UDC 57.032

NGS-based identification of druggable alterations and signaling pathways – hepatocellular carcinoma case report

E. A. Kotelnikova1,5, M. D. Logacheva1,2, E. R. Nabieva2, M. A. Pyatnitskiy3,5,
D. V. Vinogradov1,5, A. S. Makarova2,4,5, A. V. Demin5, A. G. Paleeva5, O. S. Kremenetskaya5,6,
A. A. Penin1,2,7, A. V. Klepikova1,2, A. S. Kaskanov2, D. A. Shavochkina4, N. E. Kudashkin4,

1 A. A. Kharkevich Institute for Information Transmission Problems,
19/1, Bolshoy Karetny per. Moscow, Russian Federation, 127051
2 A. N. Belozersky Institute of Physico-Chemical Biology, M. V. Lomonosov Moscow State University,
1/40, Leninskie gory, Moscow, Russian Federation, 119992
3 Orekhovich Institute of Biomedical Chemistry,
10/8, Pogodinskaya Str., Moscow, Russian Federation, 119121
4 N. N. Blokhin Russian Cancer Research Centre,
24, Kashirskeoye shosse, Moscow, Russian Federation, 115478
5 ZAO Personal Biomedicine,
124/17, Prospekt Mira, Moscow, Russian Federation, 129164
6 Center for Theoretical Problems of Physicochemical Pharmacology RAS,
4, Kosygin Str., Moscow, Russian Federation, 119991
7 Biological Faculty, M. V. Lomonosov Moscow State University
1/12, Leninskie Gory, Moscow, Russian Federation, 119991

ekotelnikova@gmail.com

Aim. To identify potential cancer driving or clinically relevant molecular events for a patient with hepatocellular carcinoma. Methods. In order to achieve this goal, we performed RNA-seq and exome sequencing for the tumor tissue and its matched control. We annotated the alterations found using several publicly available databases and bioinformatics tools. Results. We identified several differentially expressed genes linked to the classical sorafenib treatment as well as additional pathways potentially druggable by therapies studied in clinical trials (Erlotinib, Lapatinib and Temsirolimus). Several germline mutations, found in XRCC1, TP53 and DPYD, according to the data from other clinical trials, could be related to the increased sensitivity to platinum therapies and reduced sensitivity to 5-Fluorouracil. We also identified several potentially driving mutations that could not be currently linked to therapies, like deletion in CIRBP, SNVs in BTG1, ERBB3, TCF7L2 et al. Conclusions. The presented study shows the potential usefulness of the integrated approach to the NGS data analysis, including the analysis of germline mutations and transcriptome in addition to the cancer panel or the exome sequencing data.

Keywords: NGS, cancer, systems biology, pathways, pharmacogenetics, personalized medicine

Introduction

Carcinogenesis is considered to be caused by alterations in specific genes associated with dysfunction of regulatory networks [1]. Therefore, reconstruction of regulatory interactions is necessary for understanding the processes of carcinogenesis in addition to the identification of molecular targets for the
antineoplastic drugs. The systems biology analysis of transcriptomic data makes it possible to identify and interpret the effects of mutations and gene expression deregulation. In cancer research, the goal of systems biology is to decipher the impact of genetic and epigenetic aberrations in cancer cells on their homeostasis, intercommunication and response to possible treatments [2]. This approach is particularly important for precision oncology, since each tumor is unique in terms of genetics and pathological regulation of signaling pathways. The reconstruction of the patient-specific signaling pathways could help clinicians to identify the most effective treatment.

One of the interdisciplinary tools of system biology is known as the next-generation sequencing (NGS) technology. NGS platforms perform massively parallel sequencing, so millions of DNA fragments are sequenced at a time. Such large-scale sequence analysis of the genome and transcriptome is vital for developing effective strategies in personalized cancer therapy. Specifically, this NGS-oriented approach is important for choosing between the treatment schemes, when selecting patients are likely to benefit from targeted therapies [3]. The personalized NGS-based analysis promotes clinical decisions when standard therapy does not give the expected results or leads to tumor resistance.

Hepatocellular carcinoma (HCC) is one of the most often diagnosed types of liver cancer and occupies the 6th place in frequency of all cancer types [4]. In this work we aimed to identify potential cancer driving or clinically relevant molecular events for a patient with HCC using NGS technology.

Materials and Methods

Samples collection and extraction of RNA/DNA

Genomic DNA and total RNA were isolated from fresh-frozen samples of hepatitis-negative HCC and adjacent non-cancerous tissue liver using Wizard SV Genomic DNA Purification System, Promega and PureLink RNA Mini Kit, Life Technologies with DNase treatment, respectively. Samples were collected from 66 years old male patient with histologically verified moderately differentiated HCC after tumor resection with informed consent, conforming to the ethical guidelines of the 1975 Declaration of Helsinki.

RNA quality was checked using Agilent 2100 Bioanalyzer; only samples with RIN (RNA integrity number) > 7 were taken for analysis. Before library preparation, ribosomal RNA was removed using Ribo-Zero Gold rRNA Removal Kit (Epicentre). rRNA-depleted RNA was then processed using TruSeq Stranded mRNA Library Prep Kit (Illumina). Libraries were sequenced on HiSeq2000 instrument with TruSeq v. 3 chemistry. Read length was 101 from each end of the fragment.

Read processing

Before calling SNVs and indels, sequencing reads were trimmed [5] and aligned to the hg19 reference genome with bowtie2 [6]; the alignment was thereupon deduplicated, indel-realigned and base-quality recalibrated [7].

SNV and indel calling

In order to identify somatic and germline single nucleotide variants insertions/deletions, we ran VarScan2 [8] in the somatic mode on tumor and control samples. The discovered variants were annotated using the Annovar [9]. The following parameters were used:

• VarScan p-value < 0.05 (somatic p-value for somatic variants, variant p-value for germline variants)
• Fraction of reads with alternative allele found in tumor sample > 20 %
• Variant belonging to exonic or splicing region
• >10 reads for alternative allele in tumor sample

Identification of damaging mutations

In order to assess mutation impact upon a protein function we utilized MutationAssessor [10] and PolyPhen2 [11]. Additionally CHASM [12] software was used to differentiate between potential driver and passenger mutations. The following filters were applied: MutationAssessor score classification is high, low or medium OR Polyphen2 class is “delete-
rious”, OR CHASM score is less than 0.5, OR mutation is “nonsense”.

**Differential expression**

For the differential expression analysis we followed the popular protocol [13], using Tophat2 for reads mapping and DESeq [14] for discovering genes with significantly different expression levels. We used 0.05 as a threshold for p-value, and left only genes for which expression levels ratio between normal and cancer tissues exceed 2. We also calculated logratio for each gene as log2 (expr. in tumor)/(expr. in normal).

![Fig. 1. The distribution of genes with altered expression across different cancer hallmark processes.](image)
Results and Discussion

RNA differential expression data analysis

As a result of RNA-seq data analysis we have identified 497 upregulated and 359 downregulated differentially expressed genes with FDR<0.05. No clear markers of pharmacological response (either FDA or preclinical) were found among them. In order to get indirect evidences about favorable pharmacological interventions we have classified obtained genes using different cancer hallmark processes (see Fig. 1) and checked the expression of genes, related to the pathways implicated in HCC treatment responses.

Sorafenib is a multikinase inhibitor and the first target drug approved by the FDA for the HCC treatment [15]. In the studied tumor sample, PDGFα gene is upregulated relative to the control values, supporting the potential activation of the PDGF-signaling. We checked the CTD database [16] in order to define other cancer-driving differentially expressed genes, potentially affected by sorafenib action. Among the overexpressed genes is BIRC5 which is a negative regulator of apoptosis that prevents apoptotic cell death and that can be down-regulated by sorafenib [17]. Sorafenib can also inhibit HCC cell proliferation by blocking RAS/RAF/MAPK and PI3K/AKT/mTOR pathways activated by overexpressed growth factor EGF [18]. However, the genes described above could not be used for evaluation of sorafenib effectiveness in this case.

Alternatively, overexpressed EGF gene is a marker of EGFR/ERBB cascade activation with down-stream PI3K/AKT1/mTOR and JAK/STAT signaling. In general, these cascades could be targeted by EGFR and ERBB2-inhibiting drugs Erlotinib and Lapatinib [19]. A drug specific for PI3K/AKT1/mTOR inhibition, Temsirolimus, could be specifically important because of the sorafenib ineffectiveness for this cascade. We further discuss the EGFR cascade and the corresponding drugs below in the context of the found genetic alterations.

Table 1. Identified somatic variants

<table>
<thead>
<tr>
<th>Chromosome position</th>
<th>Gene symbol</th>
<th>Normal haplotype</th>
<th>Tumor haplotype</th>
<th>Aminoacid change</th>
<th>Effect predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr10_123324040</td>
<td>FGFR2</td>
<td>C/C</td>
<td>C/A</td>
<td>V55F</td>
<td>MA</td>
</tr>
<tr>
<td>chr12_56493724</td>
<td>ERBB3</td>
<td>G/G</td>
<td>G/A</td>
<td>D1014N</td>
<td>CHASM</td>
</tr>
<tr>
<td>chr12_92537924</td>
<td>BTG1</td>
<td>T/T</td>
<td>T/A</td>
<td>K150*</td>
<td>Nonsense</td>
</tr>
<tr>
<td>chr7_94041987</td>
<td>COL1A2</td>
<td>G/G</td>
<td>G/A</td>
<td>G499D</td>
<td>MA</td>
</tr>
<tr>
<td>chr10_114912166</td>
<td>TCF7L2</td>
<td>A/A</td>
<td>A/T</td>
<td>Q355H</td>
<td>MA</td>
</tr>
<tr>
<td>chr15_75091004</td>
<td>CSK</td>
<td>G/G</td>
<td>G/C</td>
<td>G22R</td>
<td>MA, CHASM, PP2</td>
</tr>
<tr>
<td>chr19_1271421</td>
<td>CIRBP</td>
<td>G/G</td>
<td>G/A</td>
<td>G102S</td>
<td>MA</td>
</tr>
<tr>
<td>chr2_132240363</td>
<td>TUBA3D</td>
<td>A/A</td>
<td>A/G</td>
<td>Y432C</td>
<td>MA, PP2</td>
</tr>
<tr>
<td>chr2_43520196</td>
<td>THADA</td>
<td>G/G</td>
<td>G/T</td>
<td>A1532D</td>
<td>MA</td>
</tr>
<tr>
<td>chr2_55867797</td>
<td>PNPT1</td>
<td>C/C</td>
<td>C/G</td>
<td>V705L</td>
<td>MA</td>
</tr>
<tr>
<td>chr4_164085514</td>
<td>NAF1</td>
<td>G/G</td>
<td>G/A</td>
<td>S132L</td>
<td>MA, PP2</td>
</tr>
<tr>
<td>chr7_73442522</td>
<td>ELN</td>
<td>C/C</td>
<td>C/T</td>
<td>A2V</td>
<td>MA</td>
</tr>
<tr>
<td>chr12_124422296</td>
<td>CCDC92</td>
<td>T/T</td>
<td>T/C</td>
<td>N102S</td>
<td>MA</td>
</tr>
<tr>
<td>chr13_52516523</td>
<td>ATP7B</td>
<td>T/T</td>
<td>T/A</td>
<td>K930N</td>
<td>MA</td>
</tr>
<tr>
<td>chr19_53014336</td>
<td>ZNF578</td>
<td>T/T</td>
<td>T/A</td>
<td>N234K</td>
<td>MA, PP2</td>
</tr>
<tr>
<td>chr5_141335930</td>
<td>PCDH12</td>
<td>G/G</td>
<td>G/A</td>
<td>S496L</td>
<td>MA, PP2</td>
</tr>
<tr>
<td>chr8_142200495</td>
<td>DENND3</td>
<td>C/C</td>
<td>C/T</td>
<td>H1040Y</td>
<td>MA</td>
</tr>
<tr>
<td>chr9_97081966</td>
<td>FAM22F</td>
<td>G/G</td>
<td>G/A</td>
<td>P472S</td>
<td>MA</td>
</tr>
</tbody>
</table>
Somatic SNVs and InDels

Exome sequencing revealed 9250 SNVs in the exonic or splicing regions, 77 somatic and 9173 germline variants. In order to identify somatic SNVs, potentially driving the cancer progression, we first filtered out dbSNP and silent mutations, leaving 23 missense or nonsense SNVs. Among these variants in the exonic or splicing regions, we identified 18 (see Table 1), predicted to be damaging by at least one of these tools: PolyPhen2 (PP2), MutationAccess (MA) or CHASM (see Materials and Methods).

Using filtering, described in Materials and Methods, we have also identified 3 deletions in exonic regions, described in Table 2.

Somatically disturbed molecular pathways

All somatic SNVs and indels were manually curated in order to identify possible cancer driving pathways and potential pharmacological interventions. Some of the examples are presented below.

ERBB3 and EGFR pathway

EGFR/ERBB1, ERBB2 and ERBB3 comprise an EGFR family of tyrosine kinases. Interacting with corresponding ligands and forming the functional homo and hetero-dimers, EGFR/ERBB-receptors could transfer the signal inside the cell, regulating proliferation, migration and apoptosis. ERRB3, mutated in the studied tumor sample, can bind to the ligands but does not have its own kinase activity. Thus, ERBB3 could activate the downstream signaling only in complex with other ERBB receptors [20].

Mutation in ERBB3 is found as potentially driving by CHASM and statistically significant overexpression of EGF as well as less significant but coordinated overexpression of other members of this cascade, could characterize the aberrant activation of this mechanism in studied tumor.

The main signaling cascades activated downstream of EGFRs are PI3K/AKT1, MAP-kinase, and JAK/STAT (see Fig. 2). The activation of these cascades leads to the inhibition of apoptosis, uncontrolled cells proliferation and other pro-oncogenic processes. This activity can be suppressed by EGFR and ERBB2 inhibitors – Erlotinib and Lapatinib [19]. There are several ongoing clinical trials, where these drugs are used as a second line therapy of HCC or in combination with sorafenib.

Alternatively, taking into account the PDGFA overexpression, the switch to the MTOR signaling is one of the probable scenarios. This cascade and its downstream targets could be suppressed by Temsirolimus. It could be specifically important because the mTOR activity is not targeted by standard sorafenib treatment. There are several clinical trials, where temsirolimus is used in combination with sorafenib for HCC treatment (NCT01008917).

BTG1 – potential driver

The gene BTG1 interacts with several nuclear receptors that could regulate differentiation of the cells [21], see Fig.3. The somatic nonsense mutation K150*(chr12: 92537924) in BTG1 is probably damaging. It leads to the partial deletion of C-terminal region that is necessary for the BTG1 accumulation in nucleus and interaction with other proteins [22]. Among the negative targets of BTG1 are antiapoptotic genes MPP9, BCL2 and CCND1, that could switch the tumor cells behavior towards the proliferative mode in response to the damaging BTG1 mutation. Additionally, BTG1 is shown to be downregulated in HCC [23]. Summarizing, these evidences support the hypothesis about BTG1 as a driver gene in the case studied.

Table 2. Somatic indels

<table>
<thead>
<tr>
<th>Chromosome position</th>
<th>Gene symbol</th>
<th>Normal haplotype</th>
<th>Tumor haplotype</th>
<th>Variant Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr19_1271419</td>
<td>CIRBP</td>
<td>G/G</td>
<td>G/-</td>
<td>Frame_Shift_Del</td>
</tr>
<tr>
<td>chr11_64032972</td>
<td>PLCB3</td>
<td>CCT/CCT</td>
<td>CCT/-</td>
<td>In_Frame_Del</td>
</tr>
<tr>
<td>chr12_56559304</td>
<td>SMARCC2</td>
<td>C/C</td>
<td>C/-</td>
<td>Frame:Shift:Del</td>
</tr>
</tbody>
</table>

440
FGFR2
FGFR2, a receptor tyrosine kinase, regulates proliferation, differentiation, migration and apoptosis. FGFR2 expression in HCC is associated with unfavorable prognosis [24]. The detected SNV in FGFR2 – V144F is considered as damaging. Possible activation of FGFR2 cascade provided by its ligands FGF2 and FGF7 may suggest tyrosine-kinase inhibitors therapy.

CIRBP
Somatic deletion in the CIRBP gene alters the polypeptide chain starting with the 101st residue, damaging the RGG domain that operates mRNA stability.
and modulates translation of CIRBP targets. Some convincing confirmation of CIRBP mutation supported by transcriptomes data (see Fig. 4) and variability of the processes regulated by CIRBP allows us to suppose that the mutation in question may play definite role in carcinogenesis.

**Germline SNVs**

Among the 9173 found germline SNVs in exonic regions we identified those 13 variants (Table 3) which were relevant to the drug toxicity and resistance according to PharmGKB database [25].

In the studied case a possible effect of TP53 and DPYD germline mutations on tumor sensitivity to 5-fluorouracil was analyzed using information from scientific literature. Somatic SNV in the gene DPYD (C29R) activates the DPYD enzyme, which rapidly converts 5-FU to its inactive metabolite 5-dihydrofluorouracil [26]. The identified TP53 polymorphism (R72P) also reduces the efficacy of the 5-FU therapy [27]. Accordingly, the use of 5-FU therapy is likely to be ineffective in this case (see Fig. 5). SNV in the gene XRCC1 (R399Q) could be related to sensitivity to platinum therapies [28]. Other germline SNVs also might be associated with therapy toxicity and adverse drug reactions. SNV in the gene MTHFR (E429A) might be associated with an increased risk of myelosuppression in the patients treated with methotrexate [29]. SNV in CDA (K27Q) was shown to be associated with an increased severity of hematological toxicity, including neutropenia, in patients with pancreatic neoplasms treated with gemcitabine or cytarabine [30]. SNV in XPC (Q902K), SLC22A2 (S270A), XRCC1 (R194W), LRP2 (K4094E) might be associated with an increased risk of drug toxicity when treated with cisplatin [31–33]. SNV in UMPS (G213A) could be related with the
increased likelihood of drug toxicity when treated with fluorouracil and leucovorin. \textit{ERBB2} polymorphism (I625V) may be associated with cardiotoxicity under trastuzumab treatment. \textit{SLC19A1} polymorphism (H27R) might be related with drug toxicity under methotrexate and mercaptopurine treatment.

**Table 3. Identified germline variants**

<table>
<thead>
<tr>
<th>Chromosome position</th>
<th>Symbol</th>
<th>Normal haplotype</th>
<th>Tumor haplotype</th>
<th>Relevant drugs</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr1_11854476</td>
<td>MTHFR</td>
<td>T</td>
<td>G</td>
<td>T</td>
</tr>
<tr>
<td>chr1_20915701</td>
<td>CDA</td>
<td>A</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>chr1_98348885</td>
<td>DPYD</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>chr2_170010985</td>
<td>LRP2</td>
<td>T</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>chr3_14187449</td>
<td>XPC</td>
<td>G</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td>chr3_124456742</td>
<td>UMPS</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>chr6_160670282</td>
<td>SLC22A2</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>chr7_7579472</td>
<td>TP53</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>chr7_37879588</td>
<td>ERBB2</td>
<td>A</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>chr19_44055726</td>
<td>XRCC1</td>
<td>T</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>chr19_44057574</td>
<td>XRCC1</td>
<td>G</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>chr21_46957794</td>
<td>SLC19A1</td>
<td>T</td>
<td>C</td>
<td>T</td>
</tr>
</tbody>
</table>
**Conclusion**

The presented study shows the potential usefulness of the integrated approach to the NGS data analysis, including the analysis of germline mutations and transcriptome in addition to the genome sequencing data. The expression profile of tumor genes corresponds to the spectrum of inhibitory activity of the Sorafenib. Additionally, the potentially effective drugs are Carboplatin, Oxaliplatin, Cisplatin (an increased sensitivity to platinum drugs is associated with the polymorphism in XRCC1); Temsirolimus (inhibitor of PI3K/AKT/mTOR signaling); Erlotinib, Lapatinib (inhibitors of ERBB cascades). 5-Fluorouracil therapy is potentially ineffective in connection with the identified polymorphisms in the TP53 and DPYD genes.

**Funding**

The work was partly supported by grant from Russian Ministry of Education and Science (contract 14.607.21.0049, RFMEFI60714X0049).
### REFERENCES

ідентифікували декілька потенційно драй-верних мутацій, які на цей час не можуть бути пов’язані з тера-піями, наприклад делєції у CIRBP, заміни в BTG1, ERBB3, TCF7L2 тощо. Висновки. Запропоноване дослідження демон-струє потенційну корисність інтегрованого підходу до NGS аналізу даних, в тому числі аналізу гермінативних мутацій та транскрипту до додатку до використання онкологічних ген-них панелей або даних секвєнування екзому.

Найважливіші результати. Дане дослідження демон-струє потенційну корисність інтегрованого підходу до NGS аналізу даних, в тому числі аналізу гермінативних мутацій та транскрипту до додатку до використання онкологічних генних панелей або даних секвєнування екзому.

Ключові слова: NGS, онкологія, системна біологія, сиг-нальні шляхи, фармакогенетика, персоналізована медицина

Ідентифікація крінічно значущих порушення і сигнальних каскадів на основі NGS на прикладі клінічного випадку гепатоцелюлярної карциноми

О. А. Котельникова, М. Д. Логacheva, Е. Р. Набєева, Н. А. Пятницкий, Д. В. Виноградов, А. С. Макарова, А. В. Демин, А. Г. Палеева, О. С. Кременецька, А. А. Пенин, А. В. Кліпікова, А. С. Касьянов, Д. А. Шавочкина, Н. Е. Кудашкин, Ю. И. Патютко, Н. С. Мюге, А. С. Кондрашов, Н. Л. Лазаревич

Мета. Ідентифікувати потенційно онкодрайверні або клінічно значущі молекулярні події у пацієнта з гепатоцелюлярною карциномою. Методи. РНК- та екзомне секвєндування пухлини та транскрипту в додаток до використання онкологічних генних панелей і біоінформатичних програм. Результати. Ми виявили декілька генів, що диференційно виражені у пухляні, які відповідають ключовим роль в механізмах росту та розповсюдження гепатоцелюлярної карциноми. Висновки. Запропоноване дослідження демонструє потенційну корисність інтегрованого підходу до NGS аналізу даних, в тому числі аналізу гермінативних мутацій та транскрипту до додатку до використання онкологічних генних панелей або даних секвєнування екзому.

Ключові слова: NGS, рак, системна біологія, сигнальні шляхи, фармакогенетика, персоналізована медицина

Received 10.10.2015