Genomics Data 6 (2015) 1–3

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

Genomic and transcriptome profiling identified both human and HBV genetic variations and their interactions in Chinese hepatocellular carcinoma

Hua Dong [a], Ziliang Qian [a], Lan Zhang [b], Yunqin Chen [c], Zhenggang Ren [b], Qunsheng Ji [a,*]

[a] AstraZeneca Asian and Emerging Market iMed, Zhangjiang Hi-Tech Park, Shanghai, PR China
[b] Liver Cancer Institute, Zhongshan Hospital, Key Laboratory of Carcinogenesis and Cancer Invasion, Ministry of Education, Fudan University, Shanghai, PR China
[c] R&D Information, AstraZeneca, Shanghai, PR China

ARTICLE INFO

Article history:
Received 1 July 2015
Received in revised form 17 July 2015
Accepted 17 July 2015
Available online 22 July 2015

Keywords:
aCGH
RNASeq
HBV
HCC

ABSTRACT

Interaction between HBV and host genome integrations in hepatocellular carcinoma (HCC) development is a complex process and the mechanism is still unclear. Here we described in details the quality controls and data mining of aCGH and transcriptome sequencing data on 50 HCC samples from the Chinese patients, published by Dong et al. (2015) (GEO#: GSE65486). In additional to the HBV-*MLL4* integration discovered, we also investigated the genetic aberrations of HBV and host genes as well as their genetic interactions. We reported human genome copy number changes and frequent transcriptome variations (e.g. *TP53*, *CTNNB1* mutation, especially *MLL* family mutations) in this cohort of the patients. For HBV genotype C, we identified a novel linkage disequilibrium region covering HBV replication regulatory elements, including basal core promoter, DR1, epsilon and poly-A regions, which is associated with HBV core antigen over-expression and almost exclusive to HBV-*MLL4* integration.

Specifications	
Organism/cell line/tissue	*Homo sapiens* patients' hepatocellular carcinoma
Sex	43 male and 7 female patients
Sequencer or array type	Illumina HiSeq2000 Genome Analyzer, Agilent 244K array CGH platform
Data format	RNASeq: Raw data: .fastq.gz files, Mapped data: .bam files, Gene expression data: .txt files aCGH: Raw data: .CEL files, Normalized data: .txt files
Experimental factors	50 fresh-frozen HCC specimens and 5 matched adjacent normal liver tissues were used for RNA sequencing. The same tumor samples and 14 matched adjacent normal liver tissues were subjected to aCGH profiling.
Experimental features	Clinical pathological parameters include patients' tumor size, tumor number, grade, stage, cirrhosis status, AFP level, progression free survival and overall survival.
Consent	Prior written informed consent was obtained from each patient, and the study was conducted in accordance with the principles of the Declaration of Helsinki and approved by the institutional review board (IRB).
Sample source location	HCC specimens were acquired from 50 patients who were undergone surgical resection in Zhongshan Hospital, Shanghai, China.

* Corresponding author.
E-mail address: qsji18@yahoo.com (Q. Ji).

1. Direct link to deposited data

aCGH profiling data have been deposited to the GEO database and can be found here: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE65484.

RNASeq profiling data have been deposited to the GEO database and can be found here: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE65485.

Superseries accession GSE65486 links the CGH and RNA-Seq studies together: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE65486.

2. Experimental design, materials and methods

2.1. RNASeq data quality control

Clinical characteristics of HCC patients could be obtained in Table 1 of PLOS One paper [1]. Patients and their tumor specimen collection, processing and profiling steps were also described in the Materials and methods section in the paper [1]. However, the quality control steps for RNASeq data were not given in details previously. Hence, we showed the RNASeq data QC in Supplementary Table 1. Raw fastq data were mapped by ArrayStudio OSA (http://omicsoft.com/osa) [2]. Total mapped reads greater than 60 M were considered to pass QC. 5′ to 3′ ratios were used to assess the RNA sample quality.

http://dx.doi.org/10.1016/j.gdata.2015.07.018

2.2. Human genomic copy number variation in HCC

Gene copy number (GCN) change was profiled by Agilent 244K array. ArrayCGH log2 ratios greater than 0.2 and less than −0.3 were considered as the GCN gain and loss, respectively; and ArrayCGH log2 ratios greater than 0.8 and less than −0.8 were considered as amplification and deletion, respectively. Chromosome segments of the most GCN gain or loss frequency were provided in Supplementary Table 2. We then looked into genes within those segments, and found genes encoded for cell adhesion molecules, including *NCAM2*, *CDH8*, *CDH13* and *CADM2*, were GCN loss with frequency of 22%, 66%, 70% and 10%, respectively. Similarly, tumor suppressor genes, such as *RB1*, *MYH2*, *CDKN2A*, *PTEN* and *APC* were also lost with high frequency of 66%, 62%, 38%, 28% and 12%, respectively. In addition to the GCN loss, gene amplifications were also observed, which include *c-MYC*, *TPPP*, *CCND1*, *PIM1* and *c-MET* with frequency of 26%, 18%, 2%, 2% and 4%, respectively. Interestingly, *CCND1*/*FGF19*, *PIM1* and *c-MET* genes showed focal amplifications. Gene deletion was observed for *UGT2B17*, *CDKN2A*, *CDH8* and *RB1* with frequency of 44%, 8%, 6% and 2%, respectively.

2.3. Human transcriptome variation in HCC

Analysis of the RNASeq data on those 50 HCC samples revealed that the most frequently recurred cancer mutation genes with hotspot are *TP53* (44%, 22/50) and *CTNNB1* (18%, 9/50). Two mutation hotspots of *TP53*, R249S and 142Ins_fs*7 observed at frequency of 27% (6/22) and 9% (2/22) in the HCC patients, respectively. Besides the previously reported non-hotspot gene mutations, such as mutations of *AXIN1*, *MLL3* and *ARID1A* [3,4], novel gene mutations were also identified in the HCC samples, which included epigenetic related gene, virus infection/transcription related genes and others. Among the epigenetic and nucleoprotein gene, mutations occurred in *MLL2*, *MLL3*, *NPIP*, *LOC100132247* (*NPIP* like), *BPTF* and *NCPR2* with frequency of 14%, 14%, 12%, 22%, 10% and 14% in HCC, respectively. For the virus infection relevant gene, mutations were found in *HAVCR1*, *SON* and *CPD* with frequency of 18%, 12% and 10%, respectively. Furthermore, hotspot mutations were also identified for genes of *MLL3*, *LOC100132247*, *AHDC1*, *CPD*, *ARAP1* and *AGAP6*. All of these transcriptome mutations identified in HCC were summarized in Supplementary Table 3.

2.4. HBV genotype C variations and their linkage disequilibrium in HCC patients

In a previous paper, we reported that HBV transcripts were detected in 88% (44/50) of the HCC samples, of which 77% (34/44) exhibited HBV genotype C, and 23% (10/44) were HBV genotype B. Expression of HBV transcription showed a segmented pattern with four peaks identified in the HBx-pre core region, prior to HBx, pre-S2, and S regions. In those HBV expressed regions, we further identified HBV mutations which occurred frequently in the 33 HBV genotype C positive patients. Moreover, high frequency of mutations identified at the important HBV regulatory regions of core promoter (nucleotides 1751–1769), DR1 (nucleotides 1824–1835), epsilon (nucleotides 1847–1907) and poly A signal (nucleotides 1916–1921). They included 14.8% T1753C, 54.2% A1762T, 33.3% G1764A, 52.2% C1827A, 42.9% A1846C/T, 21.3% G1896A, 21.4% C1913A and 21.4% G1915C. Among those HBV mutations, A1753, A1762 and G1764 are located within HBV ORF X. The others, A1846, G1896, C1913 and G1915 are located within pre-core/core ORF, which all were relevant to HCC carcinogenesis [5]. We categorized HBV variations and high frequent mutant human genes in Fig. 1a.

To further understand HCC patients infected with the mutated HBV, HBV linkage disequilibrium was investigated in the 34 HCC patients who had HBV genotype C. A specific linkage disequilibrium (LD) region in HBV was identified in those carrying frequent HBV mutations (black box region in Fig. 1b). Interestingly, we observed 15-fold increases in the transcription of pre-core/core ORF in HCC patient harboring HBV mutants in LD region compared to those harboring the wild type HBV (Fig. 1b).

2.5. Exclusivity between HBV-MLL4 gene integration and HBV regulatory region mutations

To further understand potential interaction between HBV infection and human gene variations in HCC, we analyzed the exclusivity between HBV mutation and the host genes with high frequent variations based on known functionality, such as *TP53*, *CTNNB1*, *PTEN* and *MET*. As shown in Fig. 1a, epigenetic gene mutations occurred more frequent (65%) in the HCC patients infected with HBV genotype C compared to that (33%) in HCC patients without HBV infection. Interestingly, HBV mutants were exclusive to MLL4 fusion, but correlated with MLL2 and MLL3 mutation (proportion test, $p < 0.05$). Specifically, only 14% (1/7) of HBx A1762T/G1764A mutants were found in the MLL4 fusion positive HCC compared to 50% (13/26) of that identified in MLL4 wild-type HCC (Fisher exact test, $p = 0.195$, odds ratio = 0.17). Meanwhile, frequent mutation of *MLL2*, *MLL3* and *ARID1A* was observed in LD region mutants.

3. Discussions

HBV mutations play an important role in HBV infection, and subsequently HCC development [5]. HBV genome is evolving during chronic infection, balancing between virus replication and cellular stress induced by virus replication. As the result, mutations accumulated in HBV genome, either contributing to immune escape or enhancing virus transcription/replication. Our results validated most reported high risk HBV mutations [5–7], including HBx mutation, dual mutation of A1762T/G1764A, and point mutations of G1896A, C1653T, T1753V, and HBx5 (C1386G/A). Among them, A1762T/G1764A and G1896A were reported to enhance virus replication, which might contribute to severe hepatocyte damage and cirrhosis. HBV replication regulatory mutations are one of the high risk factors for HCC, supported by prospective studies [5]. We identified specific linkage disequilibrium within this region, which contains basal core promoter, pre-core and DR1 sequences. It was the first report on the specific linkage disequilibrium in HCC to our best knowledge. The mutations induced significant increase of HBV core antigen expression. Further analysis of clinical outcomes of the patients showed that the HBV mutations in this linkage disequilibrium associated with worse progression free survival in this cohort of HCC patients although the p-value is not significant (data not show).

Compared with other solid malignancies, development of targeted therapies against HCC has been lagging mainly due to lack of identified oncogenic drivers, which is complicated by chronic viral infection and inflammation. Therefore, understanding of interactions between viral infection and human genetic aberrations may facilitate identification of novel driver mutations for design new personalized strategy to treat HCC. The observed association between *MLL4* fusions and mutations in *MLL2*/*3* and *ARID1A* genes points out a new direction for further investigation of the sequence of the genetic aberrations and their pathological impacts on initiation, development and maintenance of HCC. The *TP53* and *CTNNB1* hotspot mutations and functions in HCC development were well studies previously [8]. Here we found new hotspots mutation of *TP53* 142Ins_fs*7, which is rarely reported. Therapeutics targeting mutant TP53 and CTNNB1 are now in early clinical development [9,10]. It is intriguing to test the efficacy of these drugs in HCC patients carrying *TP53* or *CTNNB1* mutations.

In summary, by aCGH and RNASeq profiling, we found both human and HBV genome and transcriptome variations in Chinese HCC. For HBV, a novel linkage disequilibrium region covering HBV replication regulatory elements was identified, including basal core promoter, *DR1*, epsilon and polyA regions, which was associated with HBV core antigen over-expression. For the human genetic lesions, 46% (N = 50) of HCC patients harbored *MLL* family mutations, including *MLL2*

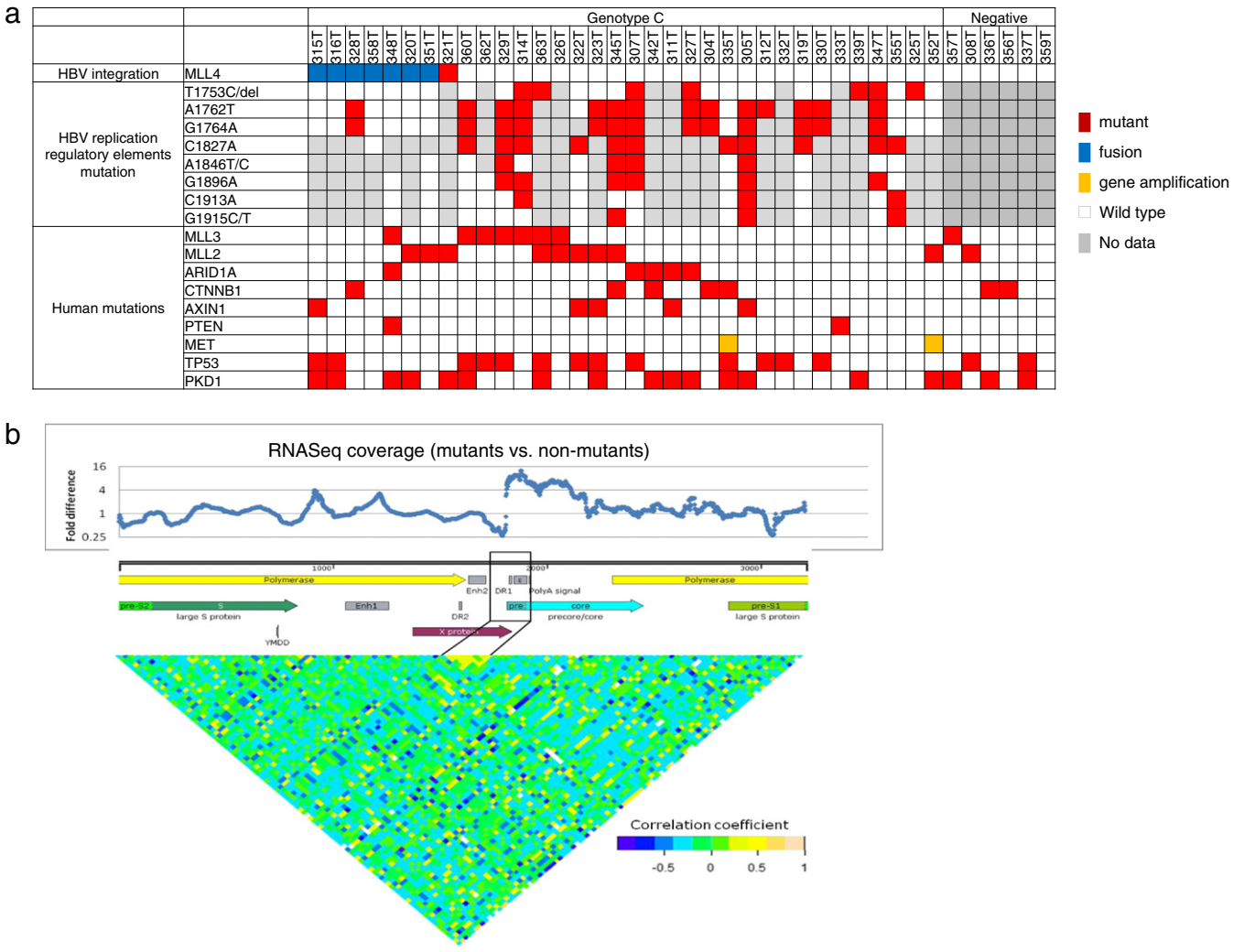

Fig. 1. (a).Overlapping view of interactions between HCC genetics lesions and HBV variants. *MLL4* fusion (blue fill), gene mutations (red fill), and *MET* amplification (orange fill) were aligned to HBV genotypes and *DR1* mutants (blank: wild type, red: mutant, gray, no reads covered). (b) Linkage disequilibrium analyses of all HBV mutations detected in current study revealed *DR1* region is functional important for HCC.

mutations, *MLL3* mutations and *HBV-MLL4* fusion. Moreover, frequent epigenetic gene variations in HCC patients, which were enriched in HBV genotype C infection subgroup suggested that *MLL4* and epigenetic mutations could be potential targets for treatment of HCC patients who are HBV type C positive.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2015.07.018.

References

[1] H. Dong, L. Zhang, Z. Qian, X. Zhu, G. Zhu, et al., Identification of HBV-MLL4 integration and its molecular basis in Chinese hepatocellular carcinoma. PLoS One 10 (2015) e0123175.
[2] J. Hu, H. Ge, M. Newman, K. Liu, OSA: a fast and accurate alignment tool for RNA-Seq. Bioinformatics 28 (2012) 1933–1934.
[3] C. Guichard, G. Amaddeo, S. Imbeaud, Y. Ladeiro, L. Pelletier, et al., Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. Nat. Genet. 44 (2012) 694–698.
[4] J. Huang, Q. Deng, Q. Wang, K.Y. Li, J.H. Dai, et al., Exome sequencing of hepatitis B virus-associated hepatocellular carcinoma. Nat. Genet. 44 (2012) 1117–1121.
[5] J.H. Lee, K.H. Han, J.M. Lee, J.H. Park, H.S. Kim, Impact of hepatitis B virus (HBV) × gene mutations on hepatocellular carcinoma development in chronic HBV infection. Clin. Vaccine Immunol. 18 (2011) 914–921.
[6] S. Liu, H. Zhang, C. Gu, J. Yin, Y. He, et al., Associations between hepatitis B virus mutations and the risk of hepatocellular carcinoma: a meta-analysis. J. Natl. Cancer Inst. 101 (2009) 1066–1082.
[7] Z.L. Fang, C.A. Sabin, B.Q. Dong, L.Y. Ge, S.C. Wei, et al., HBV A1762T, G1764A mutations are a valuable biomarker for identifying a subset of male HBsAg carriers at extremely high risk of hepatocellular carcinoma: a prospective study. Am. J. Gastroenterol. 103 (2008) 2254–2262.
[8] H.B. El-Serag, Hepatocellular carcinoma. N. Engl. J. Med. 365 (2011) 1118–1127.
[9] Z. Wang, Y. Sun, Targeting p53 for novel anticancer therapy. Transl. Oncol. 3 (2010) 1–12.
[10] A. Voronkov, S. Krauss, Wnt/beta-catenin signaling and small molecule inhibitors. Curr. Pharm. Des. 19 (2013) 634–664.