DEEP SEQUENCING ANALYSIS OF TRNSCICRIPTIONAL START SITES (Database of Transcriptional Start Sites, DBTSS: http://dbtss.hgc.jp)

Yutaka Suzuki[1], Riu Yamashita[2], Kenta Nakai[2], Sumio Sugano[1]
[1] Sch. of Frontier Sciences and [2] Human Genome Center,Institute of Medical Sci., University of Tokyo
Correspondence: ysuzuki@hgc.jp: 5-1-5 Kashiwanoha, Kashiwashi, Chiba, 277-8562, JAPAN

ABSTRACT

Although recent studies have revealed the complete structure of the transcripts for the major part of the human genes, they are mostly virtual transcript models deduced from collective information from hundreds of different cell types. Detailed figure of the transcriptome of a given particular cell type still remains elusive. We recently developed a method to enable deep sequencing analysis of the transcriptional start sites (TSSs) by combining our oligo-capping method and Illumina GA sequencing technology. Briefly, the 5'- and 3'-adaptor sequences necessary for the Illumina GA sequencing were introduced as the 5'-end-oligo at the RNA ligation and as the random hexamer primer at the first strand cDNA synthesis, respectively.

We applied this method to reveal comprehensive features of alterations in the transcriptional landscape of the human genome in different cellular conditions and developmental stages. We generated more than 300 million 36-base TSS-tag sequences and found novel stimulation-inducible alternative promoters as well as promoters that may drive non-protein-coding transcripts. All of the TSS tag sequences, their mapped genomic positions and tag counts are freely and publicly available from our database, DBTSS (http://dbtss.hgc.jp). Scalability of the TSS analysis described here would enable detailed analyses of dynamic regulations of human transcriptome in further different cell types in different conditions.

Data Contents: TSS Seq libraries

Cell lines

#TSS Cluster (TSC)	DLD-1		Ramos		BEAS2B		HEK293		MCF7		TIG3	
	TSS_tag	TSC	TSS_tag	TSC	TSS_tag	TSC	TSS_tag	TSC	TSS_tag	TSC	TSS_tag	TSC
RefSeq region	7679112	112267	1.4E+07	135388	2E+07	199200	8880335	98706	7476053	71949	8883872	69071
integenic region	957301	47347	2355840	58755	1744854	59529	736616	27959	796086	23914	783112	18450
antisense of RefSeq region	213119	17516	408592	19757	623077	23836	117363	8914	82131	6491	217908	7294

Tissues

#TSS Cluster (TSC)	brain		kidney		heart		fetal brain		fetal kidney		fetal heart	
	TSS_tag	TSC	TSS_tag	TSC	TSS_tag	TSC	TSS_tag	TSC	TSS_tag	TSC	TSS_tag	TSC
RefSeq region	7761770	297652	7285131	143101	5379867	114094	9607689	224944	6483071	166357	7982779	261567
integenic region	3472374	91915	3458501	41251	3653984	53204	1542402	67897	1592092	50950	1817819	64144
antisense of RefSeq region	602851	28972	579412	20174	397646	25522	559169	20284	487960	23870	500645	23815

TSS-Seq: Massive TSS Sequencing by "Oligo-cap ping"+ Illumina GA

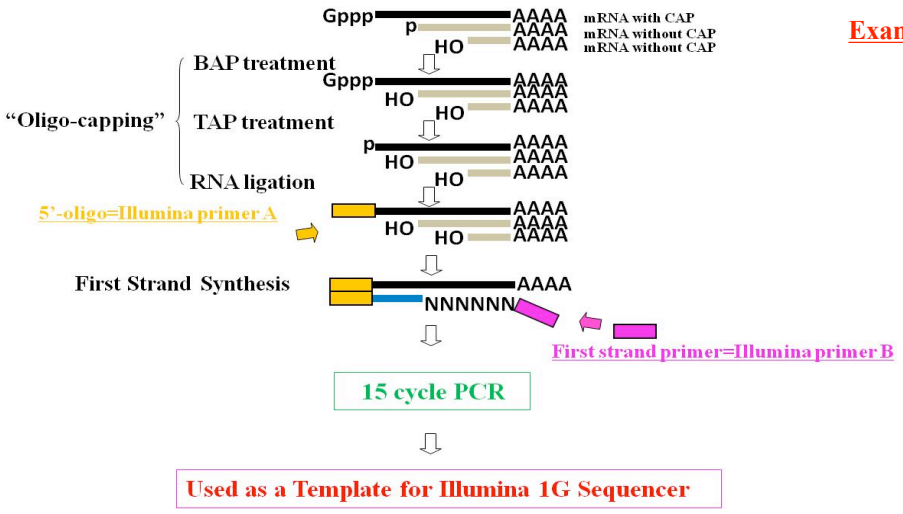


Figure 1 Scheme of 5'-end sequencing using the Illumina GA Sequencer
Adaptors containing necessary sequence for the Illumina GA sequencer are represented as yellow and pink boxes. For further information, see Supporting Online Material. Gppp: cap structure. AAA: polyA.

TSS-Seq: Cnofirmation regrading TSS positions and Expression profiles

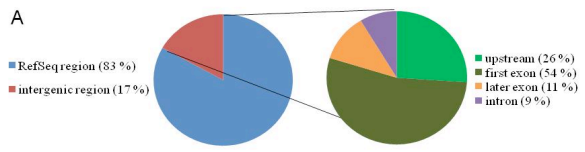


Figure 2 Validation analyses of the TSS-tag library
(A) Mapped positions of the TSS-tags relative to the RefSeq genes were evaluated. Population of the TSS-tags mapped at the corresponding positions indicated by the color bars in the margin is shown. The right circle graph shows the composition of the blue section in the left circle graph.

Database Construction (DBTSS: http://dbtss.hgc.jp)

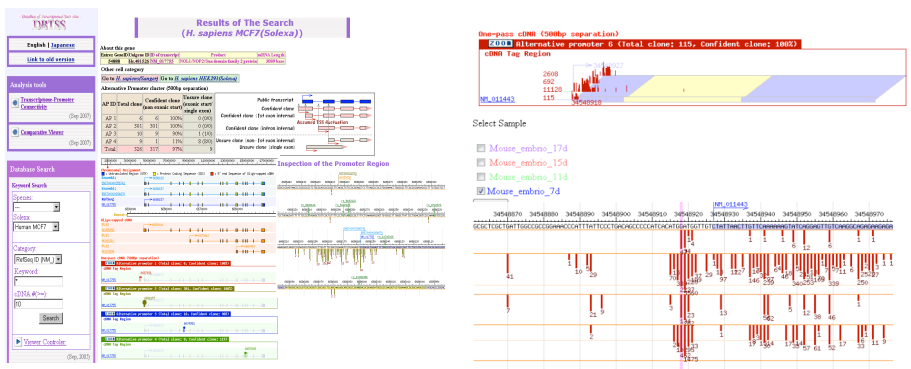
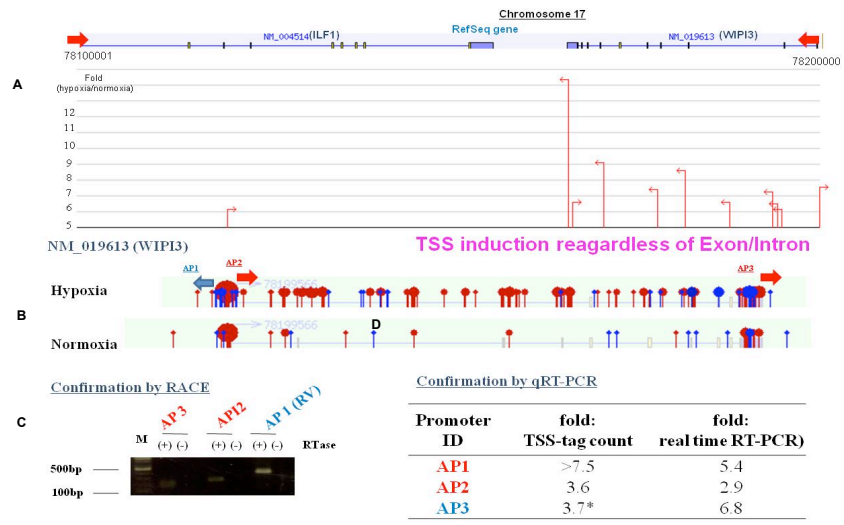


Figure 3 Update of the DataBase of Transcriptional Start Sites (DBTSS: http://dbtss.hgc.jp/)
(Left panel) Top Page and Gene viewer of the database,. (Right panel) In this upgrade, alternative promoter specific expression changes is made seachable and viewable. AP: alternative promoter.

Example of the TSS Seq Analysis (I)

Gene Expression Induction from intergenic regions (putative non-coding RNAs)



Promoter ID	fold: TSS-tag count	fold: real time RT-PCR)
AP1	>7.5	5.4
AP2	3.6	2.9
AP3	3.7*	6.8

Figure 4 Hypoxia-induced TSCs for putatively non-protein-coding RNAs
(A) Genomic positions of the regions in which activated TSSs highly concentrated Example of regions in which large numbers of transcription initiation sites were induced by hypoxia from intergenic regions (a 100 kb region in Chromosome 17). The vertical axis represents fold induction of the TSS-tag counts. (B) TSSs of the genic region of NM_019613 (WDR45-like protein gene) are shown in the bottom margin. The direction of the transcription of the RefSeq gene is represented by a red arrow. Radius of each circle represents the number of TSS-tags. Colour of each circle indicates the direction of the transcription (red: same direction with the RefSeq gene; blue: opposite direction of the RefSeq gene). Putative alternative promoters on which confirmation analysis by real-time RT-PCR is shown in (C and D) are indicated in red and blue letters (AP1-3). AP: Alternative Promoter. (C) Results of the independent oligo-cap RACE analysis for each of the Aps. (D) Real-time RT-PCR analysis of the putative alternative promoters, AP1-3, shown in (C). Fold inductions calculated by TSS-tag counts and real-time RT-PCR are shown in the third and fourth column, respectively.

Example of the TSS Seq Analysis (II)

Alernative Promoter-Specific Induction

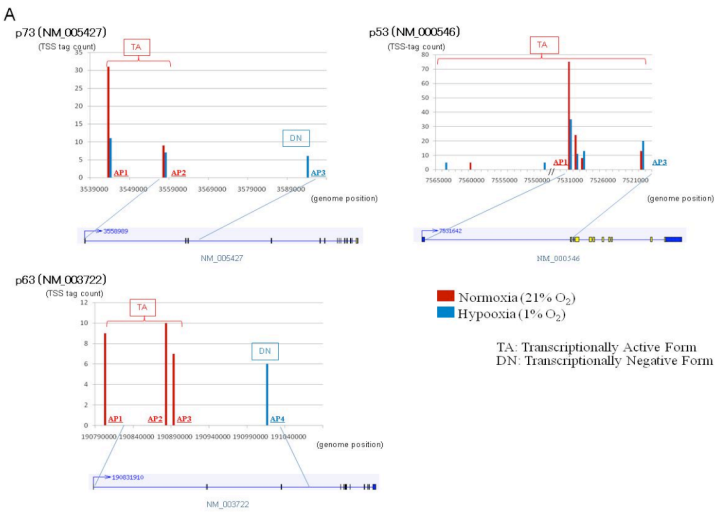


Figure 5 Hypoxia-induced TSCs in p53 family genes
(A) Count of TSS-tags mapped at the corresponding genomic regions. Red and blue solid bars represent the TSS-tag counts from normoxia and hypoxia, respectively. Exon-intron structures of the RefSeq transcripts are shown in the bottom margins. The genome regions depicted in the bar graphs are magnifications of the regions indicated by thin blue lines. Whether the transcripts from the corresponding promoters should encode the transcriptionally active (TA) or negative (DN) forms is shown in the margin. AP: alternative promoter.