

A NON-CODING CANCER MUTATION DISRUPTING AN HNF4A BINDING MOTIF AFFECTS AN ENHANCER REGULATING GENES ASSOCIATED TO THE PROGRESSION OF LIVER CANCER

M. Cavalli^{1,*}, K. Diamanti¹, G. Pan¹, M.J. Dabrowski^{1,2}, J. Komorowski^{2,3,4}, C. Wadelius^{1,*}

¹Uppsala University, 752 37, Uppsala, Sweden

²Institute of Computer Science, Polish Academy of Sciences, 01-248 Warszawa, Poland

³Swedish Collegium for Advanced Study, 752 38, Uppsala, Sweden

⁴Washington National Primate Research Center, Seattle, WA 98121, USA

Background: Somatic mutations in coding regions of the genome may result in non-functional proteins that can lead to cancer or other diseases, however cancer mutations in the non-coding regions have rarely been studied and the interpretation of their effects is difficult. Non-coding mutations might act by breaking or creating transcription factor binding motifs in promoters, enhancers or silencers resulting in altered expression of target gene(s). A high number of mutations have been reported in coding and non-coding regions in cells of liver cancer. Hepatocyte nuclear factor 4 α is a transcription factor that regulates the expression of several genes in liver cells, while the motifs it binds are frequently mutated in promoters and enhancers in liver cancer. **Aim:** The aim of the study is to evaluate the genetic effects of a non-coding somatic mutation frequently observed in liver cancer. **Materials and Methods:** We evaluated experimentally the effects of a somatic mutation frequently reported in liver cancer as a motif-breaker for the binding of hepatocyte nuclear factor 4 α . The effects of the mutation on protein binding and enhancer activity were studied in HepG2 cells via electrophoresis mobility shift assay and dual luciferase reporter assays. We also studied genome-wide promoter-enhancer interactions performing targeted chromosome conformation capture in liver tissue to identify putative target genes whose expression could be altered by the mutation. **Results:** We found that the mutation leads to reduced protein binding and a decrease in enhancer activity. The enhancer harboring the mutation interacts with the promoters of *ANAPC13*, *MAP6D1* and *MUC13*, which have been implicated in liver cancer. **Conclusions:** The study highlights the importance of non-coding somatic mutations, vastly understudied, but likely to contribute to cancer development and progression. **Key Words:** motif-breaking mutations, gene regulation, liver cancer.

DOI: 10.32471/exp-oncology.2312-8852.vol-43-no-1.15925

The vast majority of somatic mutations are located in non-coding segments of cancer genomes. The genomic location of a mutation can provide clues regarding its contribution to cancer. Most experimental research has been focused on the effect of mutations in coding regions, such as missense, nonsense or splice-site mutations, which are, arguably, easier to interpret. However, the consequences of mutations in the non-coding regions remain largely unknown.

The large-scale efforts of the ENCODE project [1] and the Roadmap Epigenomics project [2] have identified active non-coding regulatory elements in multiple cell types and tissues enabling investigations of the role of regulatory mutations in cancer. Moreover, recent contributions from the Pan-Cancer Analysis of Whole Genomes project provided further evidence on the importance of noncoding mutations [3].

The majority of somatic regulatory mutations act by breaking or creating transcription factor (TF) binding motifs in promoters, enhancers or silencers [4].

A somatic mutation affecting the binding of a TF to an enhancer may lead to downstream alteration of the expression of target gene(s) that is often challenging to identify.

Genome-wide promoter-enhancer interactions have been studied with high-throughput chromosome conformation capture (HiC), a chromosome conformation capture technique that allows defining topologically associated domains (TADs) and long-range chromatin interactions [5]. Targeted chromosome conformation capture (Capture-C, CHi-C or HiCap) has been developed to further explore promoter-enhancer interactions to a greater resolution [6–8].

The most common type of liver cancer is hepatocellular carcinoma (HCC) that originates in the hepatocytes, the main cell type of the organ. Major predisposing factors are viral infection by either hepatitis B virus or hepatitis C virus or excessive alcohol intake leading to inflammation that induce liver fibrosis and eventually scarring of the liver tissue (cirrhosis). Prevention includes reducing the risk of cirrhosis by diminishing alcohol intake, reducing obesity and diabetes by diet control, and preventive vaccination against hepatitis B virus or treatment of liver disease. Hepatitis viral infections promote genome-wide accumulation of somatic mutations both in protein-coding and non-coding regions with a similar frequency [9]. HCC is known to have a high tumor mutational burden (median 3.6/Mb) resulting in ~10,000 mutations per tumor [10].

We have previously developed a bioinformatics strategy to identify motif-breaking regulatory mutations in gastrointestinal cancers. Motifs for hepatocyte nuclear factor 4 α (HNF4 α) and other liver-relevant TFs (e.g. FOXA1, FOXA2) were frequently mutated in promoters

Submitted: October 26, 2020.

*Correspondence: E-mail: marco.cavalli@igp.uu.se; claes.wadelius@igp.uu.se

Abbreviations used: DBD – DNA binding domain; HCC – hepatocellular carcinoma; HiC – high-throughput chromosome conformation capture; HNF4 α – hepatocyte nuclear factor 4 α ; TF – transcription factor.

and enhancers in HCC [4]. We detected an increased frequency of motif-breaking somatic mutations in HCC affecting key positions in the HNF4α binding motif. Out of 39 recurrent genome-wide HNF4α motif-breaking mutations, the most common were located at positions 3, 13 and 14 of the motif (Figure, C).

Here we experimentally validated the effect of a motif-breaking somatic mutation in a non-coding regulatory

element and identified its putative target gene(s) by studying long range chromatin interactions using HiCap.

MATERIALS AND METHODS

Cell culture. HepG2 cells were cultured in RPMI 1640 medium supplemented with 10% non-inactivated FBS, L-glutamine and 10,000 units penicillin and 10 μg streptomycin/mL (Sigma-Aldrich) at 37 °C with 5% CO₂.

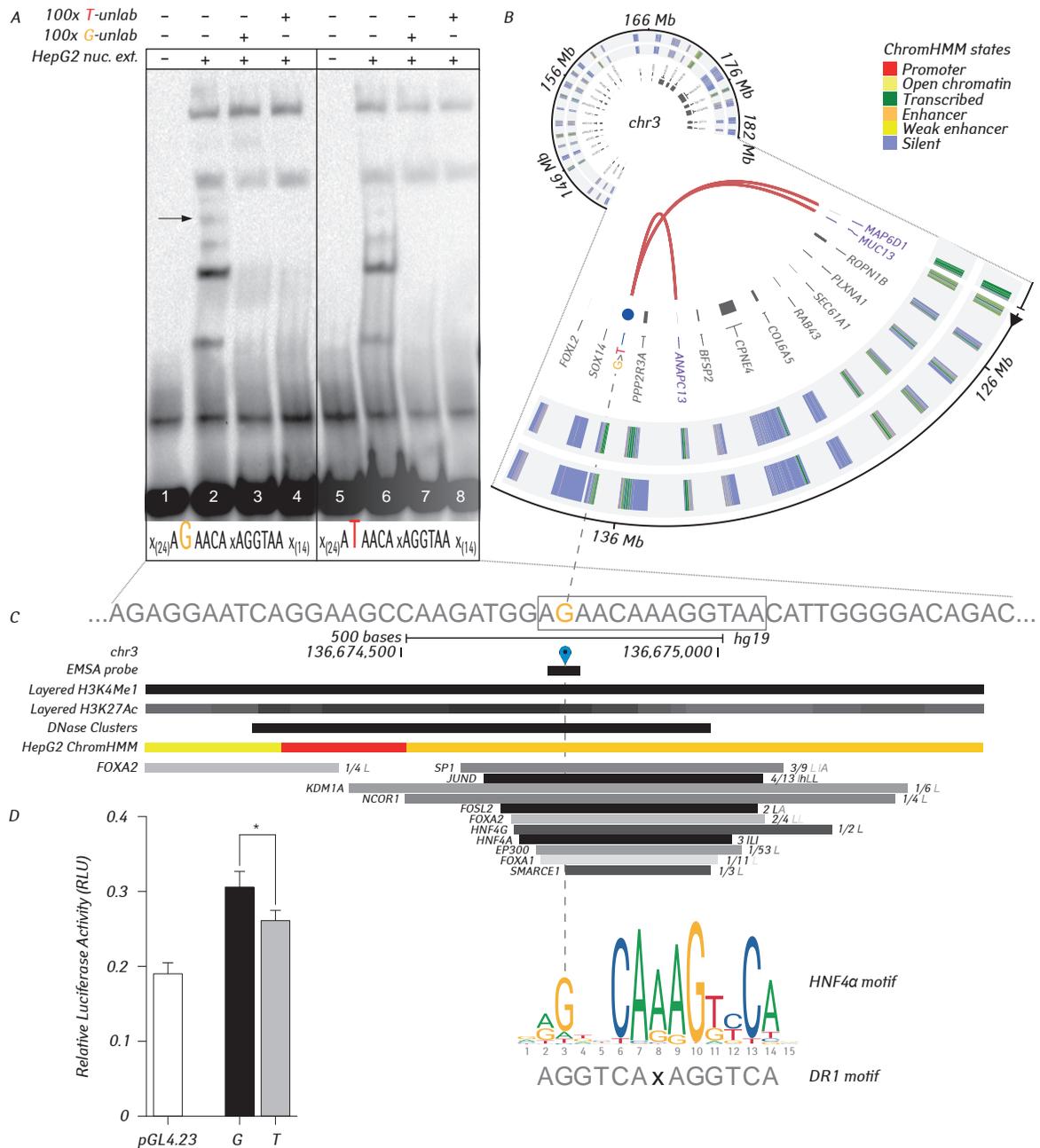


Figure. Effect of an HNF4α motif-breaking somatic mutation on TF binding and activity. **A.** EMSA for the G- and T-alleles. Lanes 3, 4 and 7, 8 represent a competition assay where a 100-fold molar excess of unlabeled probes was added to biotinylated probes and HepG2 nuclear protein extract. **B.** Circos plot illustrating interactions (red connections) of the active enhancer element harboring the HNF4α motif breaking somatic mutation G>T. From the outside in: the two external tracks show manually curated ChromHMM annotations for adult liver and HepG2 overlapping gene probes and the enhancer containing the G>T mutation. The inner track shows the experimental HiCap interactions for the enhancer. The triangle and the perpendicular line in the outer line of the plot mark the start and end of genomic coordinates. **C.** The genomic landscape of the G>T motif-breaking mutation. The UCSC genome browser tracks represent from the top: (i) the WT sequence for the EMSA probe; (ii) the coordinates of the EMSA probes, (iii) epigenetic markers of active enhancers (H3K4me1 and H3K27ac) and open chromatin (DNase clusters) from the ENCODE project, (iv) ChromHMM annotations in HepG2, (v) liver-specific transcription factor bindings from ChIP-seq experiments from the ENCODE project with the coloring (light grey to dark grey) proportional to the signal strength observed in different cell lines (cell abbreviations can be found at: <https://tinyurl.com/watv2v7>) and (vi) the HNF4α motif at the specified genomic coordinates. **D.** Dual luciferase assay testing the enhancer activity of two constructs with the G- and mutated T-alleles with respect to the empty control vector pGL4.23 (* $p < 1 \times 10^{-3}$)

Construction of cloning plasmids and luciferase report assays. All luciferase expression constructs were built based on pGL4.23 from Promega. Genomic sequences surrounding HNF4 α somatic mutation were amplified by Phusion Hot Start Flex DNA polymerase (NEB) using HepG2 genomic DNA as template (Table). The amplified fragments were inserted upstream of the minimal promoter sequence of pGL4.23 by SLiCE cloning methods [11]. The fragment harboring the somatic mutation (G>T) was constructed by site-directed mutagenesis using Phusion Hot Start Flex DNA polymerase (NEB).

HepG2 cells were transfected one day after plating with approximately 70% confluence in 96-well plate. Each well was transfected with 100 ng of firefly luciferase reporter vector harboring the reference and mutated alleles together with 1 ng of renilla luciferase reporter vector pGL4.74. Twenty-four hours after transfection, firefly and luciferase activity were measured by Dual-Luciferase[®] Reporter (DLR[™]) Assay System (Promega) on an Infinite[®] M200 PRO reader (TECAN) following instructions provided by the manufacturer. The ratios of firefly luciferase activity to renilla luciferase activity were calculated and expressed as Relative Luciferase Units in Figure. All data originated from six replicates, and *p*-values comparing Relative Luciferase Units difference between alleles were calculated using two-tailed *t*-test.

Electrophoresis mobility shift assay (EMSA).

Oligonucleotide probes were designed with the chr3:136674758 G>T somatic mutation flanked by 25 bp in both cold and 5'-biotinylated form (IDT) (Table 1). For the binding reaction, 3–6 μ g of HepG2 nuclear extract prepared from HepG2 cells using the Nuc-Buster[™] Protein Extraction kit (Novagen) were incubated with 200 fmol of each biotinylated dsDNA probe for 40 min on ice. For the competition assays, 20 pmol of unlabeled dsDNA probes were added to the binding reaction. DNA-protein complexes were cross-linked using UV-light and detected by chemiluminescence using LightShift[®] Chemiluminescent EMSA Kit (Thermo Scientific).

HiCap analysis.

Mechanically fine-ground human liver tissue was used for studying the folding of the chromatin and obtaining a list of long-range promoter-distal interactions by high throughput chromosome conformation capture coupled with subsequent targeted sequence capture (HiCap). For detailed information about the HiCap workflow and data analysis we refer to Cavalli *et al.* [12].

Table. Oligonucleotide sequences

EMSA probes	
X ₍₂₅₎ GX ₍₂₅₎	AGAGGAATCAGGAAGCCAAGATGGAGAACAAGG-TAACATTGGGGACAGAC
X ₍₂₅₎ GX ₍₂₅₎ -biotinylated	5'-biotin-AGAGGAATCAGGAAGCCAAGATGGAGAA CAAAGGTAACATTGGGGACAGAC
X ₍₂₅₎ TX ₍₂₅₎	AGAGGAATCAGGAAGCCAAGATGGATAACAAGG-TAACATTGGGGACAGAC
X ₍₂₅₎ TX ₍₂₅₎ -biotinylated	5'-biotin-AGAGGAATCAGGAAGCCAAGATGGATAAC AAAGGTAACATTGGGGACAGAC
Genomic sequence amplified to build the luciferase expression construct	
Forward	TGGCCTAAGTGGCCGTACCCATGCTGGAAGTGT-GAGCAGA
Reverse	TGTCTAGATCTTGATATCGTCTCAGGTTGATTGGTGCA
Site directed mutagenesis to create the G>T mutation	
GtoT-forward	CCAAGATGGATAACAAGGTAACATTGGGGA
GtoT-reverse	GTTACCTTTGTTATCCATCTTGGCTTCTGA

Briefly, liver cells were crosslinked in 1% formaldehyde solution, lysed, passed through a 27-gauge needle to dissociate agglomerates and incubated for 10 min on ice. The resulting nuclei solution was washed in PBS and incubated with the fast digest MboI endonuclease (ThermoFisher Scientific) to digest the chromatin at the restriction sites. Protruding 5' ends of the DNA left by the endonuclease were filled using biotinylated nucleosides followed by blunt-end chromatin complex intra-molecular ligation (proximity ligation). After ligation, the samples were de-crosslinked and remaining intact RNA was removed by treatment with RNase A (ThermoFisher Scientific). Resulting chimeric DNA constructs were subsequently sonicated to achieve fragments around 200 bp in length. KAPA HTP Library preparation kit for Illumina platforms was used to prepare NGS-compatible libraries by following manufacturer's protocol. Targeted sequence capture was performed to further enrich the libraries of interest using SureSelect XT Target Enrichment System for Illumina Paired-End Multiplexed Sequencing libraries (Agilent). Libraries were sequenced by Illumina single TruSeq LT index, paired-end sequencing on NextSeq 500 platform (Illumina).

RESULTS AND DISCUSSION

As a proof of concept of the central role of somatic mutations in non-coding regulatory elements of the genome, we experimentally investigated how a mutation in the first half site of the HNF4 α DR1 motif, alters protein binding and enhancer activity of the regulatory element harboring it.

HNF4 α belongs to the superfamily of nuclear receptors that bind specific DNA sequences consisting of two hexanucleotide half-site motifs commonly arranged in direct (DR, $\rightarrow \rightarrow$) or inverted (IR, $\rightarrow \leftarrow$) configurations with different spacing between them [13]. HNF4 α binds DR1 motif (AGGTCxAGGTCA), a direct repeat DNA element separated by one base pair [14].

The HNF4 α protein contains a zinc-finger domain and a hinge region connecting the DNA binding domain (DBD) to the ligand-binding and a dimerization region. Crystal structures [14–16] have shown that the DBD makes several contacts with the DR1 motif. These protein-DNA interactions are supported by H-bonds between amino acid chains in the DBD and both the DNA backbone or bases [16].

Mutations in the gene for HNF4 α have been implicated in several diseases. Loss-of-function mutations in the zinc-finger domain have been linked to altered transcriptional regulation in liver cancer [17] and point mutations in the DBD have been associated to Maturity Onset of Diabetes of the Young 1, hyperinsulinemic hypoglycemia [16]. Altered gene regulation associated with HNF4 α can also result from a mutation in the binding motif of the TF which can alter the TF-DNA interaction.

Mutations at position 3 G>T of the HNF4 α DR1 motif are frequently observed in HCC samples and we focused on an instance of this mutation located at chr3:136674758 (GRCh37/hg19). Based on the Roadmap Epigenomics project the genomic element harboring the mutation is annotated as an active enhancer (Figure, C) in HepG2, HCC cell line (EID: E118) and in adult liver tis-

sue (EID: E066). We studied the DNA-protein interaction by EMSA and found that the G>T mutation altered the protein binding to the mutated probe (Figure, A). We further studied the enhancer activity using a luciferase assay and verified that the regulatory element with the normal sequence has enhancer activity which is higher than the negative control pGL4.23. The enhancer activity was diminished with the introduction of the G>T mutation by site-directed mutagenesis (Figure, D).

Next, we investigated the putative target gene(s) that this enhancer would act on as a regulator based on HiCap data in liver tissue [12]. We found that the HNF4 α motif is located in an enhancer that interacts with the promoters of three genes: *ANAPC13*, *MAP6D1* and *MUC13* (Figure, B), that is consistent with other studies that have reported single enhancers regulating the activity of multiple genes [1].

ANAPC13 encodes for the Anaphase Promoting Complex Subunit 13, a controller of the cell-cycle by regulating the ubiquitin mediated degradation of B-type cyclins. ANAPC13 has been reported as an unfavorable prognostic marker for liver cancer [18] and is associated to chronic pancreatitis [19].

MAP6D1 encodes for the MAP6 Domain-Containing Protein 1, which acts as a calmodulin-regulated protein that binds and stabilizes microtubules. The expression of *MAP6D1* is up regulated in initial liver fibrosis [20] and the protein is considered a bio-marker for stages F2/F3 of liver fibrosis in a step-wise process that can lead to cirrhosis and HCC [21].

The epithelial protein mucin 13 involved in transmembrane signaling is encoded by the *MUC13* gene. Enhanced expression of *MUC13* has been reported in colorectal, stomach, ovarian cancer and liver metastatic tissue [22]. Overexpression of *MUC13* is considered a poor prognostic predictor [23] and promotes malignant growth and metastasis by upregulating/activating key oncogenes and signaling pathways [24].

CONCLUSIONS

Mutations at position 3 in the binding site for HNF4 α are recurring events in liver cancer. We have shown how such a mutation at chr3:136674758 leads to decreased protein binding to the mutated allele. The mutation is located in an enhancer active in liver cells and the mutation decreases the enhancer activity. Studies of long-range interactions suggested that the enhancer interacts with the promoters of the genes for *ANAPC13*, *MAP6D1* and *MUC13*, that are all implicated in liver cancer based on cell cycling, cell structure and cell signaling. We believe that this somatic mutation is one of many present in non-coding regulatory elements, which may contribute to cancer development and progression.

CONFLICT OF INTEREST

The author(s) declare that they have no conflict of interest.

ACKNOWLEDGEMENTS

The work was supported by grants from SciLife-Lab (CW), The Swedish Cancer Foundation [CAN 2015/759, 2018/849] (CW), The National Science Centre [DEC-2015/16/W/NZ2/00314] (JK), Institute

of Computer Science, Polish Academy of Sciences (JK), The eSSence program (JK). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

- Dunham I, Kundaje A, Aldred SF, *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012; **489**: 57–74.
- Kundaje A, Meuleman W, Ernst J, *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* 2015; **518**: 317–30.
- Rheinbay E, Nielsen MM, Abascal F, *et al.* Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature* 2020; **578**: 102–11.
- Umer HM, Cavalli M, Dabrowski MJ, *et al.* A significant regulatory mutation burden at a high-affinity position of the CTCF motif in gastrointestinal cancers. *Hum Mutat* 2016; **37**: 904–13.
- Lieberman-Aiden E, van Berkum NL, Williams L, *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 2009; **326**: 289–93.
- Sahlén P, Abdullayev I, Ramsköld D, *et al.* Genome-wide mapping of promoter-anchored interactions with close to single-enhancer resolution. *Genome Biol* 2015; **16**: 156.
- Jäger R, Migliorini G, Henrion M, *et al.* Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. *Nat Commun* 2015; **6**: 6178.
- Dryden NH, Broome LR, Dudbridge F, *et al.* Unbiased analysis of potential targets of breast cancer susceptibility loci by Capture Hi-C. *Genome Res* 2014; **24**: 1854–68.
- Weinhold N, Jacobsen A, Schultz N, *et al.* Genome-wide analysis of noncoding regulatory mutations in cancer. *Nat Genet* 2014; **46**: 1160–5.
- Chalmers ZR, Connelly CF, Fabrizio D, *et al.* Analysis of 100,000 human cancer genomes reveals the landscape of tumour mutational burden. *Genome Med* 2017; **9**: 34.
- Zhang Y, Werling U, Edelmann W. SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res* 2012; **40**: e55.
- Cavalli M, Diamanti K, Pan G, *et al.* A Multi-Omics approach to liver diseases: integration of single nuclei transcriptomics with proteomics and HiCap bulk data in human liver. *OMICS* 2020; **24**: 180–94.
- Fang B, Mane-Padros D, Bolotin E, *et al.* Identification of a binding motif specific to HNF4 by comparative analysis of multiple nuclear receptors. *Nucleic Acids Res* 2012; **40**: 5343–56.
- Lu P, Rha GB, Melikishvili M, *et al.* Structural basis of natural promoter recognition by a unique nuclear receptor, HNF4 α . *Diabetes gene product*. *J Biol Chem* 2008; **283**: 33685–97.
- Rastinejad F, Wagner T, Zhao Q, *et al.* Structure of the RXR-RAR DNA-binding complex on the retinoic acid response element DR1. *EMBO J* 2000; **19**: 1045–54.
- Chandra V, Huang P, Potluri N, *et al.* Multidomain integration in the structure of the HNF-4 α nuclear receptor complex. *Nature* 2013; **495**: 394–8.
- Taniguchi H, Fujimoto A, Kono H, *et al.* Loss-of-function mutations in Zn-finger DNA-binding domain of HNF4A cause aberrant transcriptional regulation in liver cancer. *Oncotarget* 2018; **9**: 26144–56.
- Uhlen M, Zhang C, Lee S, *et al.* A pathology atlas of the human cancer transcriptome. *Science* 2017; **357**: eaan2507.
- Wang D, Xin L, Lin J-H, *et al.* Identifying miRNA-mRNA regulation network of chronic pancreatitis based on the significant functional expression. *Medicine* 2017; **96**: 21.
- Ahmad W, Ijaz B, Hassan S. Gene expression profiling of HCV genotype 3a initial liver fibrosis and cirrhosis patients using microarray. *J Transl Med* 2012; **10**: 41.

21. **Ijaz B, Ahmad W, Das T, et al.** HCV infection causes cirrhosis in human by step-wise regulation of host genes involved in cellular functioning and defense during fibrosis: Identification of bio-markers. *Genes Dis* 2019; **6**: 304–17.

22. **Gupta BK, Maher DM, Ebeling MC, et al.** Increased expression and aberrant localization of mucin 13 in metastatic colon cancer. *J Histochem Cytochem* 2012; **60**: 822–31.

23. **Dai Y, Liu L, Zeng T, et al.** Overexpression of MUC13, a poor prognostic predictor, promotes cell growth by activating wnt signaling in hepatocellular carcinoma. *Am J Pathol* 2018; **188**: 378–91.

24. **Tiemin P, Fanzheng M, Peng X, et al.** MUC13 promotes intrahepatic cholangiocarcinoma progression via EGFR/PI3K/AKT pathways. *J Hepatol* 2020; **72**: 761–73.

МУТАЦІЇ В НЕКОДУЮЧИХ ДІЛЯНКАХ ГЕНОМУ КЛІТИН, ЩО ПОРУШУЮТЬ ЗВ'ЯЗУВАННЯ З ТРАНСКРИПЦІЙНИМ ФАКТОРОМ HNF4A, ВПЛИВАЮТЬ НА АКТИВНІСТЬ ЕНХАНСЕРІВ ГЕНІВ, ЯКІ АСОЦІЙОВАНІ З ПРОГРЕСУВАННЯМ РАКУ ПЕЧІНКИ

М. Каваллі^{1,}, К. Діаманті¹, Г. Пан¹, М.Ж. Дабровські^{1,2}, Ж. Коморовські^{2,3,4}, К. Ваделіус^{1,*}*

¹Уппсальський університет, 752 37, Уппсала, Швеція

²Інститут інформатики, Польська Академія наук, 01-248 Варшава, Польща

³Шведський колегіум перспективних досліджень, 752 38, Уппсала, Швеція

⁴Вашингтонський національний дослідницький центр приматів, Сіетл, штат Вашингтон 98121, США

Стан питання: Соматичні мутації в кодуючих ділянках гену можуть призводити до появи нефункціональних білків, що може спричинити онкологічні та інші захворювання. Стосовно мутацій у некодуючих ділянках, то їх досліджують

значно рідше, а інтерпретація ефектів, які вони спричиняють, є досить складною. Мутації в некодуючих ділянках можуть порушувати зв'язування транскрипційних факторів з відповідними промоторами, енхансерами або сайленсерами, що може призводити до зміни експресії генів, які знаходяться під контролем цих регулюючих елементів. У клітинах раку печінки виявляють велику кількість мутацій як в кодуючих, так і в некодуючих ділянках геному. Ядерний фактор гепатоцитів 4а є фактором транскрипції, що регулює експресію низки генів в гепатоцитах, причому мотиви в промоторах та енхансерах, з якими цей фактор зв'язується, досить часто містять мутації. **Мета:** Оцінити генетичні ефекти соматичних мутацій у некодуючих ділянках геному, які часто виявляють у клітинах раку печінки. **Матеріали і методи:** В експерименті досліджували ефекти соматичних мутацій в некодуючих ділянках геному клітин раку печінки на зв'язування відповідних мотивів з ядерним фактором гепатоцитів 4а. Дослідження проводили на моделі клітин НерG2 за допомогою методів зсуву електрофоретичної рухливості та подвійної люциферазної детекції. Повногеномні промоторно-енхансерні взаємодії в клітинах печінки вивчали методом фіксації конформації хромосом для виявлення можливих генів, експресія яких може змінюватися внаслідок мутацій. **Результати:** Показано, що досліджувані мутації призводять до послаблення зв'язування з білком та зниження активності енхансера. Так, мутований енхансер взаємодіє з промоторами генів *ANAPC13*, *MAP6D1* і *MUC13*, які задіяні в розвитку раку печінки. **Висновки:** Показана важливість соматичних мутацій у некодуючих ділянках геному, вивченню яких дотепер не приділяли достатньої уваги, та які, разом з тим, задіяні в розвитку та прогресуванні процесів злоякісного росту.

Ключові слова: мутації, що порушують зв'язування з транскрипційними факторами, регуляції активності генів, рак печінки.