

Libraries enriched for alternatively spliced exons reveal splicing patterns in melanocytes and melanomas

Akira Watahiki^{1,2}, Kazunori Waki^{1,3}, Norihito Hayatsu^{1,3}, Toshiyuki Shiraki^{1,3}, Shinji Kondo³, Mari Nakamura^{1,3}, Daisuke Sasaki³, Takahiro Arakawa³, Jun Kawai^{1,3}, Matthias Harbers⁴, Yoshihide Hayashizaki^{1-3,5} & Piero Carninci^{1,3}

It is becoming increasingly clear that alternative splicing enables the complex development and homeostasis of higher organisms. To gain a better understanding of how splicing contributes to regulatory pathways, we have developed an alternative splicing library approach for the identification of alternatively spliced exons and their flanking regions by alternative splicing sequence enriched tags sequencing. Here, we have applied our approach to mouse melan-c melanocyte and B16-F10Y melanoma cell lines, in which 5,401 genes were found to be alternatively spliced. These genes include those encoding important regulatory factors such as cyclin D2, Ilk, MAPK12, MAPK14, RAB4, melastatin 1 and previously unidentified splicing events for 436 genes. Real-time PCR further identified cell line-specific exons for *Tmc6*, *Abi1*, *Sorbs1*, *Ndel1* and *Snx16*. Thus, the ASL approach proved effective in identifying splicing events, which suggest that alternative splicing is important in melanoma development.

The results of the whole-genome sequencing projects have thus far identified far fewer genes than previously expected, and it is particularly puzzling that highly diverse organisms such as fruit fly, mouse and human seem to have rather similar numbers of genes. Therefore, the number of genes alone does not account for the increased complexity of higher organisms, and other regulatory principles must contribute to the use of genetic information during development and homeostasis^{1,2}. This phenomenon is attributed in part to the mechanisms of alternative splicing, which allow for combinatorial assembly of different exons from the same primary transcript. It was estimated that ~38–59% of the genes in mouse and human are subject to alternative splicing^{2–6} and that, furthermore, ~15% of the genetic disorders in humans are caused by mutations related to defects in pre-mRNA splicing^{7–9}.

Although the identification of all splice variants is important for understanding the transcriptome, only a few large-scale analyses have been made by computational means for exon prediction

within genomes or using expressed-sequence tags (ESTs) and cDNA sequences. Predicted exons were used to design exon- and intron¹⁰-specific oligonucleotide arrays to monitor pre-mRNA splicing on a genome-wide scale¹¹, including experiments using exon-junction arrays¹² or predicted exons on chromosome 22 (ref. 1). But these approaches are limited to predicted or known exons and do not allow new exon discovery.

To identify genes regulated at the level of alternative splicing by experimental means, we developed an approach to selectively clone differentially spliced exons from distinct biological samples into alternative splicing libraries (ASLs) and to sequence alternative splicing sequence enriched tags (ASSETs). As ASSETs encompass exons along with their related splice junctions, they not only allow exon and gene identification but also enable the analysis of different splicing types, which are beyond the reach of present computational predictions.

As cancer-related genes such as *Mdm2* (ref. 13) or *Cdkn2a*¹⁴ are regulated or affected at the level of alternative splicing in tumors, alteration of splicing might be a general mechanism contributing to tumor development. For this reason, we prepared ASLs from melanoma B16-F10Y and melanocyte melan-c cell lines to studying alternatively spliced genes and their contribution to melanoma development¹⁵. ASSETs sequencing in combination with computational annotation provides here an overview of the extent of alternative splicing in the context of tumorigenesis.

RESULTS

Alternative splicing libraries

The preparation of an ASL started from two full-length cDNA libraries derived from distinct samples. To avoid cloning intronic regions, ASLs were prepared from libraries of cytoplasmic RNA, which were further enriched by full-length selection to cover all exons within the transcripts. Single-stranded DNA (ssDNA) was prepared from the cDNA inserts of each library to obtain sets of

¹Genome Science Laboratory, RIKEN, Wako main campus, 2-1 Hirosawa, Wako, Saitama, 351-0198 Japan. ²Institute of Basic Medical Sciences, University of Tsukuba, 1-1-1 Tennodai Tsukuba, Ibaraki 305-8575, Japan. ³Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan. ⁴K.K. Dnaform, Tsukuba Branch, 3-1 Chuo 8-chome, Ami Machi, Inashiki Gun, Ibaraki, 300-0332, Japan. ⁵Science of Biological Supramolecular Systems, Graduate School of Integrated Science, Yokohama City University, 1-7-29 Suehiro-Cho, Tsurumi-Ku, Yokohama, 230-0045, Japan. Correspondence should be addressed to M.H. (matthias.harbers@dnaforminc.com) or P.C. (rgscerg@gsc.riken.jp).

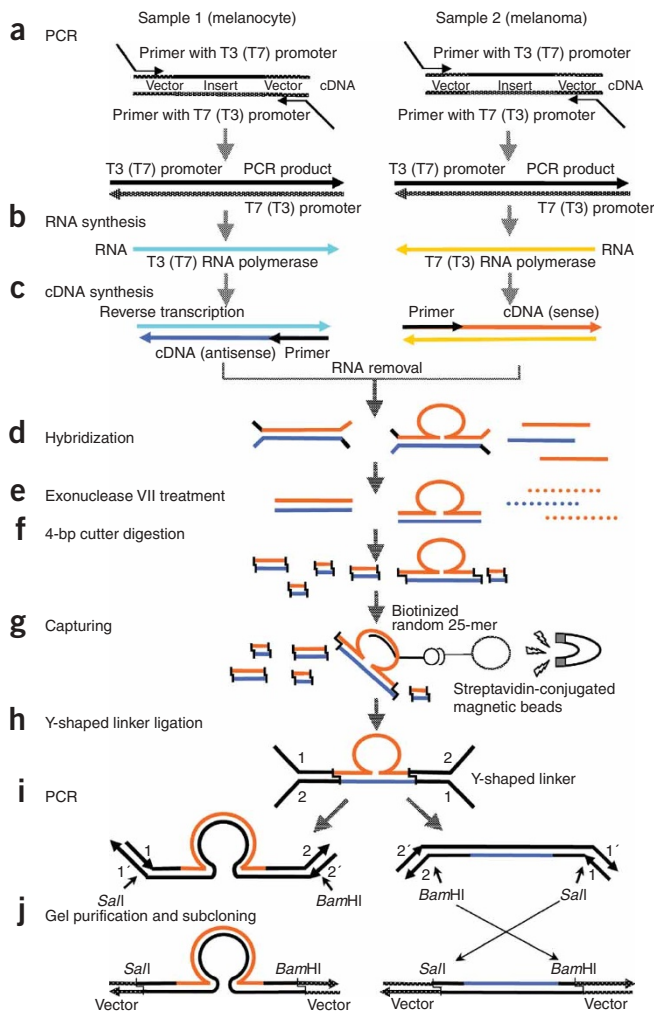


Figure 1 | Preparation of ASLs. (a–c) Preparation of ssDNA using an RNA template. (d) Hybridization of ssDNAs. (e) Removal of remaining ssDNA. (f) Digest of regions comprising dsDNA by a set of 4-bp cutters. (g) Capturing of DNA hybrids with loop structures. (h) Ligation of Y-shaped primers to isolated DNA hybrids with loop structures. (i) PCR amplification of ligation products. (j) Cloning into ASL.

sense and antisense strands (Fig. 1a–c). Complementary cDNAs formed double-stranded hybrids, which comprised sense and antisense strands derived from the two distinct libraries (Fig. 1d). Hybrids contained ssDNA regions where one cDNA contained an exon not present in the complementary cDNA. Thus, exons that were differentially spliced between the two samples formed loop structures within regions of double-stranded DNA (dsDNA) that were used for selective cloning. Unhybridized ssDNA was removed by treatment with exonuclease VII (Fig. 1e), whereas a mixture of 4-base pair (bp) cutters cleaved dsDNA into fragments mainly of 100–500 bp (Fig. 1f). The remaining DNA hybrids with single-stranded loop structures were specifically enriched by annealing to randomized oligonucleotides and captured by magnetic beads (Fig. 1g). The recovered DNA fragments were ligated to Y-shaped linkers (Fig. 1h), amplified by PCR (Fig. 1i) and cloned into ASLs (Fig. 1j). ASSETs obtained from ASLs represented individual exons derived from the loop

Table 1 | Three ASLs (L1, L2 and L3) were prepared from two parental libraries (G2: melanocyte and G3: melanoma)

Library	Conditions	
	CoT	ssDNA/dsDNA
L1	20	dsDNA
L2	1	ssDNA
L3	20	ssDNA

Library	Tissue	Cell line
G2	Melanocyte	melan-c
G3	Melanoma	B16-F10Y

CoT; Starting concentration \times time; ssDNA; single-strand DNA; ds; double-strand DNA.

structures and their related splice junctions, as the 4-bp cutters cut only within the flanking exons.

ASL preparations from melanocyte and melanoma cell lines

To test the concept of an ASL, cDNA libraries prepared from melanocyte melan-c (RIKEN ID: G2) and melanoma B16-F10Y (RIKEN ID: G3) cell lines were analyzed for differentially spliced genes, which may contribute to tumor development. Three different ASLs were prepared from the parental libraries using two different protocols (Table 1). Library L1 used PCR-derived dsDNA, which allowed for the detection of all differentially spliced transcripts within both samples. Libraries L2 and L3, prepared at two distinct hybridization conditions based on strand-specific ssDNA preparations derived from intermediate RNA templates only allowed for the detection of exons that were differentially spliced between the samples. From all ASLs and the parental libraries, individual clones were isolated and subjected to sequencing, and a total of 43,649 ASSETs and 33,602 ESTs were obtained. On average, the exons identified by ASSETs sequencing were 100–200 bp long, whereas the flanking regions covered 10–100 bp. Usually one-pass sequencing was sufficient to cover entire ASSETs where the smallest exon was only 15 bp long.

ASSETs and EST-related genes were identified and analyzed for their library distribution (Fig. 2). Libraries L2 and L3 had more genes in common with each other than with library L1, indicating that the use of dsDNA or ssDNA could drive ASL specificity, whereas no indications for preferable hybridization conditions were found. ASLs allowed identification of an additional 1,575 genes not found by ESTs, including 81 corresponding to rare transcripts (r values > 8 correspond to probabilities of $> 98\%$; ref. 16 and data not shown). Thus, sequencing of ASLs enabled a more thorough coverage of the transcriptome than was possible by EST sequencing alone. Moreover, as compared with the transcript set in Ensembl and GenBank, new splicing events of the cassette type were identified within 436 genes, demonstrating the further potential of ASLs in exon discovery (Supplementary Fig. 1 online).

Enrichment of alternatively spliced exons in an ASL

To confirm that ASLs were enriched for alternatively spliced exons, we prepared one ASL from the same parental libraries, with and without enrichment of loop structures (compare Fig. 1g), and subjected the resulting libraries to colony hybridization (Fig. 3a). Six different hybridization probes (alternative splicing, AS+) were selected from alternatively spliced exons of *Zdhhc16* (also called

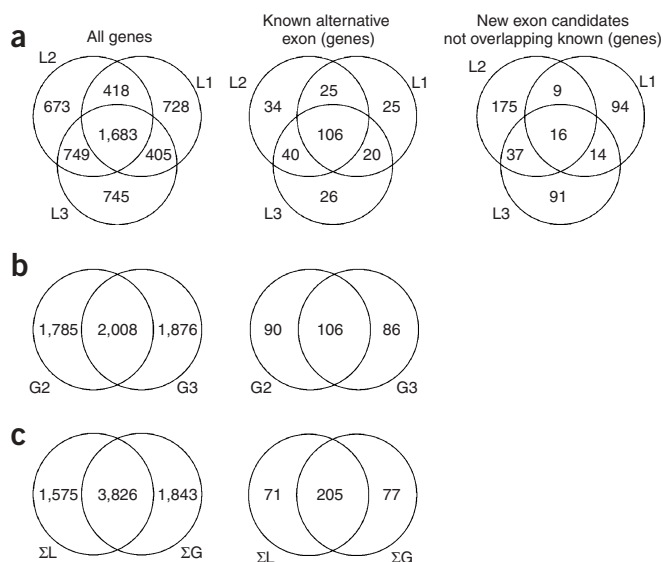


Figure 2 | Library distribution of genes identified by EST and ASSETs sequencing (Table 1). All ESTs obtained from the parental libraries and ASSETs obtained for ASLs were mapped to the mouse genome for identification of the related genes and known exons therein. The distribution of genes, known exons and new exons, as covered by the different libraries, is shown for comparison of ASLs (a), of parental libraries (b) and between ASLs and parental libraries (c).

Aph2; gi:9957214; rtsID:TB14871) and *Sorbs1* (gi:667934; rtsID:TB5104) as verified by RT-PCR. For control experiments, hybridization probes (AS⁻) were designed from invariable 3' untranslated regions of *Zdhhc16* and *Sorbs1*, for which no ASSETs had been found. AS⁺ probes recognized at least six times more colonies in the ASL library preparation subjected to enrichment of loop structures compared with the ASL library preparation omitting this step, whereas no signals were obtained with the control probe AS⁻ in the ASL library preparation subjected to enrichment of loop structures compared with some colonies detected in the ASL library preparation omitting the selection step. Statistical analysis of colony numbers from two independent experiments showed ~40-fold enrichment within ASLs by the selection of loop structures (Fig. 3b). Thus, finding the 43,649 ASSETs obtained in

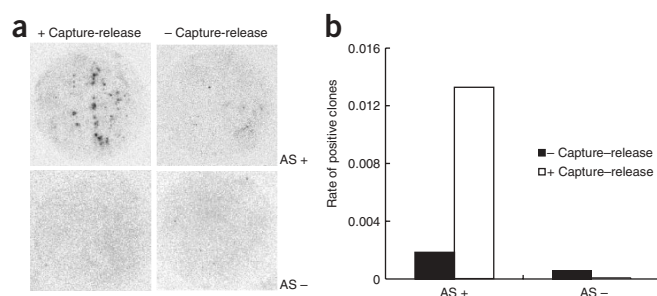


Figure 3 | Efficiency of selection step. ASLs were prepared with and without enrichment of dsDNA hybrids with loop structures and analyzed for enrichment of differentially spliced exons. (a) Images from colony hybridizations with ³²P-labeled probes comprising alternatively spliced exons (AS⁺) and invariable regions (AS⁻). (b) Ratios between positive signals obtained during colony hybridization versus total number of plated colonies.

this study would hypothetically require scanning some 1.8 million exons by EST sequencing.

Functional annotation of ASSETs and ESTs

In total, 43,649 ASSETs were obtained from three ASLs and analyzed along with 33,602 ESTs from the parental libraries to identify spliced transcripts in melanocytes and melanoma. ASSETs and ESTs were annotated by BLAST-like alignment tool (BLAT) searches in the Transcript Set within Ensembl (mouse 20.32b.1) and the mouse genome (NCBI build 32), where 29,891 ASSETs mapped, with at least 70% overlap, to unique locations in the mouse genome (Supplementary Fig. 2 online), and 37,796 mapped to the Transcript Set within Ensembl (Fig. 4 and Supplementary Table 1 online), identifying 5,401 genes, which were further annotated by Gene Ontology terms (Supplementary Table 2 online). Information on all ASSETs and ESTs is summarized on our website (see URL in Methods) and in Supplementary Table 3 online, providing gene, transcript and peptide identification numbers and Gene Ontology identification numbers and terms.

Classification of splicing types

To identify splicing types detectable by ASLs, all ASSETs were aligned against each other to identify pairs having at least one matching region to one or more ASSETs (alignments can be found on our website). For this analysis, all ASSETs were treated equally as the characterization of splice types is unrelated to the use of dsDNA or ssDNA in an ASL preparation. ASSETs groups were defined by grouping ASSETs with overlapping regions and mapping them to the same region in the genome, for which independent ASSETs confirmed exon borders. The multiple alignment program¹⁷ was used for visual inspection of all ASSETs groups. This program identified 662 alternative splicing events, which were defined as high-quality targets and were further annotated for their splice type (Supplementary Table 4 online). Among these targets eight different situations could be distinguished (Fig. 5). About 31% of the exons were of the cassette type and ~11% most likely were retained introns (defined by intronic

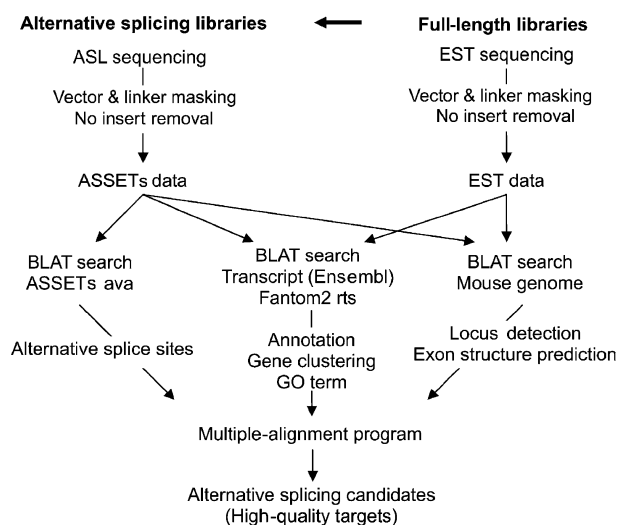


Figure 4 | Computational analysis of ASSETs. Principle of data analysis of ASSETs. Flowchart of the data analysis. ava, all versus all; GO, gene ontology.

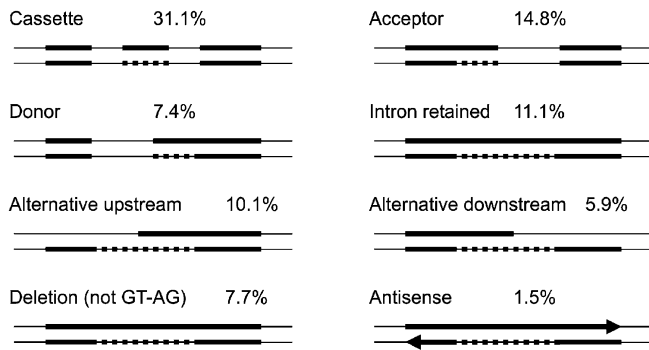


Figure 5 | Alternative splicing patterns. High-quality targets were analyzed for their splicing types. Schematic representation of the splicing types found in ASSETs.

GT-AG rule). Only 7.7% of the exons detected by ASSETs (deletion type) did not satisfy the GT-AG splicing consensus. In ten cases, sense and antisense pairs were detected, where a complementary region was spliced out from only one of the two strands. Also, the location of ASSETs within transcripts was analyzed (**Supplementary Table 5** online), and 79% of the exons were located within coding regions.

Differential expression of exons identified by ASSETs

To confirm that ASSETs comprised specific exons, 96 high-quality targets were selected at random, and exon-specific primers were designed for RT-PCR (**Supplementary Fig. 3** online). For exons of the cassette type, primers were chosen from upstream and downstream exons, whereas for other types one primer was selected from a neighboring exon and the second primer from an invariable part of the exon in question. Out of 96 RT-PCR reactions, 70 detected two different splicing patterns, whereas for 24 reactions only one splice variant was observed, and for two samples no amplification was found. RT-PCR experiments confirmed alternative splicing in melanoma and melanocytes for important genes such as dickkopf homolog 3 (*Dkk3*), tuberous sclerosis 2 (*Tsc2*), Btg3-associated nuclear protein (*Banp*) and Hermansky-Pudlak syndrome 1 (*Hps1*). Although little is known about the mouse *Dkk3*, products of the dickkopf gene family function as morphogens for the WNT pathway¹⁸. The *Tsc2* gene product is involved in tumor suppressor activity and is one of two genes associated with the tuberous sclerosis complex, where its GTPase-activating protein domain targets the small GTPase Rheb (Ras homologue enriched in brain)¹⁹. For *Banp*, at least two alternative splicing variants have been identified, one of which was associated with decreasing tumor growth²⁰. *Hps1* is involved in Hermansky-Pudlak syndrome (HPS or pale ear), which is an autosomal recessive disorder causing

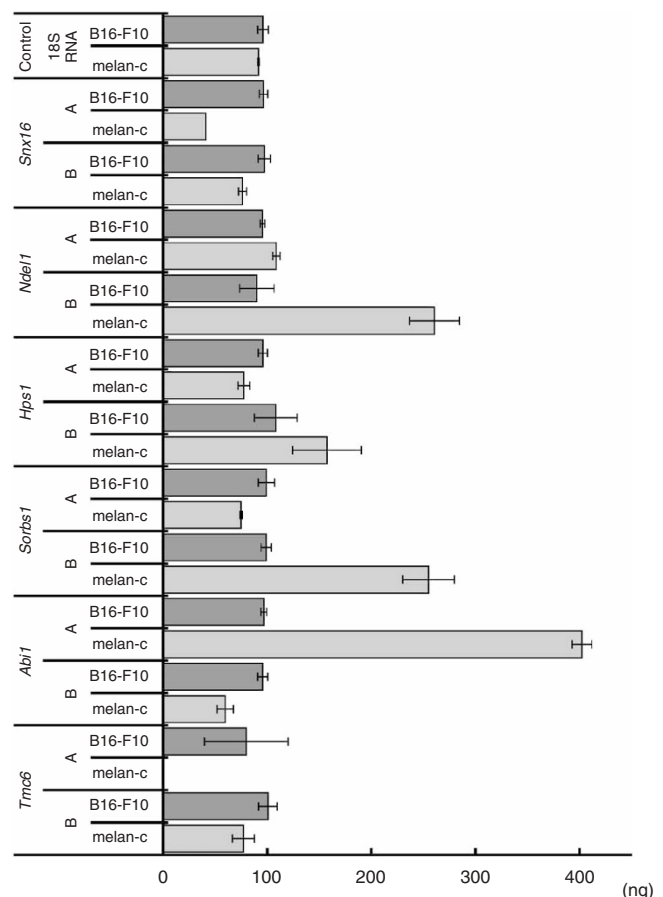
Figure 6 | Real-time PCR on selected ASSETs. Real-time PCR was done for six mRNAs, where for each mRNA a constitutive exon (A) and an alternatively spliced exon (B) were analyzed in melan-c and B16-F10Y mRNA preparations. The mean amounts \pm s.d. of RNA from three independent experiments were plotted along with 18S rRNA (control). *Snx16*: A, exons 6–7; B, exons 6– Δ 7–8. *Ndel1*: A, exons 8–9; B, exons 8–cassette–9. *Hps1*: A, exons 17–18; B, exons long17–18. *Sorbs1*: A, exons 11–12; B, exons 11–long12. *Abi1*: A, exons 2–cassette–3; B, exons 2–3. *Tmc6*: A, exons 15–cassette–16; B, exons 15–16.

oculocutaneous albinism due to defective biosynthesis or processing of melanosomes, and it has been related to trafficking of tyrosinase or tyrosinase-related protein 1 (ref. 21).

We carried out real-time PCR on six transcripts identified by ASSETs to confirm melanoma- or melanocyte-specific exons. Cytoplasmic RNA was prepared from melan-c and B16-F10 cells, and per-transcript real-time PCR was done with exon-specific primer sets for constitutive or alternatively spliced exons. Cell line-specific splicing patterns were detected for exons 15–16 within the transmembrane channel-like gene family 6 (*Tmc6*) mRNA, exons 2–3 within *Abi1* mRNA, exons 11–12 within *Sorbs1* mRNA, exons 8–9 within *Ndel1* mRNA and exons 6–8 within sorting nexin 16 (*Snx16*) mRNA (**Fig. 6**). Thus, ASSETs could identify genes specifically regulated by splicing in these cell lines, including *Tmc6*, which belongs to a larger gene family, members of which have been related to a high risk of skin carcinoma after papillomavirus infection^{22,23}; *Abi1*, which takes part in signaling and actin remodeling²⁴; *Sorbs1*, which has been related to insulin-stimulated glucose uptake and was discussed as a candidate for insulin resistance²⁵; *Ndel1*, which is involved in dynein motor activity²⁶; and *Snx16*, which belongs to a growing gene family of hydrophilic proteins, some members of which interact with various receptors, and may have a role in intracellular trafficking²⁷.

Full-length sequence comparison

After confirming that ASSETs identified specific exons, we studied their relationship to full-length cDNAs. As an example of an



extensively spliced mRNA, *Zdhhc16*, which interacts with the *c-Abl* oncogene during apoptosis²⁸, was analyzed in more detail. Nine cDNA clones identified by ESTs from the parental libraries were subjected to full-length sequencing, and different splice variants were identified by multiple alignments of full-length sequences. Also, 22 ASSETs were found, and exons were predicted by Genscan²⁹. The nine full-length cDNAs grouped into three different splice variants, which were distinct in their differential use of exons 6 and 8. The 22 ASSETs covered nine different splicing patterns, where good correlation was observed between the full-length cDNAs and ASSETs for alterations in exon 6. As outlined in **Supplementary Fig. 4** online, the number of ASSETs and full-length sequences obtained for *Zdhhc16* was not sufficient to consistently identify alterations in exons 2, 4, 8 and 12, although ASL sequencing greatly expanded the number of splice variants.

Target genes found in ASLs

Genes covered by ASSETs include those encoding key regulators of cellular control such as MAPK12 and MAPK14, the integrin-linked kinase (ILK), cyclin D2, BRPK, ADP-ribosyltransferase and melastatin 1 (**Supplementary Table 6** online). We further identified ASSETs for exon 12A of the tumor suppressor gene bridging integrator 1 (*Bin1*), whose aberrant splicing variant found in melanoma renders the protein nonfunctional³⁰. Thus, an ASL in combination with ASSETs sequencing could confirm known melanoma-related alterations in the splicing patterns of *Bin1*.

DISCUSSION

Alternative mRNA splicing is a key mechanism in higher eukaryotes for increasing proteome diversity from a limited gene repertoire^{1,2}. For a better understanding of how splice variants differ in their cellular functions, we developed ASLs and explored ASSETs derived from these libraries to locate alternatively spliced exons in a high-throughput manner using experimental means rather than computational methods.

The method of ASLs was developed to detect exons derived from internal splicing rather than from untranslated regions, as end-sequences are easily obtainable by EST sequencing³¹ or, in the case of 5' ends, by the Cap analysis gene expression (CAGE) approach³². It has been shown that ~73% of alternatively spliced exons are within CDSs², which preferentially include internal exons. The statistical analysis of the high-quality targets obtained in our study matched very well with this expectation; we found that ~79% of the ASSETs contained coding exons. Thus, ASLs allow systematic screens for alternative exon usage and its effect on coding regions. It was further important for ASL preparation that ASSETs comprise not only entire exons but also the actual splice sites, as the identification of the flanking exons greatly contributed to the analysis of splicing mechanisms and the structure of splice sites.

Another question during the development of ASLs was whether to focus on the cloning of individual exons or entire cDNAs of alternatively spliced transcripts. Here, we preferred to focus first on individual exons because their identification is sufficient for further analysis and manipulation of experimental systems. In addition to the use of realtime-PCR, individual exons can be targeted by specific antisense oligonucleotides^{33–35} or antibodies against exon-encoded protein domains^{36,37} by approaches already established for drug development and pharmacogenomics³⁸. Thus, in

combination with ASLs, exon-specific antisense oligonucleotides could be used to analyze the contribution of individual splicing events to gene function.

As an important feature of ASL preparation, we focused on the use of full-length cDNA libraries as starting materials. These libraries not only enable coverage of all internal exons within a transcript but also provide a pool for cloning individual full-length cDNAs related to ASSETs. In our initial studies on melanoma and melanocyte libraries, EST clones were prepared, but ASSETs can easily be applied to screen cDNA libraries by classical methods. Alternatively, in the future, ASLs could be prepared omitting the digestion step with 4-bp cutters to allow direct cloning and characterization of nearly full-length splice variants. When isolating full-length cDNAs, however, additional steps will be necessary to identify the alternatively spliced exons within clones, most likely including full-length sequencing.

Our characterization of thus far unknown splicing events within ASLs showed that the specific selection of loop structures comprising alternatively spliced exons was highly effective for enrichment and exon discovery. Although searching for alternatively spliced isoforms by conventional ESTs or open reading frame ESTs³⁹ are powerful means to detect alternative splicing, millions of ESTs would be required to find splicing events with a similar efficiency as in ASLs. A different approach³⁸ for cloning of alternatively spliced exons was disclosed in US patent 6,251,590, where DNA/RNA hybrids are formed and loop structures in the RNA comprising alternatively spliced exons are released by RNase H digestion of RNA within DNA/RNA hybrids. Fragments derived from the loop structures are then cloned and used for exon-specific microarray design. Apart from the lack of publicly available data, this technology cannot determine exact exon borders, as only partial exon sequences are obtained, whereas ASLs retrieve entire exons along with their flanking regions.

Although in our study of the 662 high-quality targets annotated for their splicing types ~92% followed the intronic GT-AG rule (compared with 98% previously estimated by computational means⁴⁰), ASSETs discovered a large number of thus far unknown splicing events. This observation emphasizes the need for sequencing-based methods to reliably detect splice sites. In the future, genome-wide tiling microarrays¹ may detect exon-exon junctions, although such experiments would be difficult to interpret for complex loci, because they are rather limited in their resolution and restricted to few organisms. ASSETs, however, will undoubtedly facilitate microarray design, as exon-specific or tiled microarrays rather than high-throughput sequencing could be used for evaluation of ASSETs. Similarly, ASSETs sequences could be presented on arrays to analyze splicing patterns, or they could be beneficial for other methods, such as junction probes⁴¹. Thus, we foresee comprehensive splicing studies where ASSETs could drive probe design.

ASL and ASSETs analysis was applied to melanocytes and melanomas, as in western countries the number of patients with melanoma rose faster than that with any other cancer¹⁵. Current genomics is dominated by expression profiling, despite limitations of this technique. For example, in melanomas and metastases, downregulation was observed for the tumor suppressor *Cdkn2a* and for genes involved in proliferation and growth control, such as *CDK4*, *CDK6*, *CCND1*, *RBI*, *E2F*, *ARF*, *MDM2* and *TP53* (refs. 42,43). In contrast, ASSETs derived from melanocytes and melanoma cell lines revealed a complementary pattern, excluding the above

genes at the present level of ASL sequencing. It is therefore interesting to speculate whether an increasing evolutionary pressure on regulatory networks during the development of complex organisms has preferentially expanded the number of splice variants of regulatory genes, whereas genes related to basic functions remained conserved and are preferentially regulated at the expression level. This assumption places even more emphasis on the importance of splice patterns and their contribution to regulatory processes. Here, ASSETs identified splicing events for genes of great biological relevance to melanocytes and melanoma that could be interesting targets for studies on melanoma development. The number of alternatively spliced genes covered by our ASLs supports the idea that protein variability has different patterns than previously expected from expression analysis, suggesting that the simple comparison of biological phenomena by expression profiling should be readdressed by including the detection of specific splice variants.

METHODS

Cell lines and cell culture conditions. Melanoma B16-F10 and B16-F1 cell lines were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (Gibco-BRL), 100 u/ml penicillin and 100 µg/ml streptomycin (Invitrogen) and grown in an incubator with 5% CO₂ at 37 °C. Melanocyte melan-c cells were maintained in RPMI1640 medium supplemented with 10% fetal bovine serum (Gibco-BRL) plus 100 u/ml penicillin, 100 µg/ml streptomycin (Invitrogen), 200 nM 12-0-tetradecanoyl phorbol acetate (TPA; Sigma) and 100 µM β-mercaptoethanol (Gibco-BRL) in an incubator with 10% CO₂ at 37 °C. Melan-a2 cells were maintained under the same conditions as melan-c cells, but in medium containing 200 nM TPA and 200 pM cholera toxin (Sigma) with β-mercaptoethanol omitted.

Full-length cDNA libraries. Full-length cDNA libraries were prepared as described⁴⁴.

Preparation of an ASL. A detailed protocol for the preparation of an ASL is provided in **Supplementary Methods** online.

Computational analysis. A detailed description of the computational analysis is provided in **Supplementary Figure 2** online. Sequence alignments are available from our website.

Colony hybridization, RT-PCR, real-time PCR and sequencing methods. Colony hybridization, RT-PCR, real-time PCR and sequencing methods are described in **Supplementary Methods** online, and lists of primers and probes are given in **Supplementary Tables 7–9** online. **Supplementary Fig. 5** online provides further information on the real-time PCR experiments.

URL. Our project website is available at <http://genome.gsc.riken.jp/splicing/>.

Accession numbers. ESTs from melanocytes: BB751502–BB762024, BB851120–BB851347, BY151892–BY152033 and BY465889–BY474653. ESTs from melanomas: BB762025–BB772933, BB851348–BB858427, BY152034–BY153202 and BY474654–BY481097. ASSETs: AK176923–AK220149.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank E.V. Sviderskaya for providing cells from the Wellcome Trust Functional Genomics Cell Bank, W.J. Pavan for supplying RNA and encouragement, T. Hayashi for support and encouragement, H. Nishibe for secretarial assistance and N. Harigai for technical assistance. A Research Grant for the RIKEN Genome Exploration Research Project from the Ministry of Education, Culture, Sports, Science, and Technology of the Japanese Government to Y.H. and a RIKEN Presidential Research Grant for Innovation Tool Development to P.C. supported this study.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 9 July; accepted 27 September 2004

Published online at <http://www.nature.com/naturemethods/>

- Shoemaker, D.D. *et al.* Experimental annotation of the human genome using microarray technology. *Nature* **409**, 922–927 (2001).
- Zavolan, M. *et al.* Impact of alternative initiation, splicing, and termination on the diversity of the mRNA transcripts encoded by the mouse transcriptome. *Genome Res.* **13**, 1290–1300 (2003).
- Brett, D. *et al.* EST comparison indicates 38% of human mRNAs contain possible alternative splice forms. *FEBS Lett.* **474**, 83–86 (2000).
- Croft, L. *et al.* ISIS, the intron information system, reveals the high frequency of alternative splicing in the human genome. *Nat. Genet.* **24**, 340–341 (2000).
- Liang, F. *et al.* Gene index analysis of the human genome estimates approximately 120,000 genes. *Nat. Genet.* **25**, 239–240 (2000).
- Okazaki, Y. *et al.* Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**, 563–573 (2002).
- Krawczak, M., Reiss, J. & Cooper, D.N. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum. Genet.* **90**, 41–45 (1992).
- Caceres, J.F. & Kornblihtt, A.R. Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends Genet.* **18**, 186–193 (2002).
- Faustino, N.A. & Cooper, T.A. Pre-mRNA splicing and human disease. *Genes Dev.* **17**, 419–437 (2003).
- Barras, J.D. & Beggs, J.D. Splicing goes global. *Trends Genet.* **19**, 295–298 (2003).
- Castle, J. *et al.* Optimization of oligonucleotide arrays and RNA amplification protocols for analysis of transcript structure and alternative splicing. *Genome Biol.* **4**, R66 (2003).
- Johnson, J.M. *et al.* Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302**, 2141–2144 (2003).
- Bartel, F., Taubert, H. & Harris, L.C. Alternative and aberrant splicing of MDM2 mRNA in human cancer. *Cancer Cell* **2**, 9–15 (2002).
- Loo, J.C. *et al.* Germline splicing mutations of CDKN2A predispose to melanoma. *Oncogene* **22**, 6387–6394 (2003).
- Houghton, A.N. & Polsky, D. Focus on melanoma. *Cancer Cell* **2**, 275–278 (2002).
- Stekel, D.J., Git, Y. & Falciani, F. The comparison of gene expression from multiple cDNA libraries. *Genome Res.* **10**, 2055–2061 (2000).
- Huang, X. On global sequence alignment. *Comput. Appl. Biosci.* **10**, 227–235 (1994).
- Zom, A.M. Wnt signalling: antagonistic Dickkopfs. *Cur. Biol.* **11**, R592–R595 (2001).
- Zhang, Y. *et al.* Rheb is a direct target of the tuberous sclerosis tumour suppressor proteins. *Nat. Cell Biol.* **5**, 578–581 (2003).
- Kaul, R. *et al.* Direct interaction with and activation of p53 by SMAR1 retards cell-cycle progression at G2/M phase and delays tumor growth in mice. *Int. J. Cancer* **103**, 606–615 (2003).
- Sarangarajan, R., Budev, A., Zhao, Y., Gahl, W.A. & Boissy, R.E. Abnormal translocation of tyrosinase and tyrosinase-related protein 1 in cutaneous melanocytes of Hermansky-Pudlak Syndrome and in melanoma cells transfected with anti-sense HPS1 cDNA. *J. Invest. Dermatol.* **117**, 641–646 (2001).
- Keresztes, G., Mutai, H. & Heller, S. TMC and EVER genes belong to a larger novel family, the TMC gene family encoding transmembrane proteins. *BMC Genomics* **4**, 24 (2003).
- Ramoz, N. *et al.* Mutations in two adjacent novel genes are associated with epidermodysplasia verruciformis. *Nat. Genet.* **32**, 579–581 (2002).
- Sini, P., Cannas, A., Koleske, A.J., Di Fiore, P.P. & Scita, G. Abl-dependent tyrosine phosphorylation of Sos-1 mediates growth-factor-induced Rac activation. *Nat. Cell Biol.* **6**, 268–274 (2004).
- Lin, W.H. *et al.* Cloning, mapping, and characterization of the human sorbin and SH3 domain containing 1 (SORBS1) gene: a protein associated with c-Abl

- during insulin signaling in the hepatoma cell line Hep3B. *Genomics* **74**, 12–20 (2001).
26. Liang, Y. *et al.* Nudel functions in membrane traffic mainly through association with Lis1 and cytoplasmic dynein. *J. Cell Biol.* **164**, 557–566 (2004).
 27. Haft, C.R., de la Luz Sierra, M., Barr, V.A., Haft, D.H. & Taylor, S.I. Identification of a family of sorting nexin molecules and characterization of their association with receptors. *Mol. Cell Biol.* **18**, 7278–7287 (1998).
 28. Li, B., Cong, F., Tan, C.P., Wang, S.X. & Goff, S.P. Aph2, a protein with a zf-DHHC motif, interacts with c-Abl and has pro-apoptotic activity. *J. Biol. Chem.* **277**, 28870–28876 (2002).
 29. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94 (1997).
 30. Ge, K. *et al.* Mechanism for elimination of a tumor suppressor: aberrant splicing of a brain-specific exon causes loss of function of Bin1 in melanoma. *Proc. Natl. Acad. Sci. USA* **96**, 9689–9694 (1999).
 31. Carninci, P. *et al.* Targeting a complex transcriptome: the construction of the mouse full-length cDNA encyclopedia. *Genome Res.* **13**, 1273–1289 (2003).
 32. Shiraki, T. *et al.* Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc. Natl. Acad. Sci. USA* **100**, 15776–15781 (2003).
 33. Sazani, P. & Kole, R. Therapeutic potential of antisense oligonucleotides as modulators of alternative splicing. *J. Clin. Invest.* **112**, 481–486 (2003).
 34. Vacek, M., Sazani, P. & Kole, R. Antisense-mediated redirection of mRNA splicing. *Cell. Mol. Life Sci.* **60**, 825–833 (2003).
 35. Mercatante, D.R. & Kole, R. Control of alternative splicing by antisense oligonucleotides as a potential chemotherapy: effects on gene expression. *Biochim. Biophys. Acta* **1587**, 126–132 (2002).
 36. Le Fur, N., Kelsall, S.R., Silvers, W.K. & Mintz, B. Selective increase in specific alternative splice variants of tyrosinase in murine melanomas: a projected basis for immunotherapy. *Proc. Natl. Acad. Sci. USA* **94**, 5332–5337 (1997).
 37. Le Fur, N., Silvers, W.K., Kelsall, S.R. & Mintz, B. Up-regulation of specific tyrosinase mRNAs in mouse melanomas with the c2j gene substituted for the wild-type tyrosinase allele: utilization in design of syngeneic immunotherapy models. *Proc. Natl. Acad. Sci. USA* **94**, 7561–7565 (1997).
 38. Bracco, L. & Kearsy, J. The relevance of alternative RNA splicing to pharmacogenomics. *Trends Biotechnol.* **21**, 346–353 (2003).
 39. Camargo, A.A. *et al.* The contribution of 700,000 ORF sequence tags to the definition of the human transcriptome. *Proc. Natl. Acad. Sci. USA* **98**, 12103–12108 (2001).
 40. Buset, M., Seledtsov, I.A. & Solovyev, V.V. SpliceDB: database of canonical and non-canonical mammalian splice sites. *Nucleic Acids Res.* **29**, 255–259 (2001).
 41. Wang, H. *et al.* Gene structure-based splice variant deconvolution using a microarray platform. *Bioinformatics* **19** Suppl 1, I315–I322 (2003).
 42. Carr, K.M., Bittner, M. & Trent, J.M. Gene-expression profiling in human cutaneous melanoma. *Oncogene* **22**, 3076–3080 (2003).
 43. Chin, L. The genetics of malignant melanoma: lessons from mouse and man. *Nat. Rev. Cancer* **3**, 559–570 (2003).
 44. Bowtell, D. & Sambrook, J. (eds.). *DNA Microarrays: A Molecular Cloning Manual* (Cold Spring Harbor Laboratory Press, New York, 2003).