

## Supplementary Tables and Figures Legends

**Supplementary Table 1.** Ovarian tumor tissue sources. Cancerous ovaries (n = 21) were obtained from patients with an average age of 59 years and a median of 60 years. Normal ovaries (n = 25) were obtained from patients with an average age of 60 years and a median of 57 years. The patients from whom cancerous ovaries and normal ovaries were collected showed comparable mean ages (analyzed by ANOVA with a level of significance of 5 %). n. d. not determined.

**Supplementary Table 2.** Properties of ovarian cancer-specific alternative splicing (AS) variants. Epithelial ovarian cancer (EOC) associated splice events are presented in alphabetical order. First columns show gene symbol and gene name. Column 3 shows the ASE type and direction of the shift between normal and cancer tissue. Positive numbers indicate a shift to the larger isoform in cancer, negative numbers indicate a shift toward the shorter isoform in cancer. Column 4 describes the location and effect of the splicing event with respect to the reading frame. Column 5 gives an indication of the protein's function. Column 6 indicates whether the EOC-associated splicing pattern is likely cancer-specific (CS) or epithelium-specific (ES). Unavailable data are indicated as not determined (n. d.). Column 7 indicates conservation of the splice event between man and mouse. Columns 8 and 9 indicate the mean  $\Psi$  values in normal (n=25) and cancer tissue (n=21) respectively. The final column indicates the PCR primers used to detect the AS event.

Klinck et al., *Cancer Res* 68(3) 657-63, 2008.

**Supplementary Figure 1.** Layered and integrated system for splicing annotation (LISA).

**(A)** Overview of the LISA annotation process. Transcript information for each selected gene is uploaded into the LISA database from AceView. The system automatically designs PCR primers and PCR reactions to cover all putative exon-exon junctions. RNA is reverse transcribed in bulk, using a mixture of random hexamers and oligo-d(T) nucleotides. The experiment design is sent to an automated platform that performs the PCR reactions in 384 well plates and subsequently separates and quantifies the resulting amplicons by microcapillary electrophoresis. Digitized experimental data is merged with the transcript input for analysis. **(B)** A transcript map for each input gene is generated by the LISA from publicly available databases. The map generated for the STIM1 gene is shown as an example. Each variant transcript from AceView is shown on a separate line and named (left) as per the AceView convention. Exons are shown as pale blue boxes, and the intervening introns are shown as thin black lines (not to scale). The displayed intron and exon sizes change automatically to allow clear illustration of the primer locations and whether exons from different variants overlap or not. The relative locations of forward (green) and reverse (red) primers are shown as vertical lines. The designed PCR reactions are shown below each targeted transcript (black horizontal lines). Other relevant information, such as coding regions (dashed horizontal lines) or protein functional domains (Pfam (Sonnhammer, EL, Eddy, SR, Durbin, *Proteins* 1997;28(3):405-20) pink horizontal lines) are mapped onto this representation and can be accessed for subsequent data analysis. **(C)** Capillary electropherograms of the PCR reaction spanning AS event 1 of the STIM1 gene shown in (B) in a normal ovary and an epithelial ovarian cancer (EOC) pool. Each reaction is compared to internal markers at 15 ( $M_{15}$ ) and 7000 ( $M_{7000}$ ) base pairs. Amplicon sizes and concentrations relative to the markers are measured and digitized. The fluorescence

Klinck et al., Cancer Res 68(3) 657-63, 2008.

signals of the short and long isoforms are indicated. **(D)** Graphical representation of the results obtained from the PCR reactions targeting the STIM1 AS event 1 in 4 RNA sources. Each row represents data from a single RNA source, and each column represents a single PCR reaction spanning an AS event. In this case, 4 independent reactions covered the same AS event. Electropherograms were analyzed for the presence of expected amplicon sizes. The most intense amplicon signal for each reaction in all sources was identified. The ratio of this amplicon relative to the total expected amplicon concentration was calculated and expressed as the percentage splicing index,  $\Psi$ . This representation allows rapid visual inspection of the tissue-specific AS profile of any gene of interest.

**Supplementary Figure 2.** Alternative splicing annotations generated by the LISA. **(A)** Exon sizes (black rectangles) are proportional to the square root of their lengths; introns (white rectangles) are not to scale. Each intron is uniquely labelled, and transcript names are retained from AceView. All putative exon-exon junctions are automatically assigned (uppercase letters in the intervening introns). By comparing all available transcripts, the alternative splicing (AS) events are identified, classified by type and listed beneath the transcript representation (labeled by roman numerals). Each AS event or combination of closely clustered events is represented as a red box flanked by its upstream and downstream exon boundaries. **(B)** LISA-generated summary of RNA source-specific detection of exon-exon junctions for the SHMT1 gene. The exon-exon junction analysis was performed for each of 2 ovarian cancer tissue pools and 2 normal tissue pools. Columns represent the data for each exon-exon junction defined in (A), and each row represents a different RNA source. Junctions are classified as *detected* (green), *not detected* (red) or, when data is ambiguous, *not determined* (white). The

Klinck et al., Cancer Res 68(3) 657-63, 2008.

confidence level of each assignment is represented by the intensity of the blue background. Highest confidence is displayed as dark blue and lowest confidence as white. Data from all RNA sources were subjected to unsupervised clustering and displayed with a dendrogram (left) representing the similarities between the RNA sources used. The confidence level was derived from an heuristic scoring system based on the number of reactions covering the exon-exon junction. In the case of SHMT1, 2 predicted exon junctions, labelled N and X, out of 28 were *not detected* in pools of RNA obtained from normal and ovarian serous tissues of 16 individuals with a high degree of confidence (all primer sets in agreement). On the other hand, 3 exon-exon junctions were found to be tissue type-dependent with moderate confidence level (columns A, K and AB). Three exon-exon junctions were not determined in one RNA source (OVN Pool 3 in I, L and M) and one junction was not detected with high confidence in any RNA source tested (column B). **(C)** RNA source-based display of AS events. The data generated in (B) was further analyzed to assess the state of each detected AS event. AS events defined in (A), (columns), are classified as yielding a long form, a short form or both for each RNA source (rows). The presence (green) or absence (red) of the long form is indicated in the left semicircle for each event in each RNA source, likewise for the short form's state which is indicated in the right semicircle. Data from each RNA source is clustered and presented with its corresponding dendrogram (left) to emphasize RNA source similarities.

**Supplementary Figure 3.** Individual splicing distribution graphs for 64 ASEs tested in 25 normal and 21 cancer tissues. The 48 cancer markers are shown above 16 candidates that failed the validation.  $\Psi$  value distributions are shown as a histogram for

Klinck et al., Cancer Res 68(3) 657-63, 2008.

normal (blue) and cancerous tissues (red). Derived distribution curves are superimposed. Uncorrected P values for the distributions are also indicated.

**Supplementary Figure 4.** Quantitative PCR for a sample of 11 validated genes in 2 ovarian normal and 2 serous tumor pools. Relative quantitation (RQ), calculated using the  $\Delta\Delta C_t$  method (Pfaffl, M.W., Nucleic Acids Res 2001;29(9):e45), relative to Normal Ovary Pool 1 which was normalized to a value of 1.