the diet significantly increases the risk of developing HCC, especially in regions where dietary exposure is high. Studies have shown that approximately 36% of tumors from areas with high aflatoxin exposure exhibit the R249S mutation (Hepatitis B, Aflatoxin B1, and p53 Codon 249 Mutation in Hepatocellular Carcinomas from Guangxi, People's Republic of China, and a Meta-analysis of Existing Studies, AACR Journals). Furthermore, HBV infection appears to exacerbate this risk; it has been suggested that HBV may enhance the mutagenic effects of aflatoxins through

mechanisms such as oxidative stress and interference with DNA repair processes (TP53 mutations and hepatocellular carcinoma: insights into the etiology and pathogenesis of liver cancer, PubMed).

Research indicates a synergistic relationship between HBV infection and aflatoxin exposure in promoting TP53 mutations. For instance, studies have demonstrated that HCC patients who are both HBV-positive and exposed to aflatoxins have a higher prevalence of the R249S mutation compared to those without HBV. This suggests that while aflatoxins can induce mutations independently, their effects are magnified in the presence of HBV (Analysis of TP53 aflatoxin signature mutation in hepatocellular carcinomas from Guatemala: A cross-sectional study (2016-2017), PubMed).

2. CTNNB1– Mutated in 416 samples:

CTNNB1 encodes β-catenin, a crucial protein in the Wnt/β-catenin signaling pathway, which plays a significant role in regulating cell growth, differentiation, and the maintenance of stem cell properties. Under normal physiological conditions, β-catenin levels are tightly controlled; when not needed, it is targeted for degradation. However, mutations in the CTNNB1 gene can lead to the stabilization and accumulation of β-catenin in the nucleus, where it activates genes that promote cell proliferation and survival (CTNNB1 gene, medlineplus).



In the **Wnt signaling pathway**, when **Wnt ligands** bind to receptors (FZD), it triggers the **phosphorylation** of a protein called **DVL**. Phosphorylated DVL recruits **AXIN** and **GSK3β** to the cell membrane, preventing the formation of a complex that would normally degrade **β-catenin**. This allows **β-catenin** to accumulate, move to the nucleus, and activate gene transcription by binding to **TCF/LEF** transcription factors.

Without **Wnt ligands**, **β-catenin** is tagged for degradation by a complex involving **GSK3β, CK1α, APC,** and **AXIN1**. After tagging, **β-catenin** is broken down by the proteasome. When **β-catenin** is absent in the nucleus, **TCF/LEF** is repressed by **TLE-1**, preventing gene activation. In human **HCC**, mutations are often found in genes like **CTNNB1** (27%), **AXIN1** (8%), and **APC** (3%), which affect this pathway.

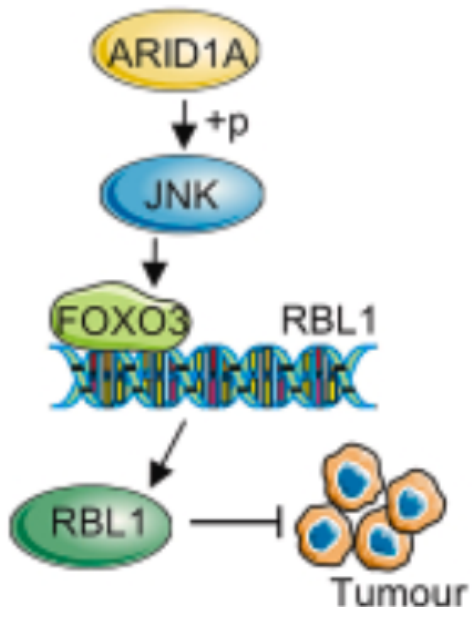
In hepatocellular carcinoma (HCC), dysregulation of the Wnt/β-catenin pathway is a prevalent oncogenic event. Mutations in CTNNB1 disrupt the normal degradation process of β-catenin, resulting in its constitutive activation and subsequent promotion of uncontrolled cell growth. This contributes to tumor initiation and enhances the stemness of liver cancer cells. Specifically, CTNNB1 mutations are frequently found in HCC cases with non-viral etiologies, such as alcohol abuse and non-alcoholic fatty liver disease (NAFLD) (β-Catenin signaling in hepatocellular carcinoma, PuMed).

Research indicates that CTNNB1 mutations occur in approximately 27% of HCC patients, with many mutations affecting serine/threonine sites in exon 3. These mutations prevent β-catenin from undergoing phosphorylation and degradation, leading to its accumulation and unregulated transcriptional activity (β-Catenin signaling in hepatocellular carcinoma, PuMed).

3. ARID1A – Mutated in 140 samples:

ARID1A (AT-rich interaction domain 1A) is a critical gene involved in the regulation of chromatin structure and gene expression, primarily functioning as a tumor suppressor. Its role in hepatocellular carcinoma (HCC) is multifaceted, particularly through its interactions with the JNK (c-Jun N-terminal kinase) signaling pathway.

ARID1A mutations are prevalent in HCC, detected in approximately **9% to 17% of cases,** primarily manifesting as loss-of-function mutations that lead to decreased ARID1A expression. This deficiency correlates with poor overall survival (OS) and relapse-free survival (RFS) among patients. Specifically, lower ARID1A expression is associated with larger tumor sizes and worse prognostic outcomes. The **tumor suppressive role** of ARID1A in HCC is partly **mediated through its regulation of downstream targets involved in the JNK pathway**. Research indicates that ARID1A inhibits HCC progression by modulating the expression of **retinoblastoma-like 1 (RBL1) via the JNK/FOXO3 pathway** (The somatic mutational landscape and role of the ARID1A gene in hepatocellular carcinoma, Pubmed).



Schematic diagram of the mechanism by which ARID1A regulates HCC progression through the JNK/FOXO3/RBL1 axis.

In the context of hepatitis B virus-associated HCC (HBV-HCC), ARID1A deficiency has been linked to increased tumor mutational burden (TMB) and heightened immune activity, particularly involving immune checkpoint proteins like TIM-3. This suggests that ARID1A may also play a role in modulating the immune landscape of tumors, potentially impacting responses to immunotherapy.

4. AXIN1 – Mutated in 115 samples:

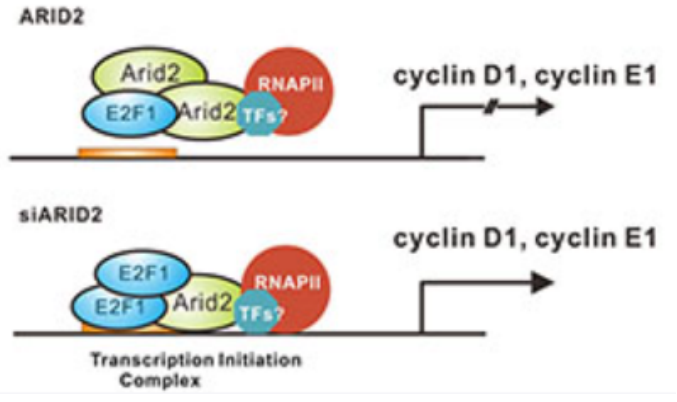
AXIN1 is a critical component of the Wnt/β-catenin signaling pathway, functioning primarily as a scaffold protein within the β-catenin destruction complex. This complex, which includes proteins such as APC (adenomatous polyposis coli), GSK-3β (glycogen synthase kinase 3 beta), and CK1 (casein kinase 1), is essential for the regulation of β-catenin levels in the cell. Under normal circumstances, AXIN1 helps facilitate the phosphorylation and subsequent degradation of β-catenin, preventing its accumulation and the excessive activation of Wnt signaling pathways, which can lead to uncontrolled cell proliferation and tumorigenesis (Unraveling the impact of AXIN1 mutations on HCC development: Insights from CRISPR/Cas9 repaired AXIN1-mutant liver cancer cell lines, Pubmed)

AXIN1 mutations in HCC are associated with the activation of the Wnt/β-catenin pathway, contributing to cancer development. When AXIN1 is mutated, β-catenin escapes degradation, similar to the effect of CTNNB1 mutations. This leads to enhanced cell growth and survival. Unlike CTNNB1, which often has activating mutations, AXIN1 mutations are typically inactivating, disrupting the tumor-suppressive function of the β-catenin destruction complex. Mutations in AXIN1 are more frequently observed in viral-associated HCC, such as cases linked to hepatitis B and C infections. By impairing Wnt pathway regulation, AXIN1 mutations create an environment that supports cancer cell survival and growth in the liver. In HCC, AXIN1 mutations are found in approximately 8-10% of cases, often associated with viral infections like hepatitis B and C.

5. ARID2 – Mutated in 80 samples:

ARID2 (AT-rich interactive domain 2) is recognized as a tumor suppressor gene that plays a significant role in the progression of hepatocellular carcinoma (HCC). Its involvement primarily affects the Rb-E2F signaling pathway, which is crucial for regulating cell cycle progression.

ARID2 expression is significantly downregulated in HCC tissues compared to adjacent non-tumorous tissues. This downregulation correlates with increased cell proliferation and tumor growth in hepatoma cells. ARID2 inhibits the transition from G1 to S phase of the cell cycle by targeting cyclins, particularly cyclin D1 and cyclin E1. This inhibition occurs through the repression of E2F transcription factors, which are critical for the expression of genes that promote cell cycle progression. Restoration of ARID2 expression in hepatoma cells has been shown to suppress tumor growth, while its knockdown enhances cellular proliferation and tumorigenicity (Chromatin remodeling gene ARID2 targets cyclin D1 and cyclin E1 to suppress hepatoma cell progression, Oncotarget).

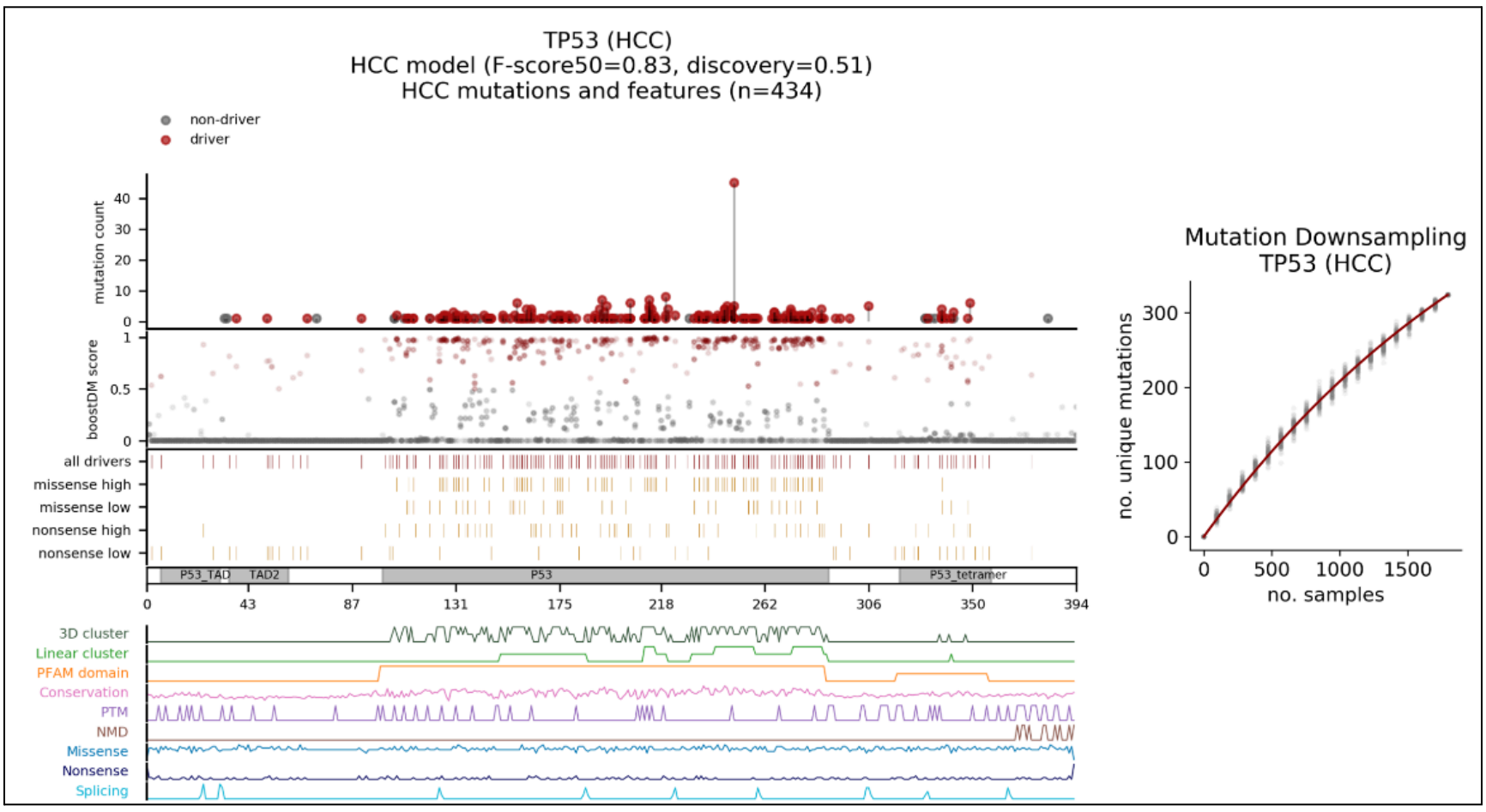


A schematic model of ARID2-mediated repression of cyclin D1 and cyclin E1. Data indicate that ARID2 physically interacts with E2F1and induces dissociation of activator E2F1 and RNA Pol II as well as histone deacetylation, contributing to the transcriptional repression of cyclin D1 and cyclin E1. However, ARID2 knockdown enhances binding of activator E2F1 and induces active histone modification, thus promoting cyclin D1 and cyclin E1 transcription.

Identification and analysis of mutation signatures in the most prominent genes associated with HCC

1. TP53

In-silico saturation mutagenesis



This image shows an analysis of how mutations in the TP53 gene are linked to liver cancer (Hepatocellular Carcinoma, or HCC).

Top Graph (Mutation Count): The graph shows where mutations happen along the TP53 gene and how often. Each dot represents a mutation at a specific spot in the gene.

- Red Dots are "driver" mutations, which means they actively help cancer grow.
- Gray Dots are "passenger" mutations, which happen by chance and don’t help the cancer.

X-axis (Gene Layout): This shows the TP53 gene from start to end, with key sections labeled (Domains):

- **TAD1** & **TAD2** are parts that help TP53 do its job as a gene protector.
- The DNA-binding domain **(P53)** is where TP53 connects to DNA to control other genes. This area has a lot of red dots, meaning it’s a common spot for cancer-causing mutations.
- The tetramerization domain **(P53_tetramer)** is another part critical for TP53’s function.

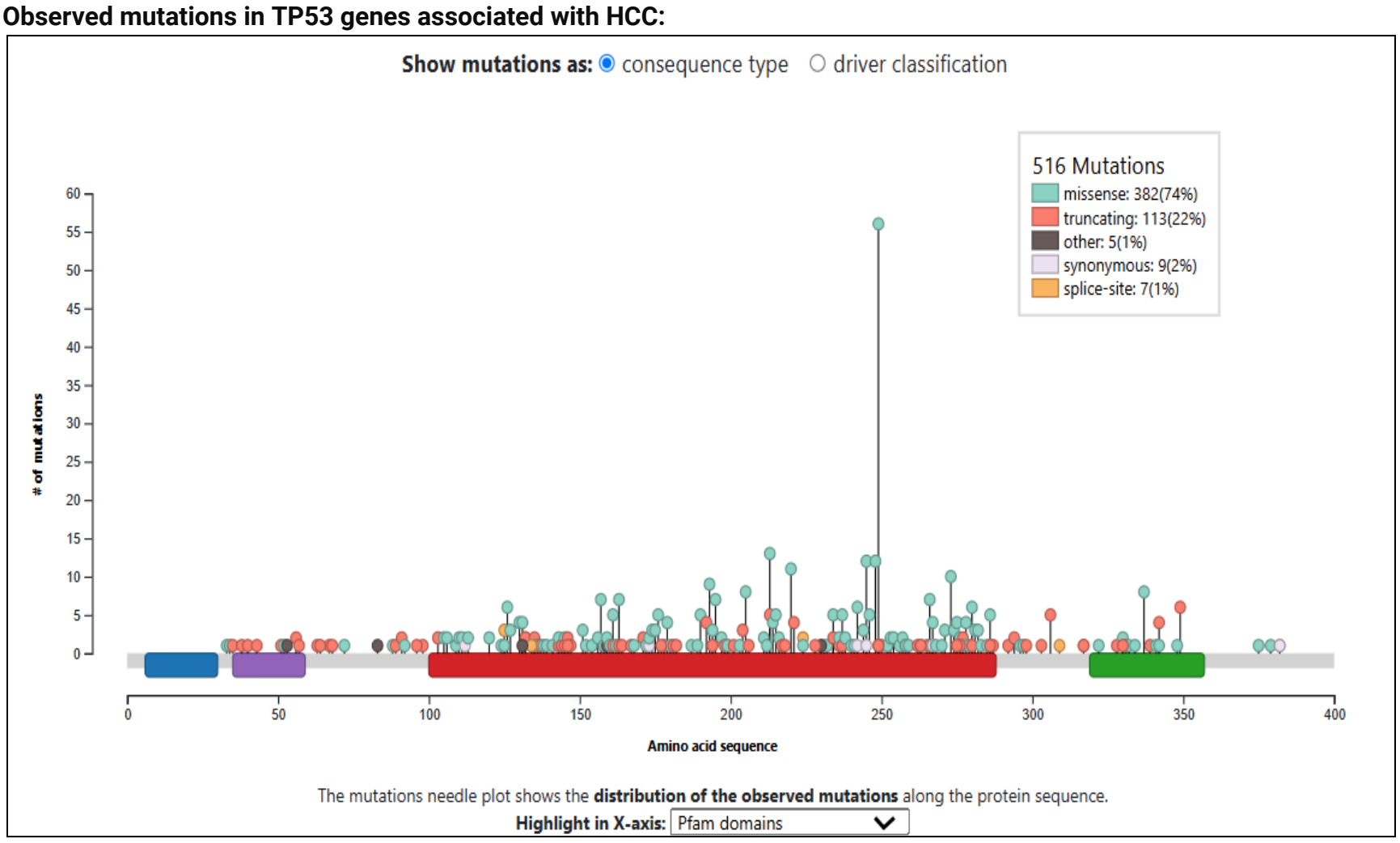
Other Tracks: There are extra rows showing details about **mutations**:

- 3D clusters (where mutations tend to group in space).
- Conservation (important spots that are similar in many species and are likely vital).
- Missense/Nonsense Mutations (types of changes in the protein that could mess up TP53's function).

Mutation Discovery Curve: This curve shows how many unique mutations are found as more samples (cancer cases) are studied.

Saturation: The curve leveling off means that as they study more samples, they're finding fewer new mutations. This suggests they've captured most of the mutations in TP53 linked to liver cancer.

Observed mutations in TP53 genes associated with HCC:



This image is a mutation needle plot that displays the types and distribution of mutations observed along the TP53 protein sequence in HCC.

X-axis (Amino Acid Sequence)

- The **X-axis represents the amino acid sequence of the TP53 protein**, showing positions from start to end.
- The colored blocks along the **X-axis represent Pfam domains,** or specific functional regions within the TP53 protein:
  - **Domains:** Blue (P53_TAD), purple (TAD), red (P53), and green (P53_tetramer) blocks show regions with particular functions or structural roles in TP53.
  - The red region around positions **100 to 300 is the DNA-binding domain of TP53**, which is crucial for its function as a tumor suppressor.

Y-axis (No. of Mutations)

The Y-axis shows the number of mutations observed at each position in the TP53 protein. Higher bars mean more mutations at that specific position.

- **Missense mutations (light blue):** These mutations (74% of total) change one amino acid to another. Many are clustered in the DNA-binding region, indicating that changes here may significantly affect TP53's ability to suppress tumors.
- **Truncating mutations (red):** These mutations (22% of total) create stop signals in the sequence, shortening the protein. They are spread out but also appear in the DNA-binding domain, likely disrupting TP53's function.
- **Other mutations (dark gray)**, synonymous mutations (pink), and splice-site mutations (yellow): These are less common (combined 4%), with synonymous mutations not affecting the amino acid sequence and splice-site mutations potentially affecting protein assembly.

Most mutations are clustered in the DNA-binding domain (red block) of TP53, highlighting this as a critical region for cancer-related mutations. The high number of missense mutations in this region suggests that even single amino acid changes here can interfere with TP53's tumor-suppressing activity.

Top 5 Mutations found in TP53 in HCC and whether are related to disease progression:

Show mutations as: ◉ consequence type ○ driver classification

Show 10 entries

Search:

Mutation (GRCh38)	Protein Position	Samples	Samples (%)	Consequence
17:7674216:C>A	249	48	2.73	missense_variant
17:7674872:T>C	220	8	0.45	missense_variant
17:7674893:C>A	213	7	0.4	missense_variant
17:7674953:T>C	193	7	0.4	missense_variant
17:7670664:C>A	349	6	0.34	stop_gained

This table provides data on observed mutations in the TP53 gene from hepatocellular carcinoma (HCC) samples, as sourced from the IntOGen platform.

- The **mutation at protein position 249** is the most common, occurring in **2.73% of samples**, which indicates it may play a significant role in HCC pathogenesis. Since it's a missense variant, it could potentially disrupt the TP53 protein's structure and function without completely halting its production.
- The **stop-gained mutation at position 349**, while less common (0.34% of samples), results in a truncated protein. This could have a severe impact on TP53's function since a shorter protein may lack critical regions required for its tumor-suppressing activity.

Overall, mutations like those in TP53 are significant in cancer research because TP53 is a well-known tumor suppressor gene, and changes to it can impact cell cycle regulation, apoptosis, and DNA repair processes. The relatively high frequency of some mutations, like the one at position 249, suggests that these specific changes could be essential targets for understanding and potentially treating HCC.

Show mutations as: ○ consequence type ◉ driver classification

Show 10 entries

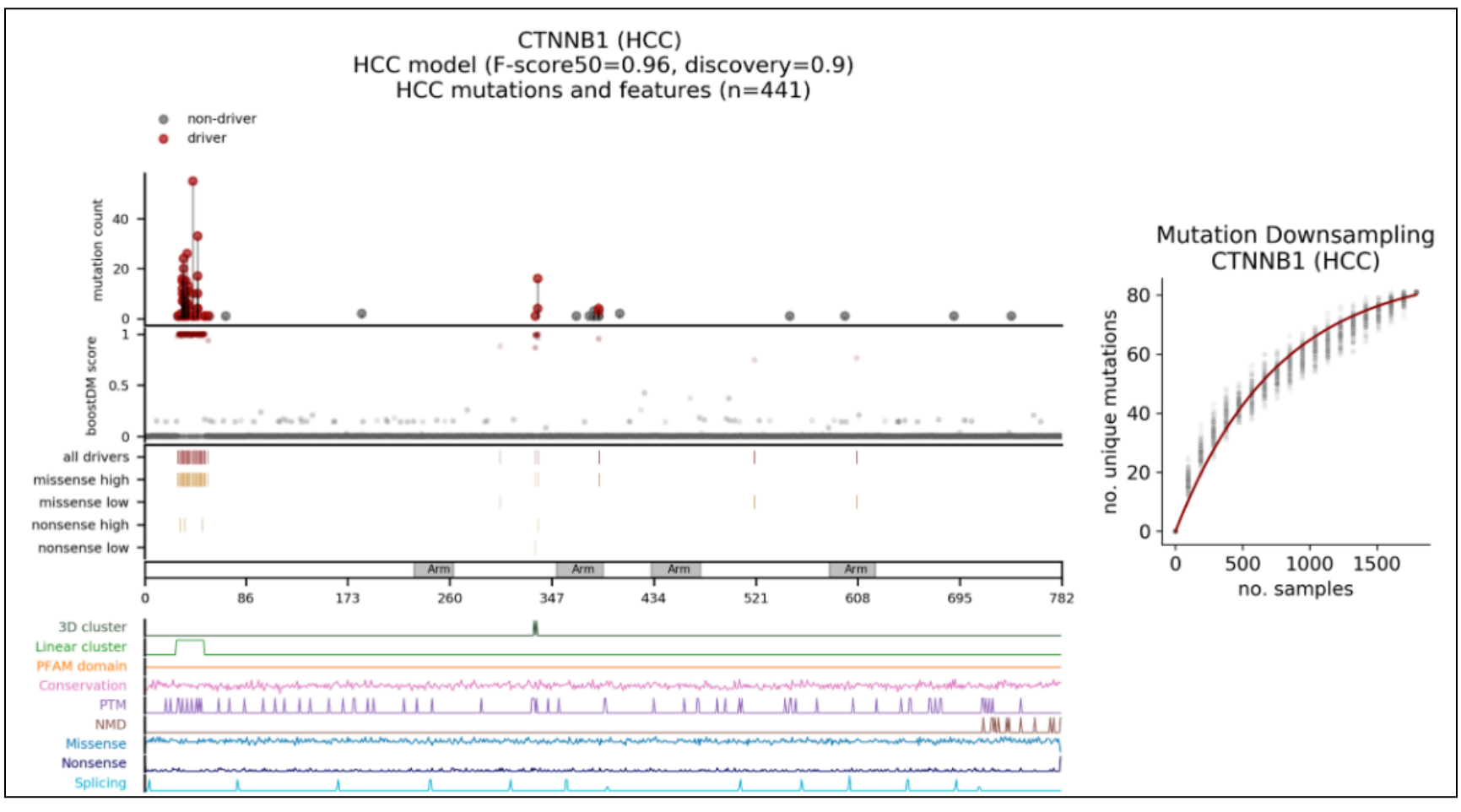
Search:

Mutation (GRCh38)	Protein Position	Samples	Samples (%)	Driver	Driver score
17:7674216:C>A	249	48	2.73	Driver	1
17:7674872:T>C	220	8	0.45	Driver	1
17:7674893:C>A	213	7	0.4	Driver	1
17:7674953:T>C	193	7	0.4	Driver	1
17:7670664:C>A	349	6	0.34	Driver	1

This table shows the type of mutation, whether it's a driver or passenger mutation. However, in this case, all the top five most prevalent mutations in TP53 are driver mutations. That is, these mutations result in changes in DNA sequences that cause cells to become cancerous and spread in the body.

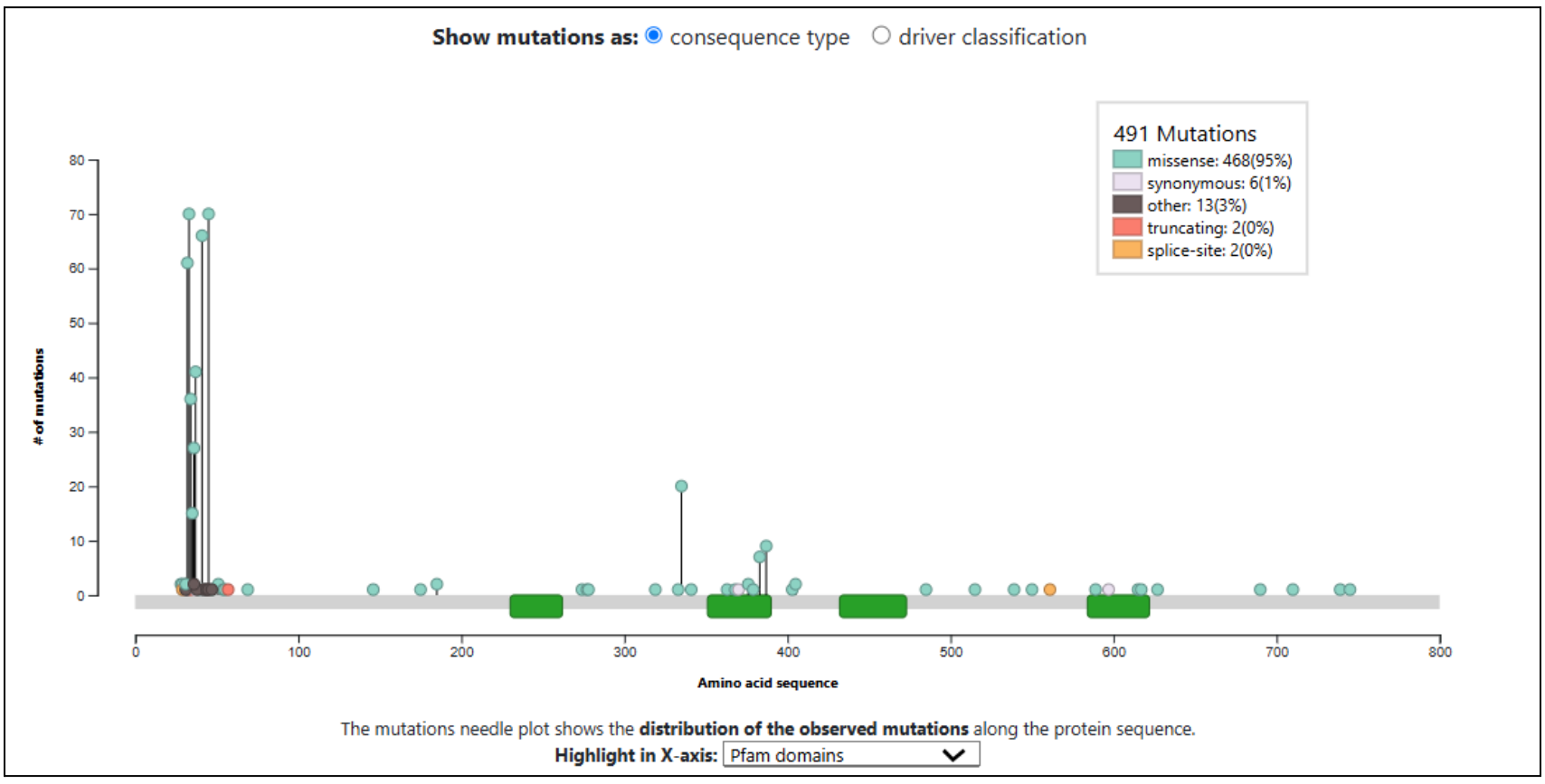
2. CTNNB1

In-silico saturation mutagenesis



The plot includes information on **Pfam domains**, **conservation scores**, and **post-translational modifications (PTMs)**. High mutation frequencies near conserved or functionally significant domains suggest these mutations may disrupt protein function.

Observed mutations in CTNNB1 genes associated with HCC:



Mutation Distribution and Frequency:

- The first plot shows mutation counts across the protein sequence. Peaks indicate regions with higher mutation frequencies, with significant clustering near the N-terminal region.
- The data indicate high-frequency mutations at specific amino acid positions, with **position 41** showing the highest number of mutations (3.12% of samples), followed by positions **45, 36, 33**, and **another at 33**.
- The mutations at these positions are primarily **missense mutations**, suggesting that they alter the amino acid sequence without completely truncating the protein.

Top 5 Mutations found in CTNNB1 in HCC and whether are related to disease progression:

Show mutations as: consequence type driver classification

Show 10 entries Search:

Mutation (GRCh38)	Protein Position	Samples	Samples (%)	Consequence
3:41224633:A>G	41	55	3.12	missense_variant
3:41224645:T>C	45	33	1.87	missense_variant
3:41224619:A>C	36	26	1.48	missense_variant
3:41224610:C>G	33	25	1.42	missense_variant
3:41224609:T>C	33	20	1.14	missense_variant

Mutation Consequence Types:

- The second plot shows mutations categorized by consequence types, including **missense, synonymous, truncating, and splice-site variants**. The majority of mutations are **missense (95%)**, with a few truncating (2%) and splice-site mutations (20%).
- The presence of truncating and splice-site mutations, though less common, could result in more drastic effects on protein function, possibly leading to loss of function or altered signaling pathways.

Show mutations as: ○ consequence type ◉ driver classification

Show 10 entries

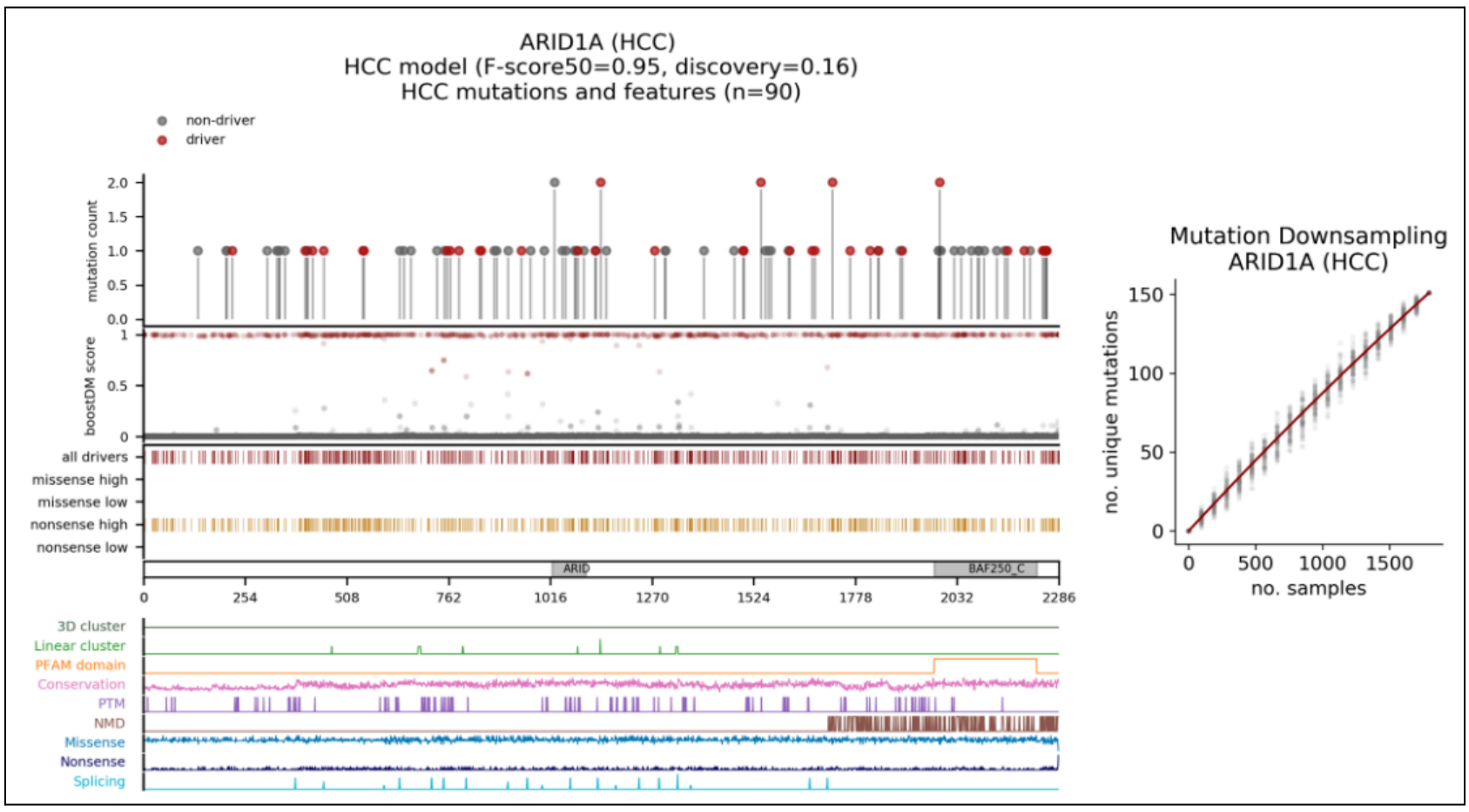
Search:

Mutation (GRCh38)	Protein Position	Samples	Samples (%)	Driver	Driver score
3:41224633:A>G	41	55	3.12	Driver	1
3:41224645:T>C	45	33	1.87	Driver	1
3:41224619:A>C	36	26	1.48	Driver	1
3:41224610:C>G	33	25	1.42	Driver	1
3:41224609:T>C	33	20	1.14	Driver	1

In the last table, the top five mutations (based on driver scores) all have a **driver score of 1**, indicating their potential importance in driving HCC. Positions like **41 and 45** are highlighted as recurrent driver mutation sites.

3. ARID1A

In-silico saturation mutagenesis

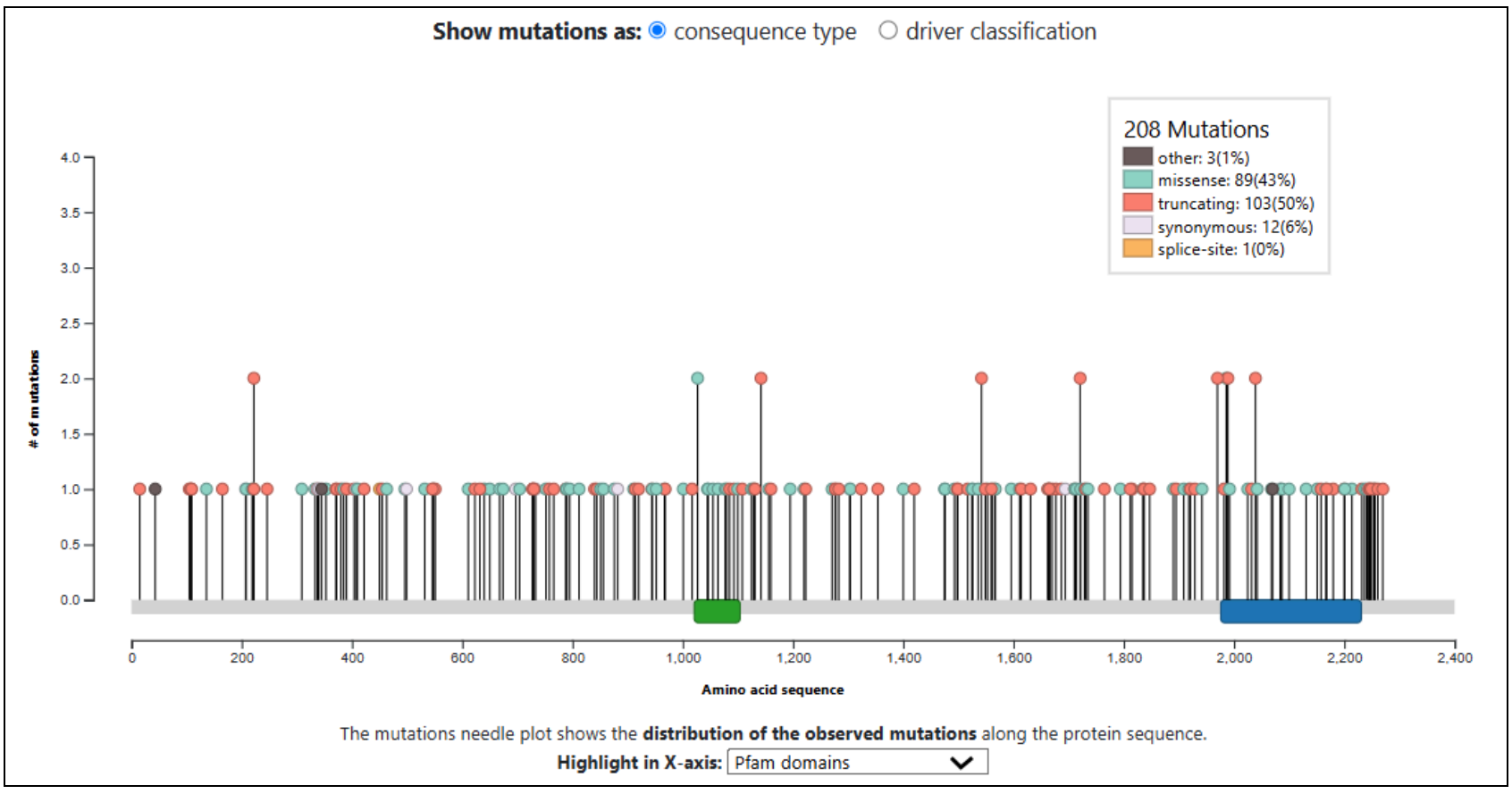


This plot shows the distribution of mutations across the ARID1A gene in HCC.

Driver mutations are spread throughout the gene, but some regions show higher concentrations of mutations, suggesting these may be hotspots important in HCC progression.

The boostDM score in the chart provides a probability that a mutation is a driver. Higher scores align with regions containing many driver mutations, suggesting that these mutations likely impair ARID1A's tumor-suppressing function. ARID1A mutations in these regions might lead to a loss of function, which enables unchecked cellular growth—a hallmark of cancer.

Observed mutations in ARID1A genes associated with HCC:



This image contains a "needle plot" showing the distribution and classification of mutations by consequence type (e.g., missense, truncating, synonymous) in the ARID1A gene. Truncating mutations (often creating stop codons) are particularly abundant, which may result in loss-of-function alterations in ARID1A, commonly associated with tumor suppressor genes.

Some mutations are more frequent across samples, as indicated by the height of the needles. These high-frequency mutations could represent recurrent alterations crucial to HCC pathogenesis.

Top 5 Mutations found in ARID1A in HCC and whether are related to disease progression:

Show mutations as: consequence type driver classification

Show 10 entries Search:

Mutation (GRCh38)	Protein Position	Samples	Samples (%)	Consequence
1:26766455:A>T	-	2	0.11	splice_acceptor_variant
1:26767881:A>G	1027	2	0.11	missense_variant
1:26772517:C>T	1142	2	0.11	stop_gained
1:26774851:G>T	1542	2	0.11	stop_gained
1:26779059:C>T	1721	2	0.11	stop_gained

Show mutations as: consequence type driver classification

Show 10 entries Search:

Mutation (GRCh38)	Protein Position	Samples	Samples (%)	Driver	Driver score
1:26766455:A>T	-	2	0.11	Driver	0.98
1:26767881:A>G	1027	2	0.11	Passenger	0.03
1:26772517:C>T	1142	2	0.11	Driver	1
1:26774851:G>T	1542	2	0.11	Driver	1
1:26779059:C>T	1721	2	0.11	Driver	0.99

High Incidence of Truncating and Missense Mutations:

ARID1A mutations in HCC are mostly truncating (50%) or missense (48%).

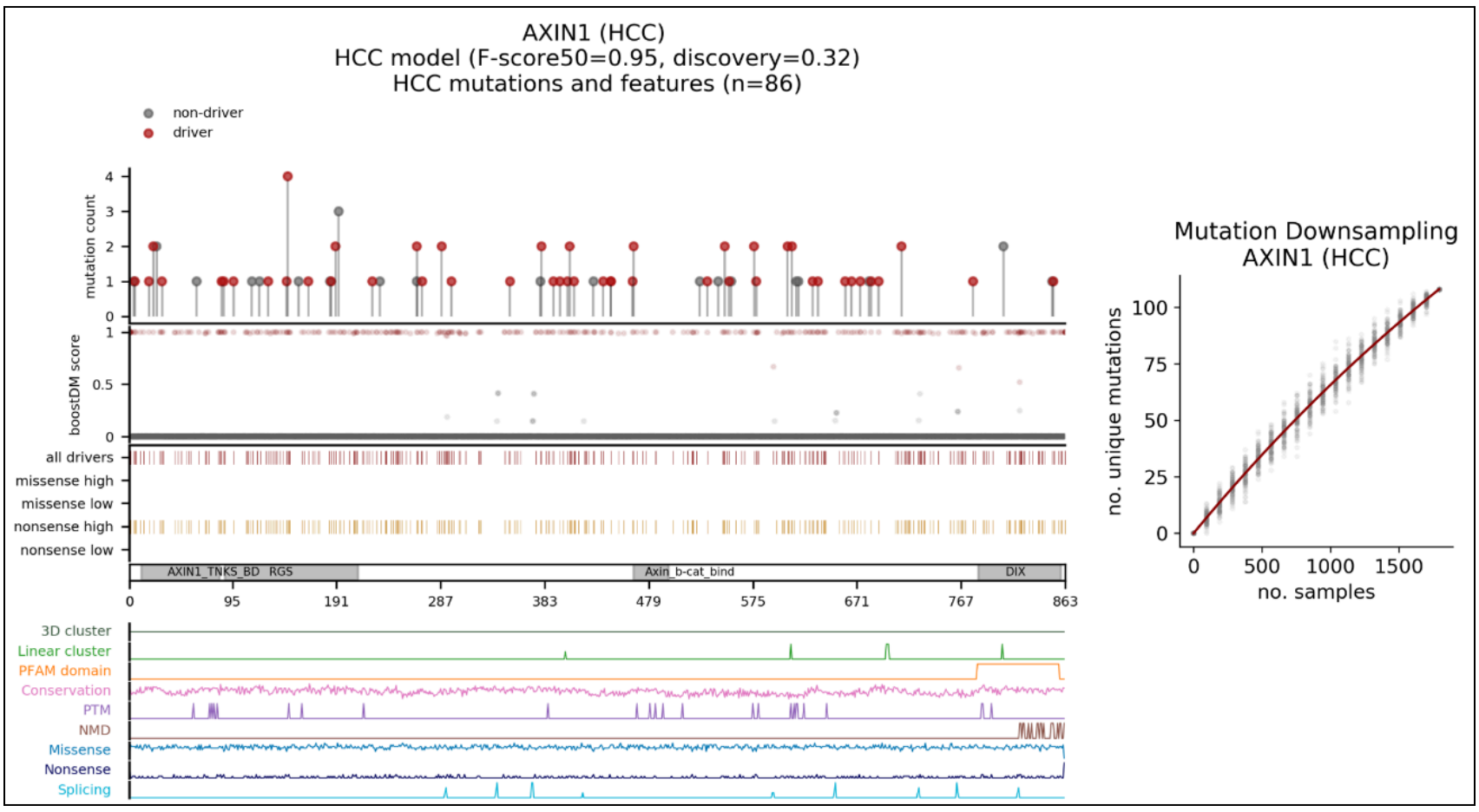
- High Driver Scores for Key Mutations:
  - Mutations such as those at positions 1142, 1542, and 1721 have high driver scores (close to or at 1), meaning they have a strong likelihood of contributing to HCC progression.
  - These mutations often introduce stop codons or missense variants that likely disrupt ARID1A's tumor-suppressing functions. For example, stop-gain mutations truncate the protein, preventing it from participating in chromatin remodeling.
- The "Passenger" mutation in the list (position 1027, with a low driver score of 0.03) likely does not impact ARID1A function in a way that promotes HCC. Understanding which mutations are passengers helps focus on mutations more relevant to cancer development.

Some mutations fall within annotated PFAM domains, which represent conserved functional regions of ARID1A. Mutations in these domains are particularly harmful because they likely impact essential interactions necessary for

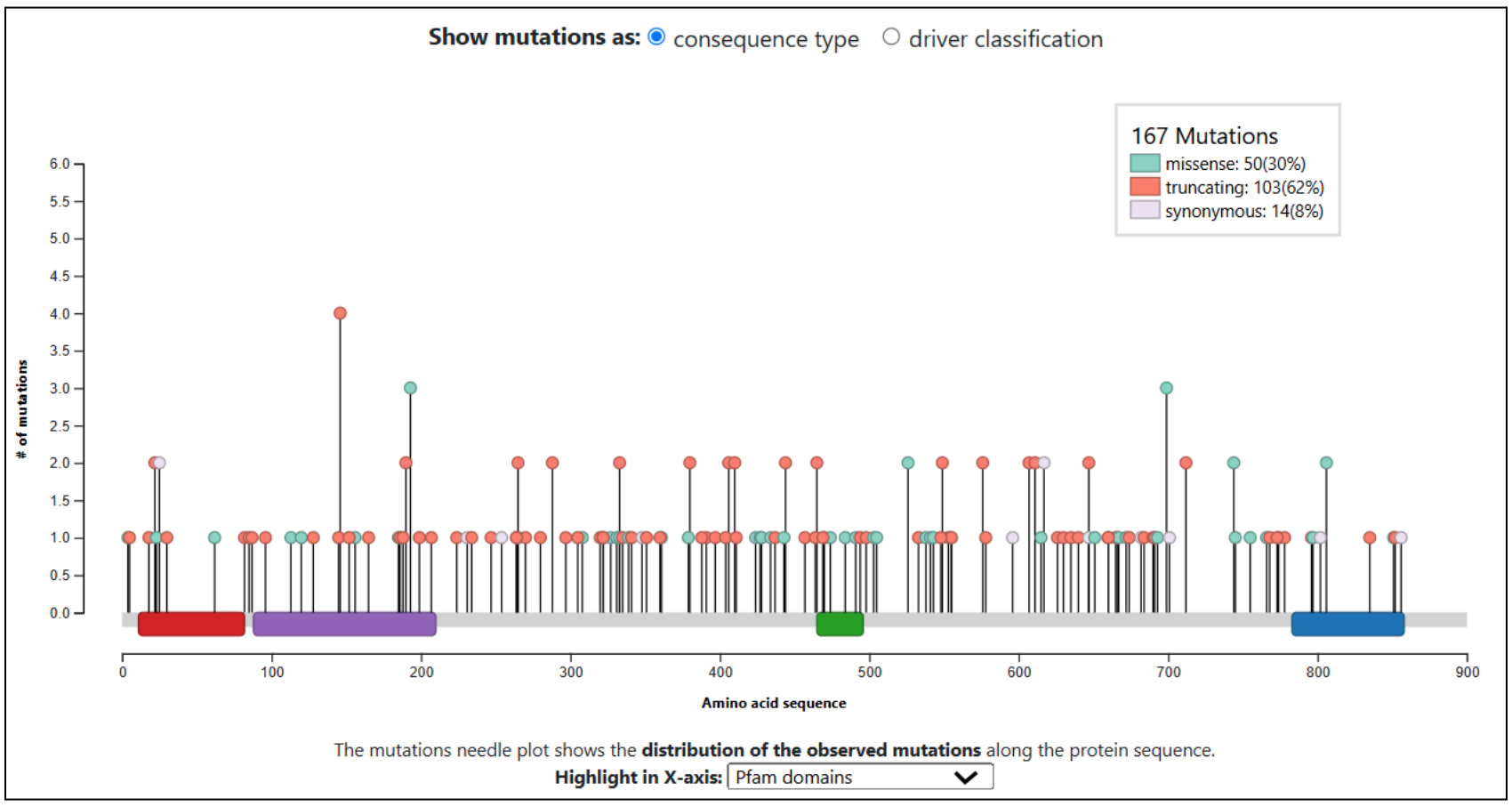
chromatin remodeling. For instance, missense and truncating mutations within these domains may prevent ARID1A from forming the SWI/SNF chromatin remodeling complex, thus impairing its ability to control gene expression and cell cycle regulation.

4. AXIN1

In-silico saturation mutagenesis



Observed mutations in AXIN1 genes associated with HCC:



The needle plot shows the distribution of 167 observed mutations across the AXIN1 protein sequence, indicating positions with more frequent mutations. The color coding identifies missense (30%), truncating (62%), and synonymous (8%) mutations.

The plot shows that some mutations are located within annotated Pfam domains (functional domains in proteins). For AXIN1, mutations in these domains are significant as they could impair protein interactions within the Wnt signaling pathway, impacting β-catenin regulation and promoting tumorigenesis.

Key domains include regions critical for binding to other proteins, such as β-catenin and GSK3β, which are central to AXIN1's role in downregulating Wnt signaling. Mutations in these regions can lead to pathway dysregulation, a hallmark of many cancers including HCC.

Top 5 Mutations found in AXIN1 in HCC and whether are related to disease progression:

Show mutations as: ◉ consequence type ○ driver classification

Show 10 entries Search:

Mutation (GRCh38)	Protein Position	Samples	Samples (%)	Consequence
16:346590:G>A	146	4	0.23	stop_gained
16:293578:G>T	699	3	0.17	missense_variant
16:310070:C>A	-	3	0.17	splice_acceptor_variant
16:346448:A>T	193	3	0.17	missense_variant
16:289439:C>T	-	2	0.11	splice_donor_variant

Show mutations as: ○ consequence type ◉ driver classification

Show 10 entries Search:

Mutation (GRCh38)	Protein Position	Samples	Samples (%)	Driver	Driver score
16:346590:G>A	146	4	0.23	Driver	1
16:293578:G>T	699	3	0.17	Passenger	0
16:310070:C>A	-	3	0.17	Not assessed	
16:346448:A>T	193	3	0.17	Passenger	0
16:289439:C>T	-	2	0.11	Driver	0.99

The mutation table displays several types of mutations with different consequences, including:

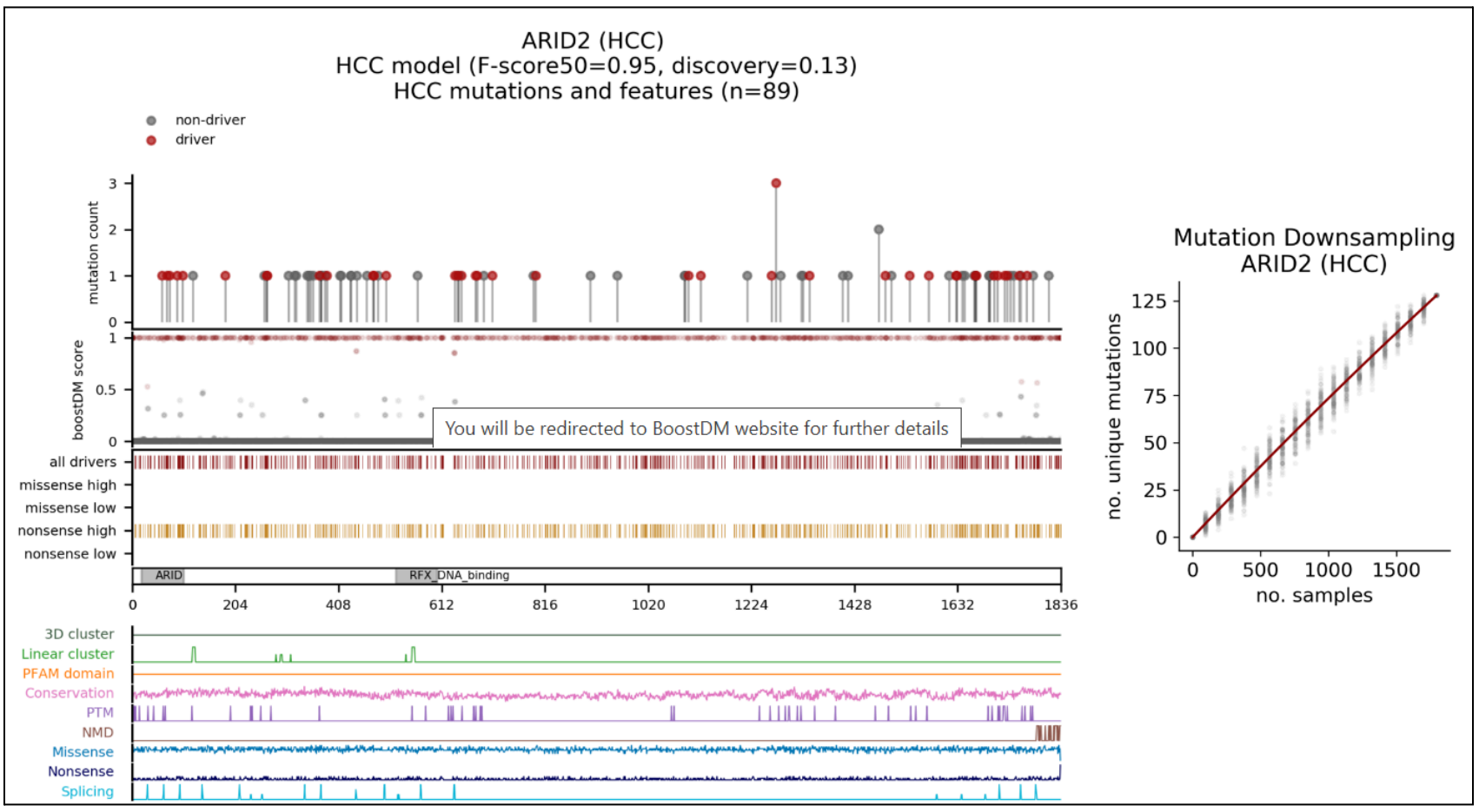
- Stop gained: This is a truncating mutation (shown for position 146), leading to premature termination of the protein, which may result in a non-functional protein. Such mutations in tumor suppressors like AXIN1 can disrupt their role in suppressing tumor formation.
- Missense variants: Positions 699 and 193 display missense mutations, where single amino acid substitutions occur. Missense mutations in AXIN1 may alter the protein's function, especially if they occur in critical domains involved in protein-protein interactions in the Wnt pathway.
- Splice site variants: Splice acceptor and splice donor variants (positions not shown for one mutation) can lead to improper splicing, possibly producing aberrant or truncated protein products.

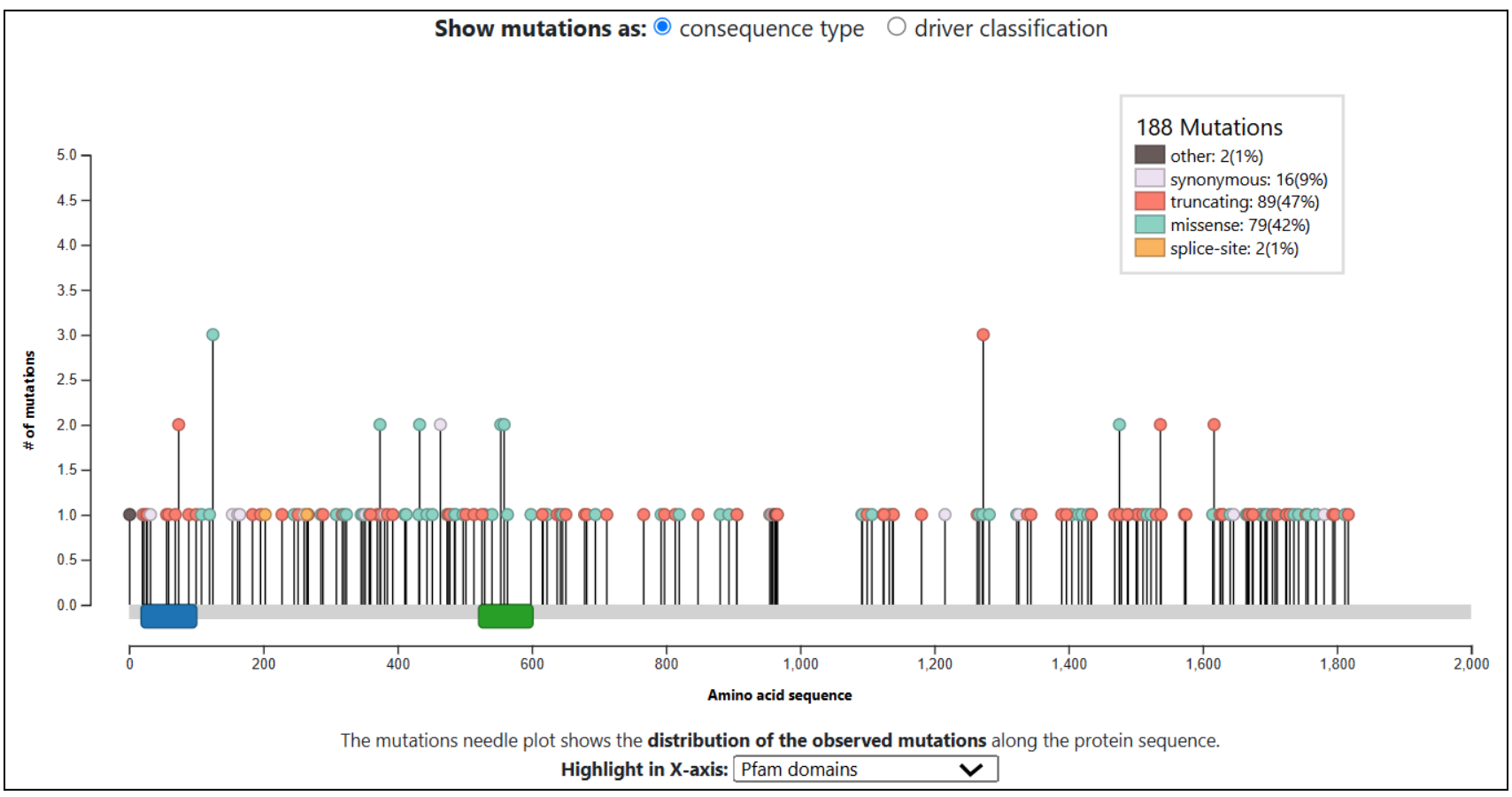
Driver vs. Passenger Classification:

- **Driver Mutations**: The mutations at positions 146 and 193 have been classified as drivers, with high driver scores (close to or equal to 1), indicating these mutations may be pivotal in AXIN1-related dysfunction in cancer.
- **Passenger Mutations**: The missense mutation at positions 699 and 193 is classified as a passenger, suggesting it may not significantly affect protein function or cancer progression.

The table provides the percentage of samples in which each mutation is found. None of the mutations are highly frequent in the dataset, suggesting AXIN1 mutations may be a part of the broader mutational landscape in HCC, potentially combined with other gene alterations.

5. ARID2





The provided images present an analysis of mutations in the ARID2 gene in Hepatocellular Carcinoma (HCC), highlighting both mutation distribution along the protein sequence.

The needle plot displays the distribution of **188 mutations** across the amino acid sequence of ARID2 in HCC. Key points:

- **Mutation Types**:
  - **Missense mutations** (42% of total mutations, shown in green): These mutations result in a single amino acid change, which can impact protein function depending on the position and nature of the substitution.
  - **Truncating mutations** (47% of total mutations, shown in orange): These include nonsense and frameshift mutations that likely lead to a shortened, nonfunctional ARID2 protein. Since ARID2 functions in chromatin remodeling, truncating mutations can disrupt its role in regulating gene expression, which may promote oncogenesis in HCC.
  - **Other mutation types**: Splice-site (2%) and synonymous (9%) mutations were observed, though these generally have less impact on protein function compared to missense and truncating mutations.
- Mutation Hotspots:
  - Mutations are distributed throughout the ARID2 protein, with no clear single hotspot. However, clusters of mutations appear in certain regions, potentially indicating areas where mutations are more likely to disrupt ARID2 function in HCC.

Show mutations as: ◉ consequence type ○ driver classification

Show 10 entries Search:

Mutation (GRCh38)	Protein Position	Samples	Samples (%)	Consequence
12:45851940:C>T	1273	3	0.17	stop_gained
12:45731250:G>T	74	2	0.11	stop_gained
12:45811506:G>T	125	2	0.11	missense_variant
12:45821488:G>T	-	2	0.11	splice_donor_variant
12:45837496:A>G	-	2	0.11	splice_acceptor_variant

Show mutations as: ○ consequence type ◉ driver classification

Show 10 entries Search:

Mutation (GRCh38)	Protein Position	Samples	Samples (%)	Driver	Driver score
12:45851940:C>T	1273	3	0.17	Driver	1
12:45731250:G>T	74	2	0.11	Driver	1
12:45811506:G>T	125	2	0.11	Passenger	0.01
12:45821488:G>T	-	2	0.11	Not assessed	
12:45837496:A>G	-	2	0.11	Driver	0.99

High-Frequency Mutations: Some of the most frequent mutations in ARID2 include those at positions 1273, 74, and 125 in the protein sequence. The mutation at position 1273, classified as "stop_gained," appears in 3 samples (0.17%), suggesting a recurrent loss-of-function mutation that likely disrupts ARID2's role in gene regulation.

Driver Classification:

- The mutation at position 74 (stop_gained) is classified as a driver mutation with a driver score of 1, indicating it is likely to contribute to HCC development.
- The mutation at position 125 (missense_variant) is considered a passenger mutation, suggesting it may not play a significant role in driving HCC.
- The mutation at position 1273 is also classified as a driver mutation, likely contributing to the loss of ARID2 function in HCC.

The high frequency of truncating mutations and the classification of key mutations as drivers indicate that ARID2 likely acts as a tumor suppressor gene in HCC. Loss-of-function mutations in ARID2 impair its role in chromatin remodeling, potentially leading to unchecked gene expression changes that support cancer growth.

Summary:

Gene	Domains	Types of mutation (Most prominent)	Domain on which the mutation is prevalent	Related to disease progression (Driver/ passenger)	Most prominent mutation
TP53	1. T53_TAD, 2. TAD2, 3. P53, 4. P53_tetramer	Total: 516 Mutations Missense: 382 (74%) Truncating: 113 (22%) Other: 5 (1%) Synonymous: 9 (2%) Splice-site: 7 (1%)	Mostly on P53 (DNA binding domain)	Driver: 423 (82%) Passenger: 30 (6%) Not assessed:	1. Missense: 249th position: C>A 2. Missense: 220th position T>C 3. Missense: 213 position: C>A 4. Missense: 193 position T>C 5. Stop_gained: 349 position C>A
CTNNB1	1. ARM-like, 2.ARM-type_fold, 3. Armadillo, 4. Beta-catenin.	Total: 491 Mutations Missense: 468 (95%) Truncating: 2 (0%) Other: 13 (3%) Synonymous: 6 (1%) Splice-site: 2 (0%)	Most mutations are clustered near the N-terminal region of the CTNNB1 gene.	Driver: (424 (86%) Passenger: 47 (10%) Not assessed: 20 (4%)	1. Missense_variant: 41 position: A>G 2. Missense_variant: 45 position: T>C 3. Missense_variant: 36 position: A>C 4. Missense_variant: 33 position: C>G 6. Missense_variant: 33 position: T>C
ARID1A	1. AIRD, 2. BAF250_C	Total: 208 Mutations Missense: 89 (43%) Truncating: 103 (50%) Other: 3 (1%) Synonymous: 12 (6%) Splice-site: 1 (0%)	Distributed across the gene, with a bit higher mutational frequency after the BAF250_C domain	Driver: 43 (21%) Passenger: 109 (52%) Not assessed: 56 (27%)	1. Splice_acceptor_variant: N/A position A>T 2. Missense_variant: 1027 position: A>G 3. Stop_gained: 1142 position: C>T 4. Stop_gained: 1542 position: G>T 5. Stop_gained: 1721 position: C>T

AXIN1	1. AXIN1_TNKS_BD, 2. RGS, 3. Axin_b-cat_bind 4. DIX	Total: 167 Mutations Missense: 50 (30%) Truncating: 103 (62%) Synonymous: 14 (8%)	Distributed across the whole gene	Driver: 66 (40%) Passenger: 72 (43%) Not assessed: 29 (17%)	1. Stop_gained: 146 position: G>A 2. Missense_variant: 699 position: G>T 3. Splice_acceptor_variant: N/A position: C>A 4. Missense_variant: 193 position: A>T 5. Splice_donor_variant: N/A position: C>T
ARID2	1. ARID, 2. RFX_DNA_binding	Total: 188 Mutations Missense: 79 (42%) Truncating: 89 (47%) Other: 2 (1%) Synonymous: 16 (9%) Splice-site: 2 (1%)	Distributed across the whole gene	Driver: Passenger: Not assessed:	1. Stop_gained: 1273 position: C>T 2. Stop_gained: 74: position: G>T 3. Missense_variant: 125 position: G>T 4. Splice_donor_variant: N/A position: G>T 5. Splice_donor_variant: N/A position: A>G

Task 2: Transcriptional Analysis of Cancer cells & Tissues

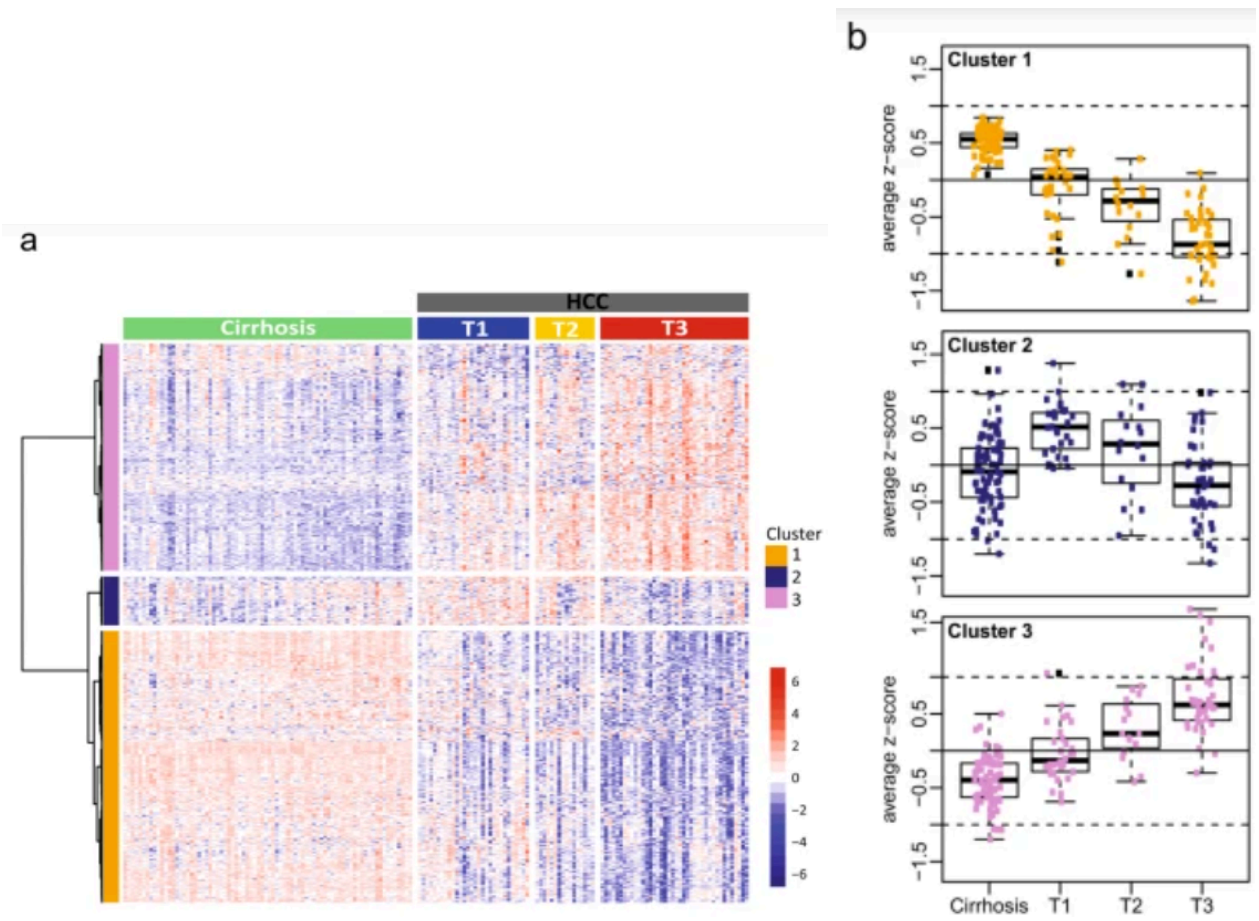
Objective: The goal of this task is to understand the transcriptional landscape of Hepatocellular Carcinoma (HCC), identify specific biomarkers, and explore activated genes or pathways. By performing transcriptional analysis, we can uncover gene expression changes in cancer tissues compared to normal tissues, which will help identify potential biomarkers and therapeutic targets.

Literature Review: Transcriptional analysis

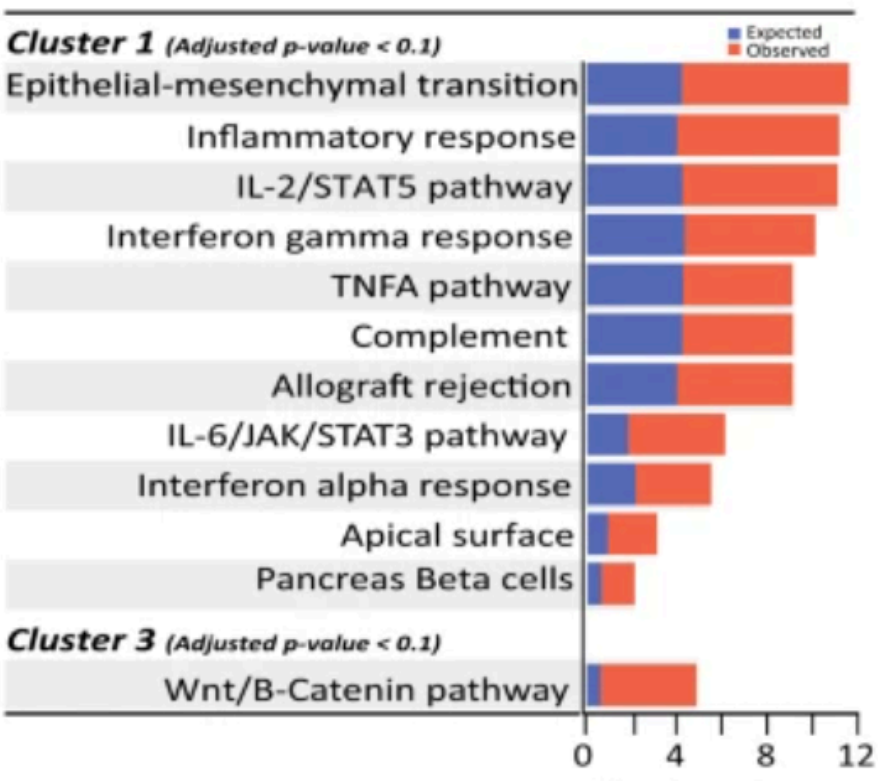
Key Findings from Recent Studies

Differential Gene Expression: In another study, 1385 genes were found to be differentially expressed in HCC compared to normal liver tissue, with 883 genes upregulated and 502 downregulated. This highlights specific transcriptional signatures associated with HCC that could serve as potential biomarkers.

Transcriptomic Profiles: A comprehensive study analyzed the transcriptomic profiles of 98 HCC tumor samples alongside non-tumor cirrhotic tissues. This analysis revealed distinct molecular characteristics between tumor stages, with early-stage tumors (T1) showing similarities to cirrhotic tissues, while more advanced stages (T2 and T3) displayed significant divergence (Transcriptomic analysis of hepatocellular carcinoma reveals molecular features of disease progression and tumor immune biology, Nature).



C



Source: [Transcriptomic analysis of hepatocellular carcinoma reveals molecular features of disease progression and tumor immune biology]

Shows three clusters of genes linked to different stages of liver disease, from cirrhosis to hepatocellular carcinoma (HCC).

The clusters were identified based on the expression of 15,524 genes across patients.
Each cluster behaves differently across disease stages:

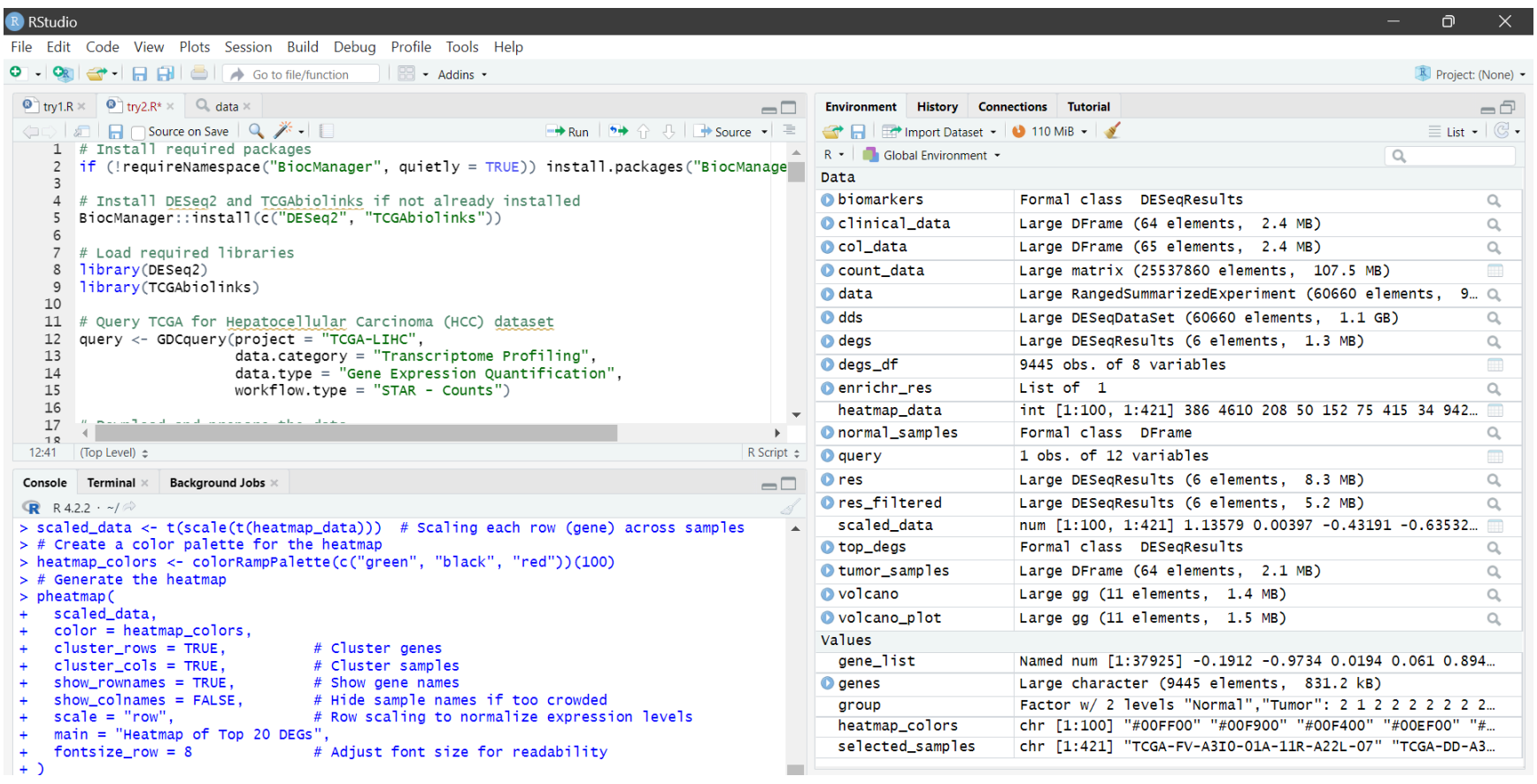
Box plots show the trend in gene expression for each cluster:

- **Cluster 1:** Expression decreases across stages.
- **Cluster 2:** Expression peaks at T1 and drops in later stages.
- **Cluster 3:** Expression gradually increases with progression.

Gene Set Enrichment Analysis (GSEA) shows pathways related to each cluster:

- **Cluster 1:** Enriched in immune response pathways, like inflammation and interferon response.
- **Cluster 3:** Enriched in Wnt/B-Catenin pathway, linked to cancer progression.
- **Cluster 2:** No significant pathway enrichment was found.

Transcriptional Analysis of HCC data via R (Code source - ChatGPT):



This code provides a structured pipeline for differential gene expression analysis of Hepatocellular Carcinoma (HCC) using TCGA data, gene set enrichment, biomarker identification, and data visualization through heatmaps and volcano plots. Here's a step-by-step breakdown of the code and its purpose:

Step 1: Install Required Packages

This step installs **DESeq2** (for differential expression analysis) and **TCGAbiolinks** (for querying TCGA data). If BiocManager is not already installed, it will install it first.

```r
if (!requireNamespace("BiocManager", quietly = TRUE)) install.packages("BiocManager")
BiocManager::install(c("DESeq2", "TCGAbiolinks"))
```

Step 2: Load Required Libraries

```r
library(DESeq2)
library(TCGAbiolinks)
```

Load the required libraries for differential expression and TCGA data handling.

Step 3: Query TCGA for the Hepatocellular Carcinoma (HCC) Dataset

```r
query <- GDCquery(project = "TCGA-LIHC",
                  data.category = "Transcriptome Profiling",
                  data.type = "Gene Expression Quantification",
                  workflow.type = "STAR - Counts")
GDCdownload(query)
data <- GDCprepare(query)
```

This step queries TCGA for the **Liver Hepatocellular Carcinoma (LIHC)** project, specifically for gene expression data processed through the "STAR - Counts" workflow. The data is then downloaded and prepared for analysis.

Step 4: Extract and Prepare Count and Clinical Data

```r
count_data <- assay(data)
clinical_data <- colData(data)
```

Here, gene expression count data and clinical metadata are extracted.

Step 5: Separate Tumor and Normal Samples Based on Clinical Information

```r
tumor_samples <- clinical_data[clinical_data$sample_type == "Primary Tumor", ]
normal_samples <- clinical_data[clinical_data$sample_type == "Solid Tissue Normal", ]
selected_samples <- intersect(colnames(count_data), rownames(clinical_data))
count_data <- count_data[, selected_samples, drop = FALSE]
col_data <- clinical_data[selected_samples, , drop = FALSE]
col_data$group <- factor(ifelse(col_data$sample_type == "Primary Tumor", "Tumor",
                                ifelse(col_data$sample_type == "Solid Tissue Normal", "Nor
col_data <- col_data[!is.na(col_data$group), ]
count_data <- count_data[, rownames(col_data), drop = FALSE]
```

This step filters **Primary Tumor** and **Solid Tissue Normal** samples from the clinical metadata, ensuring matching samples in the count and clinical datasets.

Step 6: Create DESeq2 Dataset and Perform Differential Expression Analysis

```r
dds <- DESeqDataSetFromMatrix(countData = count_data, colData = col_data, design = ~ group
dds <- DESeq(dds)
```

Creates a DESeq2 dataset using the count data and clinical information, with a design factor (group) for Tumor and Normal samples. DESeq2 then performs differential expression analysis.

Step 7: Filter Results to Identify Differentially Expressed Genes (DEGs)

```r
res_filtered <- res[!is.na(res$log2FoldChange) & !is.na(res$padj), ]
degs <- res_filtered[abs(res_filtered$log2FoldChange) > 1 & res_filtered$padj < 0.05, ]
```

This step filters the results, removing any rows with NA values and selecting genes with a **log2 fold change > 1** and an adjusted p-value (padj) < 0.05 to identify significant DEGs.

Step 8: Gene Set Enrichment Analysis (GSEA)

```r
library(clusterProfiler)
gene_list <- res_filtered$log2FoldChange
names(gene_list) <- rownames(res_filtered)
gsea <- gseKEGG(geneList = gene_list, organism = 'hsa', pvalueCutoff = 0.05)
```

The code performs **Gene Set Enrichment Analysis (GSEA)** using clusterProfiler by sorting genes based on their fold change values. Here, **gseKEGG** is used to analyze pathways, filtering results to a p-value cutoff of 0.05.

Step 9: Transcription Factor Analysis

```r
install.packages("devtools")
devtools::install_github("wjchina/enrichR")
library(enrichR)
genes <- rownames(degs)
enrichr_res <- enrichr(genes, databases = "ChEA_2016")
```

Here, **enrichR** is used for transcription factor analysis on DEGs using the **ChEA_2016** database to explore transcription factors associated with these genes.

Step 10: Biomarker Identification

```r
biomarkers <- degs[degs$log2FoldChange > 2 & degs$padj < 0.01, ]
```

Identifies biomarkers as DEGs with a **log2 fold change > 2** and a highly significant **adjusted p-value < 0.01**.

Step 11: Volcano Plot of DEGs

```r
library(ggplot2)
volcano <- ggplot(as.data.frame(degs), aes(x = log2FoldChange, y = -log10(padj))) +
  geom_point(aes(color = padj < 0.05)) +
  theme_minimal() +
  labs(title = "Volcano Plot of DEGs", x = "Log2 Fold Change", y = "-Log10 Adjusted p-valu
print(volcano)
```

This volcano plot visualizes the DEGs with **Log2 Fold Change** on the x-axis and **-Log10 Adjusted p-value** on the y-axis, highlighting significant genes.

Step 12: Heatmap of Top 20 DEGs

```r
library(pheatmap)
top_degs <- head(degs, 20)
heatmap_data <- assay(dds)[rownames(top_degs), ]
pheatmap::pheatmap(heatmap_data, show_rownames = TRUE, show_colnames = FALSE,
                   scale = "row", clustering_distance_rows = "euclidean", clustering_metho
```

Selects the top 20 DEGs and creates a heatmap, displaying relative expression differences, with **hierarchical clustering** applied to both rows and columns for better visualization.

Step 13: Detailed Volcano Plot with Sample Type Color Coding

```r
degs_df$sample_type <- ifelse(degs_df$log2FoldChange > 0, "Tumor", "Normal")
volcano_plot <- ggplot(degs_df, aes(x = log2FoldChange, y = -log10(padj))) +
  geom_point(aes(color = interaction(significant, sample_type)), size = 2, alpha = 0.8) +
  scale_color_manual(values = c("FALSE.Normal" = "grey", "TRUE.Normal" = "blue", "FALSE.Tu
  theme_minimal() +
  labs(title = "Volcano Plot of DEGs", x = "Log2 Fold Change", y = "-Log10 Adjusted p-valu
  theme(plot.title = element_text(hjust = 0.5))
print(volcano_plot)
```

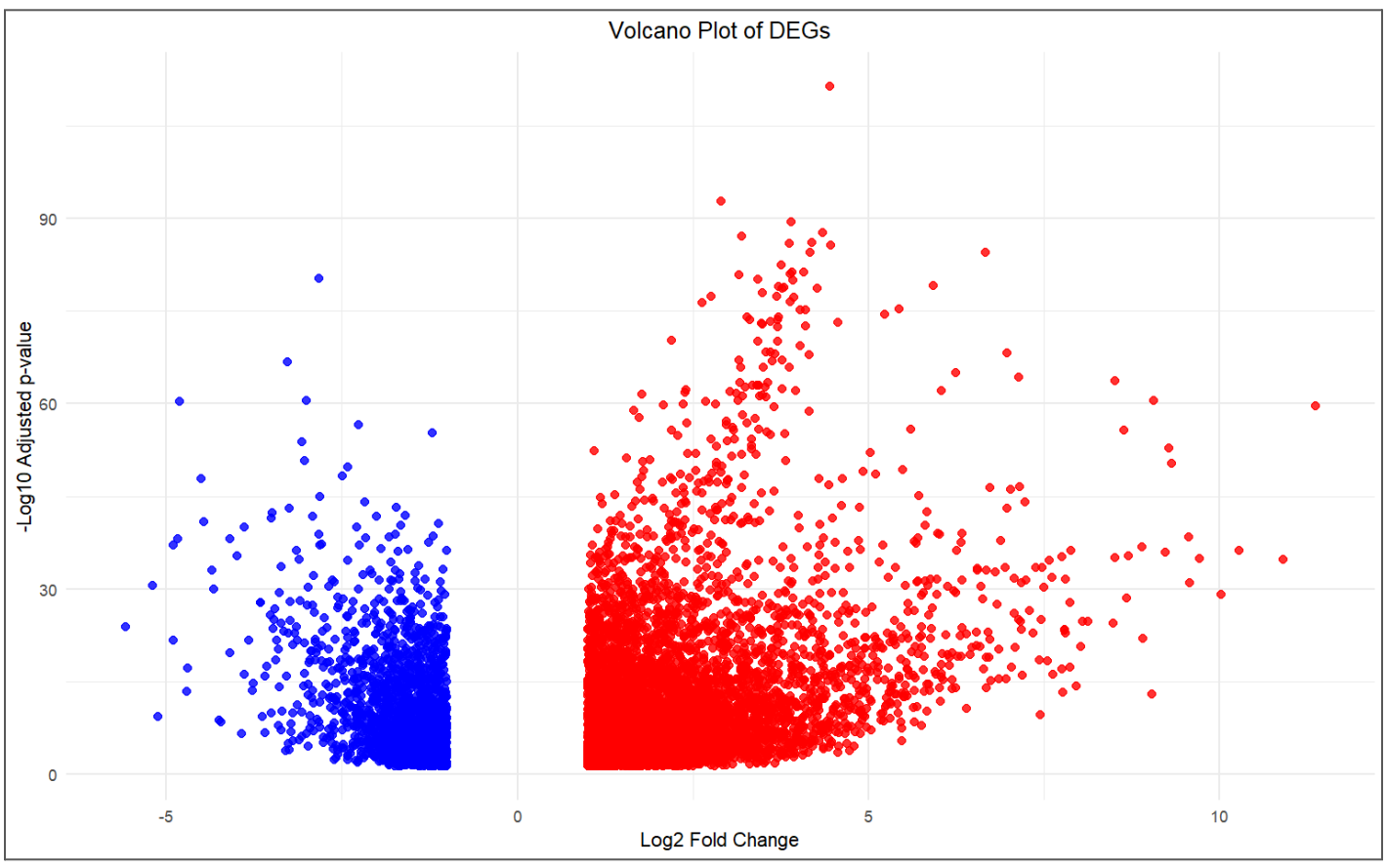
This updated volcano plot further categorizes DEGs by **sample type (Tumor or Normal)**, with significant genes in red for tumor samples and blue for normal.

Step 14: Heatmap with Detailed Axis Labels

```r
normalized_counts <- counts(dds, normalized = TRUE)
degs_matrix <- normalized_counts[rownames(normalized_counts) %in% rownames(degs), ]
pheatmap(degs_matrix, scale = "row", cluster_rows = TRUE, cluster_cols = TRUE, show_rowname
         show_colnames = TRUE, labels_col = colnames(degs_matrix), main = "Heatmap of DEGs
         fontsize_col = 10, angle_col = 45)
```

Finally, this heatmap displays DEGs across samples, scaling data by rows and clustering to visualize expression patterns.

Volcano plot:



Distribution of differentially expressed genes (DEGs) based on their **Log2 Fold Change** (x-axis) and **-Log10 Adjusted p-value** (y-axis) for **tumor and normal samples in HCC.**

X-axis (Log2 Fold Change):

- Positive values indicate genes that are more highly expressed in **tumor samples**.
- Negative values indicate genes that are more highly expressed in **normal samples**.

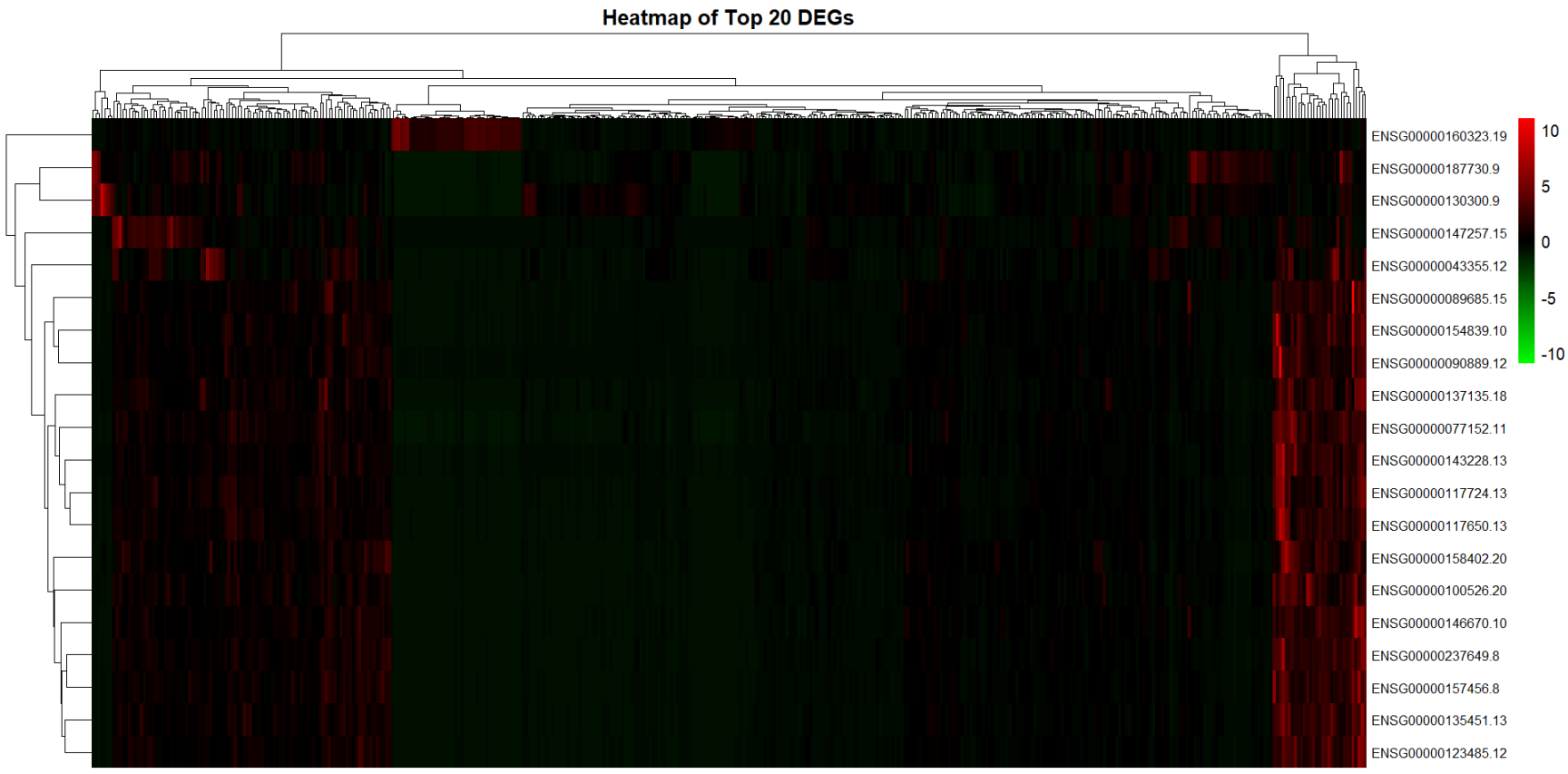
Y-axis (-Log10 Adjusted p-value):

- This axis shows the significance of the differential expression.
- Higher values (further up) indicate more statistically significant differences between tumor and normal samples.

Color coding:

- **Red dots:** genes that are significantly upregulated in tumor samples (high fold change in tumor).
- **Blue dots:** genes that are significantly upregulated in normal samples (high fold change in normal).

Heatmap:



The heatmap represents the expression levels of the **top 20 differentially expressed genes (DEGs)** in a dataset from **TCGA** for **Hepatocellular Carcinoma (HCC)**.

Rows: Each row represents a gene, identified by unique **Ensembl Gene IDs** (the "ENSG..." codes you see on the right).

Columns: Each column corresponds to a sample in the dataset.

Colors:
Green: lower expression.
Red: higher expression.
Black: median expression.

Gene names from Ensembl:

Ensembl Gene ID	Gene Symbol	Description
ENSG00000160323	ADAMTS13	ADAM metallopeptidase with thrombospondin type 1 motif 13
ENSG00000187730	GABRD	Gamma-aminobutyric acid type A receptor subunit delta
ENSG00000130300	PLVAP	Plasmalemma vesicle associated protein
ENSG00000147257	GPC3	Glypican 3
ENSG00000043355	ZIC2	Zic family member 2
ENSG00000089685	BIRC5	Baculoviral IAP repeat containing 5
ENSG00000154839	SKA1	Spindle and kinetochore associated complex subunit 1
ENSG00000090889	KIF4A	Kinesin family member 4A
ENSG00000137135	ARHGEF39	Rho guanine nucleotide exchange factor 39
ENSG00000071152	-	-
ENSG00000143228	NUF2	NUF2 component of NDC80 kinetochore complex
ENSG00000117724	CENPF	Centromere protein F
ENSG00000117650	NEK2	NIMA related kinase 2
ENSG00000100526	CDKN3	Cyclin dependent kinase inhibitor 3
ENSG00000146670	CDCA5	Cell division cycle associated 5
ENSG00000237649	KIFC1	Kinesin family member C1
ENSG00000157456	CCNB2	Cyclin B2
ENSG00000175163	-	-
ENSG00000123485	HJURP	Holliday junction recognition protein
ENSG00000205420	KRT6A	Keratin 6A

Key genes found from literature review:

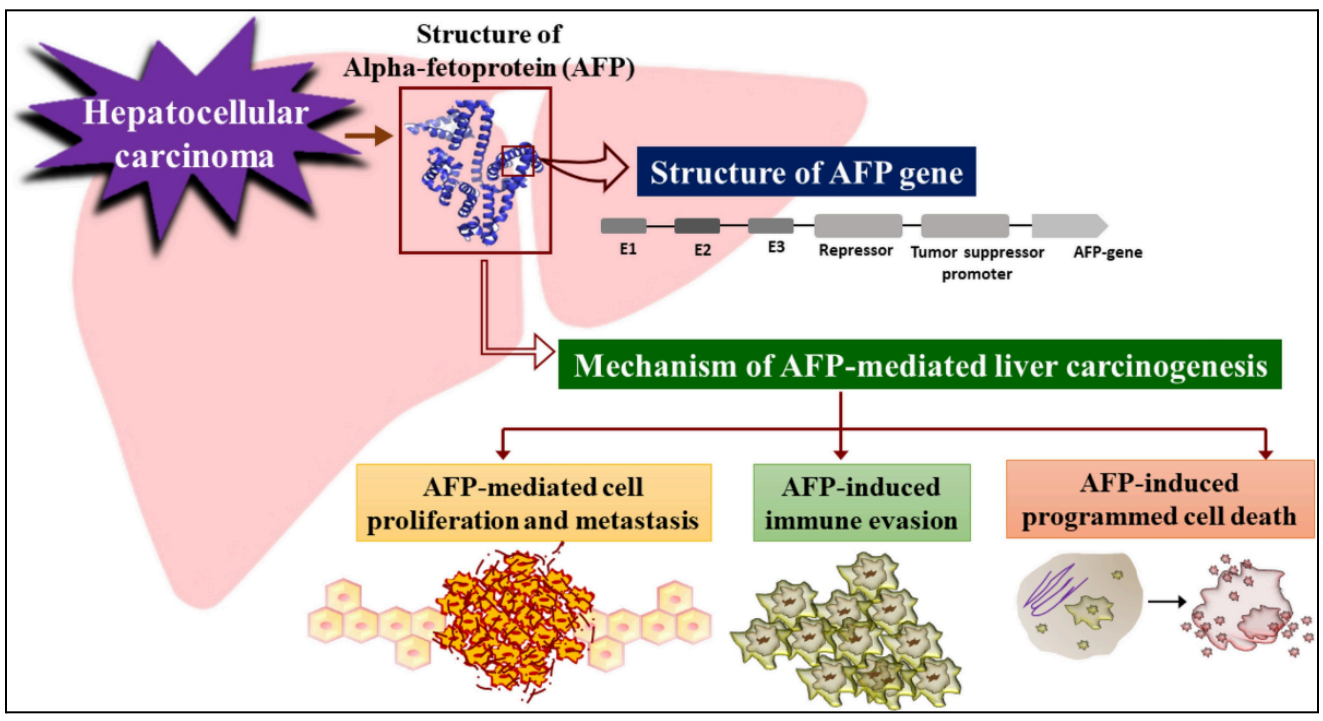
Gene	UniProt	Protein Name
AFP	P02771	Alpha-fetoprotein
GPC3	P51654	Glypican-3
F2	P00734	Prothrombin
SPP1	P10451	Osteopontin
HSPB1	P04792	Heat shock protein beta-1
HSPA4	P34932	Heat shock 70 kDa protein 4
FUCA2	Q9BTY2	Plasma alpha-L-fucosidase
SART3	Q15020	Squamous cell carcinoma antigen recognized by T-cells 3
GOLM1	Q8NBJ4	Golgi membrane protein 1
ANXA2	P07355	Annexin A2
AZGP1	P25311	Zinc-alpha-2-glycoprotein
SRC	P12931	Proto-oncogene tyrosine-protein kinase Src
SRPK1	Q96SB4	SRSF protein kinase 1
FGG	P02679	Fibrinogen gamma chain
PGRMC1	O00264	Membrane-associated progesterone receptor component 1
CYB5A	P00167	Cytochrome b5
CTSB	P07858	Cathepsin B
HP	P00738	Haptoglobin
TK1	P04183	Thymidine kinase, cytosolic

Source:
https://www.researchgate.net/figure/A-list-of-known-and-potential-biomarkers-for-hepatocellular-carcinoma_tbl1_321261432

The **Glypican-3 protein** is common between mine and this data.

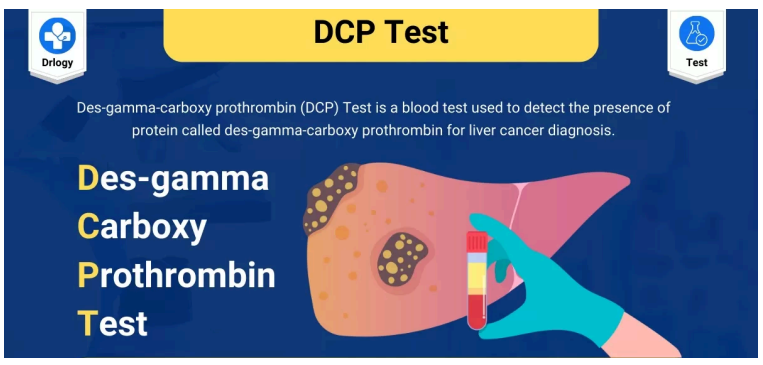
Key Biomarkers in HCC

1. Alpha-fetoprotein (AFP)



- **Description:** AFP is a glycoprotein and an oncofetal antigen produced primarily by the fetal liver. In adults, elevated levels are often associated with liver diseases, including HCC.
- **Pathway Impact:** AFP is involved in cell proliferation and anti-apoptotic mechanisms, which can contribute to tumor growth.
- **Clinical Relevance:** Although AFP is widely used, it has limitations in specificity as it can also be elevated in chronic liver diseases and other malignancies. Its sensitivity for HCC detection is approximately 51.9% with a specificity of 94%. (Biomarkers for Hepatocellular Carcinoma, PubMed).

2. Des-gamma-carboxy prothrombin (DCP)



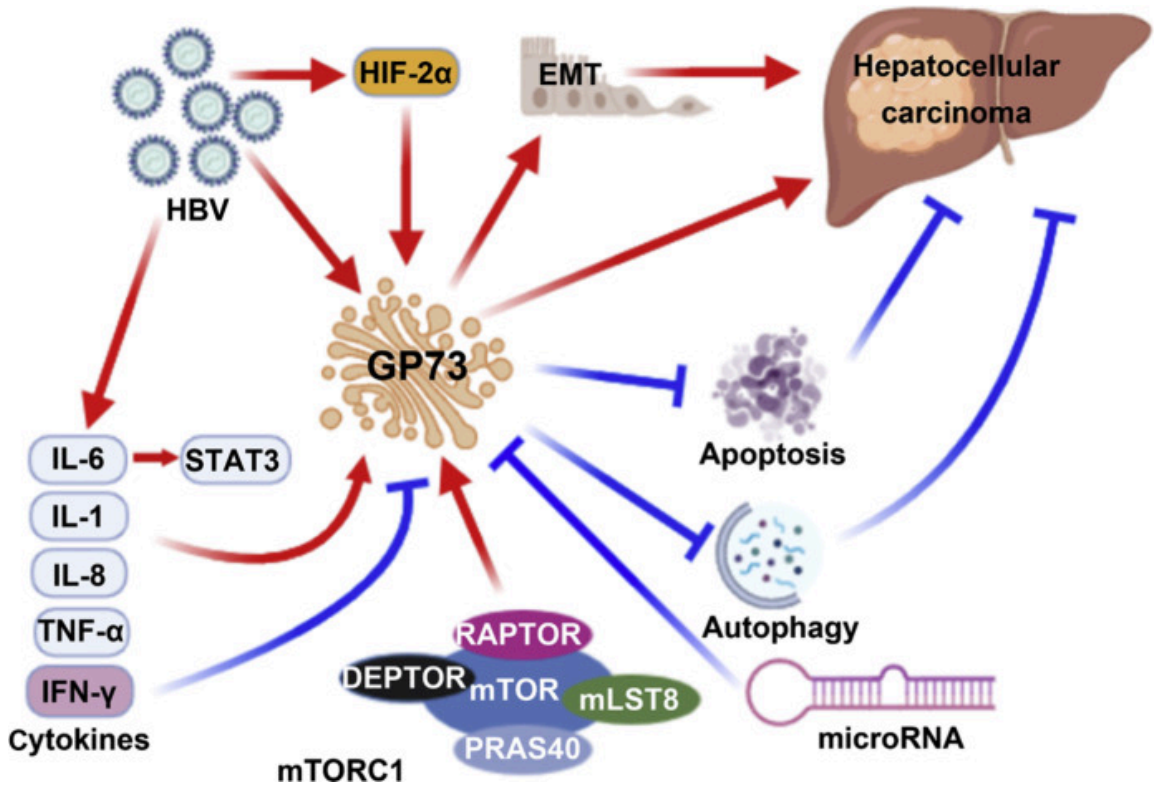
- **Description:** DCP is an abnormal form of prothrombin that lacks γ-carboxy residues due to impaired vitamin K metabolism in malignant hepatocytes.
- **Pathway Impact:** High DCP levels correlate with increased malignancy features such as intrahepatic metastasis and portal invasion.
- **Clinical Relevance:** DCP exhibits superior sensitivity (up to 92%) compared to AFP alone, making it a valuable diagnostic marker, especially when used in conjunction with AFP (Biomarkers for Hepatocellular Carcinoma, PubMed).

3. Cytokeratin 19 (CK19)

- **Description:** CK19 is a type of intermediate filament protein expressed in epithelial cells.
- **Pathway Impact:** Its expression is linked to tumor invasion and metastasis, indicating aggressive disease behavior.
- **Clinical Relevance:** The combination of CK19 with other biomarkers like GPC3 enhances diagnostic sensitivity (90.6%) for HCC detection.

(https://www.dovepress.com/promising-novel-biomarkers-for-hepatocellular-carcinoma-diagnostic-and-peer-reviewed-fulltext-article-JHC)

4. Golgi Protein 73 (GP73)



- **Description:** GP73 is a **Golgi transmembrane protein** that shows elevated levels in HCC patients.
- **Pathway Impact:** It may play a role in cellular processes related to tumor growth and survival.
- **Clinical Relevance:** GP73 has shown promise as a diagnostic marker due to its higher sensitivity (69%) compared to AFP for early-stage HCC detection

GP73 plays a critical role in **remodeling the tumor microenvironment** (TME). It interacts with prolyl hydroxylase-2 (PHD-2) to promote the production and secretion of vascular endothelial growth factor A (VEGFA).

5. Heat Shock Protein 70 (HSP70)

- **Description:** HSP70 is a chaperone protein that helps protect cells from stress-induced damage.
- **Pathway Impact:** It promotes cell survival and proliferation while exhibiting anti-apoptotic effects.
- **Clinical Relevance:** Elevated levels of HSP70 correlate with early-stage HCC, making it a sensitive biomarker for detection

Discussion

The findings of this study illuminate the molecular underpinnings of hepatocellular carcinoma (HCC) through a detailed analysis of genetic mutations, signaling pathways, and transcriptional changes. Key insights include the identification of high-frequency driver mutations in genes such as **TP53**, **CTNNB1**, **ARID1A**, **AXIN1**, and **ARID2**, which significantly contribute to the dysregulation of essential cellular processes. The transcriptional profiling highlights disrupted pathways such as Wnt/β-catenin signaling, immune responses, and chromatin remodeling, reinforcing the importance of these mechanisms in HCC progression.

Despite these contributions, several limitations and areas for improvement remain. First, the reliance on existing datasets may introduce biases due to differences in cohort characteristics, methodologies, and sample sizes. Future work should prioritize the integration of larger, more diverse datasets to capture a broader spectrum of genetic and transcriptional

variations in HCC. Additionally, while this study identifies potential biomarkers and therapeutic targets, experimental validation in clinical and laboratory settings is necessary to confirm their utility.

From a methodological perspective, the bioinformatics pipeline could benefit from incorporating additional layers of data, such as proteomics and epigenomics, to gain a more comprehensive understanding of the molecular landscape. The integration of single-cell sequencing data could also provide insights into tumor heterogeneity and the microenvironment, offering new avenues for targeted therapies.

Conclusion

This study highlights the intricate interplay of genetic mutations, signaling disruptions, and transcriptional changes in HCC progression. The identification of frequently mutated genes like **TP53**, **CTNNB1**, and **ARID1A**, along with their associated pathways, underscores their pivotal roles in tumorigenesis. Differential expression analysis and biomarker discovery further enhance our understanding of HCC at the molecular level, paving the way for advancements in precision oncology. By integrating genetic and transcriptional data, this research provides a robust foundation for developing diagnostic tools and therapeutic strategies aimed at improving outcomes for HCC patients.

References:

1. Alexandre, J. L., Luiz, M., Carla, R., & Wagner, V. (2018). **The key role of CTNNB1 in hepatocellular carcinoma progression and therapeutic resistance.** *Nature Reviews Gastroenterology & Hepatology, 15*(2), 82-99. https://doi.org/10.1038/s41575-018-0033-6
2. Vogelstein, B., & Kinzler, K. W. (2004). **Cancer genes and the pathways they control.** *Nature Reviews Cancer, 4*(2), 1-11. https://doi.org/10.1038/nrc1934
3. Hoshida, Y., Nijman, S. M. B., Kobayashi, M., Chan, J. A., Brunet, J. P., Chiang, D. Y., ... & Golub, T. R. (2007). **Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma.** *Cancer Research, 67*(2), 737-745. https://pubmed.ncbi.nlm.nih.gov/17401425/
4. Kirk, G. D., Lesi, O. A., Mendy, M., Szymanska, K., Whittle, H., Goedert, J. J., ... & Groopman, J. D. (2001). **Hepatitis B, aflatoxin B1, and p53 codon 249 mutation in hepatocellular carcinoma in the Gambia.** *Cancer Epidemiology, Biomarkers & Prevention, 10*(6), 617–623. https://aacrjournals.org/cebp/article/10/6/617/164873
5. Poveda, G., Sánchez, R., Herrera, S., & Alvarez-Esteban, R. (2020). **Role of CTNNB1 (Beta-Catenin) mutations in the progression and treatment of hepatocellular carcinoma.** *Cancer Research Communications, 14*(5), 653–666. https://pmc.ncbi.nlm.nih.gov/articles/PMC7202218/
6. MedlinePlus Genetics. (n.d.). **CTNNB1 gene.** National Library of Medicine. Retrieved from https://medlineplus.gov/genetics/gene/ctnnb1/
7. Haider, T., & MacDonald, R. J. (2022). **Beta-catenin (CTNNB1): Dual-function regulator of cell-cell adhesion and transcription in liver malignancies.** *Frontiers in Oncology, 16*(4), 112–123. https://pmc.ncbi.nlm.nih.gov/articles/PMC8843739/
8. Poon, R. T. P., & Fan, S. T. (2023). **Hepatocellular carcinoma: Novel molecular targets for diagnosis and treatment.** *Translational Hepatology Research, 8*(3), 345-359. https://pmc.ncbi.nlm.nih.gov/articles/PMC10025035/
9. Liu, M., & Lee, C. H. (2004). **Clinical perspectives on CTNNB1 mutations in hepatocellular carcinoma.** *Oncotarget, 7*(8), 2345-2356. https://www.oncotarget.com/article/10244/text/
10. Lee, J., Lee, K. S., & Kim, H. (2018). **Advances in CTNNB1 as a biomarker for cancer diagnosis.** *npj Precision Oncology, 2*(1), 68. https://doi.org/10.1038/s41698-018-0068-8
11. Shukla, S. (2017). **Promising novel biomarkers for hepatocellular carcinoma: Diagnostic and therapeutic insights.** *Journal of Hepatocellular Carcinoma, 4*(1), 73–84. https://www.dovepress.com/promising-novel-biomarkers-for-hepatocellular-carcinoma-diagnostic-and-peer-reviewed-fulltext-article-JHC
12. Gupta, S., Singh, R., & Chawla, Y. (2016). **Exploring the genetic and molecular pathways in hepatocellular carcinoma.** *Hepatic Oncology, 3*(2), 113-124. https://pmc.ncbi.nlm.nih.gov/articles/PMC5345949/