Modeling Spatial Correlation of Transcripts With Application to Developing Pancreas

Ruishan Liu, Marco Mignardi, Robert Jones, Martin Enge, Seung K Kim, Stephen R Quake, James Zou

Supplementary Information
Supplementary Notes

1 In Situ RNA Sequencing Data of Pancreas

To demonstrate the use of our computational analysis tool, we study in situ RNA sequencing of human fetal pancreatic tissue. The study was performed on tissues from de-identified donors with informed consent, and was approved by the ethics committee of the Stanford University Institutional Review Board (IRB). All methods were performed in accordance with the relevant guidelines and regulations. The data are collected at three developmental ages — 80, 87 and 117 days post-fertilization.

Experimental procedure. Anonymized fetal pancreas tissues were OCT-embedded and immediately frozen in liquid nitrogen upon arrival and stored at -80°C. Before tissue staining, blocks were cut in 10 µm-thick sections which were collected on superfrost plus slides and stored at 80°C.

All the solution used for tissue staining were prepared using DEPC-treated or RNase-free reagents. Slides were let air dry before fixation in 4% (w/v) paraformaldehyde in PBS for 45 min at room temperature. After fixation sections were washed in PBS for 5 min at room temperature and permeabilized with 0.01% (w/v) pepsin in 0.1 M HCl for 5 min at 37°C followed by wash in PBS (5 min room temperature) and a wash in EtOH 70, 85 and 100% for one min each at room temperature.

Slides were let air dry before mounting a silicon hybridization chamber (Secure-seal, SIGMA) of different size depending on the size of the tissue section. Tissues were washed once with PBST (1x PBS, 0.05% tween-20) at room temperature before incubation in 1x RT buffer (ThermoFisher) until reverse transcription.

In situ reverse transcription was carried out adding all the cDNA primers together (Supplementary Table 7) at a final concentration of 0.3 µM each, 1x RT buffer (ThermoFisher), 0.2 µg/µl BSA (NEB), 0.5 mM dNTPs (ThermoFisher), 1 U/µl RiboLock RNase Inhibitor (ThermoFisher) and 10 U/µl Revert AID H Minus Reverse Transcriptase (ThermoFisher) and incubating the tissue at 37°C overnight. After reverse transcription, slides were washed twice in PBST and post-fixation was done with 4% (w/v) paraformaldehyde in PBS for 10 min at room temperature followed by two washes in PBST 5 min at room temperature and incubation in 1x Ampligase buffer.

Ligation was carried out adding all the 5-end phosphorylated padlock probes together (Supplementary Table 8) at a final concentration of 0.05 µM each, 1x Ampligase buffer (Epicentre), 0.2 µg/µl BSA (NEB), 50 mM KCl, 20% Formamide, 0.4 U/µl RNase H (ThermoFisher) and 0.75 U/µl Ampligase DNA ligase (Epicentre) and incubating the tissue 30 min at 37°C followed by 60 min at 45°C. After ligation slides were washed twice in PBST and incubated in 1x Phi29 DNA Polymerase buffer. Rolling circle amplification (RCA) was carried out adding 5% glycerol, 0.2 µg/µl BSA (NEB), 0.25 mM dNTPs (ThermoFisher),
1x Phi29 DNA polymerase buffer (ThermoFisher) and 1 U/µl of Phi29 DNA polymerase (ThermoFisher) and incubating the tissue at 30°C overnight. After RCA slides were washed twice in PBST and detection mix consisting of 1x hybridization buffer (2x SSC, 20% formamide), 0.1 µM of each detection probe (Supplementary table XXX) and 10 ng/µl DAPI (ThermoFisher) was added for 30 min at 37°C.

Slides were washed twice in PBST before removing the hybridization chamber and washed in EtOH 70, 85 and 100% for 5 min each at room temperature. Stained sections were air-dried and mounted with SlowFade Antifade Gold mountant (ThermoFisher) and a coverslip before imaging.

Following staining cycles were done as follow: slides were dip in 70% EtOH until the coverslip fell off. Section were then washed in 85 and 100% EtOH 5 min each and air dried. UNG treatment was done adding 1x UNG buffer (ThermoFisher), 0.2 µg/µl BSA (NEB) and 0.05 U/µl UNG (ThermoFisher) for 30 min at 37°C. Uracil cleavage was followed by was in pre-warmed 65% formamide in water for 5 min at 55°C and two washes in PBST. Second and third staining cycles and nuclear staining were performed as described above.

**Image acquisition and analysis.** Imaging was carried out on a Zeiss Axiosplan epifluorescence microscope equipped with an Axiocam 506 mono camera (Zeiss) and filter-cubes for DAPI, FITC, Cy3 and Cy5. Each sample was imaged after every round of hybridization with a 20x/0.8 Plan-Apochromat objective (Zeiss) for a total of three rounds. Multiple fields of view were acquired with 10% overlapping and 10-15 z-stacks (about 0.5 µm step size). The resulting images were projected (maximum intensity projection, MIP) and automatically stitched using DAPI staining. MIP images were shade corrected using the automatic function in Zen Acquisition software (Zeiss) feeding a minimum intensity projection. Images were exported as 16bit grayscale images. Background subtraction was done using ImageJ software by measuring the fluorescence intensity outside the signals and removing it from the corresponding image.

Image registration was done in ImageJ as follow: images from the same round of hybridization, except for the DAPI staining, were combined (MIP). A first registration between images from different rounds of hybridization was done using the nuclei staining (DAPI images). A second registration was done using the MIPs from the other channels. Image alignment was done using MultiStackReg plugin in ImageJ using one round of hybridization (typically the first) as a reference and aligning the other rounds to the first. For every alignment a rigid transformation was applied, and a transformation matrix was saved and applied to the rest of the images.

Pre-aligned and background-corrected images were analyzed with CellProfiler 2.1.1 (rev 6c2d896) to identify nuclei, cells and fluorescent spots (RCA). The intensity and position of RCA products were measured using the same pipeline as in Mignardi et al. (2015) [1]. The barcode decoding was obtained using the same MATLAB script as described before (Ke et al., 2013) [2] and a quality threshold was applied to the detected barcodes as described above. All the
raw images, CellProfiler pipelines and oligonucleotide sequences are accessible in Supplementary Table 7, Supplementary Table 8 and Supplementary Table 9.

Basic statistics. The spatial distributions of transcripts from 25 different genes are retrieved. For samples collected at age 80, 87 and 117 days after fertilization, the 2D positions of SST, GLUC and INS are plotted in Supplementary Figure 1a, Supplementary Figure 1c and Figure 1a. The total number of reads after quality control is given in Supplementary Table 1.

The positions of nuclei are identified at the same time, illustrated in Supplementary Figure 1b, Supplementary Figure 1d and Figure 1b for samples at age 80, 87 and 117 days post fertilization. The total number of counts is recorded in Supplementary Table 1.

2 Algorithm for Identifying of Endocrine Islets

The endocrine islets are identified according to Algorithm 1. In this analysis, we set number of possible radius $m = 10$, maximum possible islet radius $r_{\text{max}} = 110 \mu m$, minimum possible islet radius $r_{\text{min}} = 10 \mu m$ which is around the averaged nuclei spacing.

Algorithm 1 Identification of Endocrine Islets

**Input:** Nuclei positions $(x_i, y_i)$, total number of SST, GLUC and INS transcripts in $i$th cell $n_i$ ($i = 1, \ldots, N$), maximum possible radius $r_{\text{max}}$, minimum possible radius $r_{\text{min}}$ and number of radius steps $m$.

Initiate islets region set $C = \emptyset$ and exocrine nuclei set $P_{\text{exo}} = \{i \mid 1 \leq i \leq N, n_i \geq 1\}$

for $t = 0$ to $m$ do

Islet radius $r = (1 - t/m)r_{\text{max}} + (t/m)r_{\text{min}}$.

Get the number of neighboring nuclei $n^{\text{neighbor}}_i$ within a radius $r$ for $i$th nucleus, $i \in P_{\text{exo}}$.

The temporary islets set $P_t = \{i \mid i \in P_{\text{exo}}, n^{\text{neighbor}}_i > (1/\#P_{\text{exo}}) \sum_{j \in P_{\text{exo}}} n^{\text{neighbor}}_j\}$

while $P_t$ is not $\emptyset$ do

$i^* = \text{arg max}_{i \in P_t} n^{\text{neighbor}}_i$, remove $i^*$ from $P_t$

circle $c^*$ has center $(x_{i^*}, y_{i^*})$ and radius $r$.

if $c^*$ do not intersect with $C$ then

Add $c^*$ to islets set $C$.

$P_{\text{exo}} = \{i \mid i \in P_{\text{exo}}, (x_i, y_i) \not\in c^*\}$

end if

end while

end for

We then identified clusters of endocrine cells for all three samples. Within
each sample we identified clusters of different size and the distribution of cluster size for each sample is plotted in Supplementary Figure 2. The total number of reads inside identified endocrine regions is given in Supplementary Table 5.

3 Density Profile-based Analysis

The density profile-based analysis is carried out to capture the relation between transcripts and other morphological features of the tissue such as nuclei’s position or developing pancreatic islets.

The density profiles are calculated based on kernel density estimation. Given the characteristic distance of each transcript \(x_i\) \((i = 0, ..., n)\), the density \(\rho\) at distance \(x\) is estimated as

\[
\rho(x) = \frac{1}{nw} \sum_{i=1}^{n} K \left( \frac{x_i - x}{w} \right),
\]

where \(n\) is the transcripts number, \(K\) is a kernel with bandwidth \(w\). For example, when analyzing the relation between gene expression and endocrine regions, the characteristic distance \(x\) could be the distance between the transcript and the boundary of its closest endocrine islet. For a good approximation at short distance, we use a 1D kernel with linear combination correction [3].

The difference between two density profiles \(\rho_1(x)\) and \(\rho_2(x)\) is characterized by symmetric Kullback-Leibler (KL) divergence

\[
D(\rho_1, \rho_2) = \frac{1}{2} \left( \int_{0}^{\infty} \rho_1(x) \log \frac{\rho_1(x)}{\rho_2(x)} dx + \int_{0}^{\infty} \rho_2(x) \log \frac{\rho_2(x)}{\rho_1(x)} dx \right),
\]

which averages the KL divergence from \(\rho_1\) to \(\rho_2\) and the divergence from \(\rho_2\) to \(\rho_1\).

4 Statistical Model for Spatial Correlations

4.1 Islet shape analysis.

In the experiment, the boundaries for endocrine islets can be nicely approximated by circles and the model characterizes the spatial correlations in the circular areas. We note that the statistical model does not rely on particular boundary shape and can be readily applied to non-circular cases.

To illustrate the model’s performance when fitted on non-circular regions, we repeat the spatial correlation analysis on pancreatic islets approximated by squared shapes. The endocrine islets are identified as squares, following Algorithm [3] with the radius of circles replaced by the side length of squares. Supplementary Figure [5] shows that the squared boundaries do not capture the shapes of islets well and tend to include more exocrine pancreas compared to circular boundaries. We again fit the statistical model and summarize the typical results in Supplementary Table [2] in correspondence with Table 2 in the...
most spatial correlations among genes are fitted to be close to 1 and three pairs of genes with positive spatial correlation are identified.

4.2 Evaluation on synthetic data

In this section, we evaluate our statistical model on synthetic datasets and demonstrate its power in characterizing the spatial distribution of the expression level among different genes.

Other methods. The computational analysis on the gene-gene spatial correlations has not been addressed in literature. Recent methods such as SpatialIDE identifies the spatial variation of individual genes on their own and the gene-gene spatial correlations are not discussed [4].

We propose two other methods as comparison — a baseline model with preliminary statistics and a pairwise Strauss process model.

In the baseline, the spatial correlation between type $i$ and type $j$ transcripts is defined as

$$\gamma_{ij, \text{baseline}} = \frac{\rho_i(j)}{\overline{\rho}(j)},$$

(3)

where $\rho_i(j)$ denotes the mean density of type $j$ transcripts in the neighborhood of type $i$ within radius $r$ and $\overline{\rho}(j)$ is the mean density of type $j$ transcripts in the whole analyzed region. The case $\gamma_{ij, \text{baseline}} > 1$ indicates that type $j$ transcripts are more likely to appear near type $i$, thus a clustering effect is expected. Similarly, $0 < \gamma_{ij, \text{baseline}} < 1$ represents an inhibition effect and $\gamma_{ij, \text{baseline}} = 1$ corresponds to an independent relation.

We further compare our multitype Strauss model with a pairwise Strauss model. The spatial correlation between type $i$ and $j$ are modeled with Strauss process for gene $i$ and $j$ only, ignoring the effects of other genes.

Synthetic datasets. We generate two toy spatial transcriptome datasets I and II with known gene-gene spatial correlations, as shown in Supplementary Figure 5. Here the interaction radius is set to be 1 unit and the analyzed region is a 20 units × 20 units square. The scale unit can be related to different length measure. Taking the pancreas transcriptomic dataset as an example, 1 unit can correspond to 20µm.

In dataset I, three genes A, B and C are correlated following Supplementary Figure 5a. There is a clustering effect between pairs (A, B) and (B, C), while the generation of gene B and C are independent. In the experiment, gene A is randomly sampled from a Poisson process on the squared region. Gene B and C are randomly generated in the neighborhood of gene A following Poisson process independently. The intensities for A, B and C are 0.1, 0.09 and 0.08, respectively. An example of generated spatial transcriptomic data are plotted in Supplementary Figure 5b.

In dataset II, two more genes D and E are added, as shown in Supplementary Figure 5c. There is a clustering effect between gene D and E, which are both
independent on the rest three genes. Genes A, B and C are generated following the same procedure as in dataset I. We sample gene D from a Poisson process on the whole region with intensity 0.1 and sample gene E from a Poisson process in the neighborhood of gene D with intensity 0.09, as illustrated in Supplementary Figure 5d.

**Performance.** Our model successfully captures the spatial correlations among genes and significantly outperforms the other two methods. The performance of fitting the baseline method, pairwise Strauss model and our proposed model on dataset I and II are given in Supplementary Table 3 and Supplementary Table 4, respectively.

Our statistical model is able to distinguish between spatial correlation and spatial co-occurrence, while the other two methods cannot. In the synthetic dataset I, gene B and C are independently generated, but they both correlate with gene A. Thus the two independent genes are also more likely to appear close to each other because of their common neighbor gene A. As shown in Supplementary Table 3, our statistical model correctly captures their independence as $\gamma_{ij} = 1.00 \pm 0.02$. However, the other two methods directly contribute the co-occurrence of B and C to their spatial correlation and yield $\gamma_{ij} > 1$ indicating a clustering effect.

When more types of transcripts are considered in synthetic dataset II, our statistical model continues to learn the correct correlations, as indicated by Supplementary Table 4. The independence between the two groups (A, B, C) and (D, E) is characterized by $\gamma_{ij} = 1$ with gene $i$ and $j$ from each group. The other two methods still mistakenly concludes that there is a clustering effect between the two independent genes B and C.

### 5 Full Results for Gene Correlation

In the experiment, the spatial correlation $\gamma_{ij}$ among genes within endocrine islets are fitted with our statistical model. The full results are summarized in Supplementary Table 6.
Supplementary Figure 1: In situ sequencing. (a)(c) Detected SST, GLUC and INS transcripts are plotted on xy coordinates for fetal pancreas sample at age (a) 80 days and (c) 87 days post fertilization. Identified pancreatic islets are identified by black circles. (b)(d) Identified and segmented nuclei are plotted on xy coordinates at age (b) 80 days and (d) 87 days post fertilization.
Supplementary Figure 2: Histogram of islets radius. The samples are collected at age (a) 80 days (b) 87 and (c) 117 days post fertilization.
Supplementary Figure 3: **Pancreatic islets approximated by squares.** The sample is from fetal pancreas at age 117 days post fertilization. Detected SST, GLUC and INS transcripts are plotted on xy coordinates. Computationally identified pancreatic islets are identified by black squares.
Supplementary Figure 4: **Synthetic datasets.** (a)(c) The illustration of gene correlations for (a) dataset I and (b) dataset II. The double-headed arrows connect two genes which have clustering effect with each other. If two genes are not directly connected, they are generated independently. (b)(d) The spatial transcriptomic data in (b) dataset I and (d) dataset II.
Supplementary Tables

Supplementary Table 1: Total number of different transcripts reads and nuclei at age 80, 87 and 117 days after fertilization.

|       | OGG1 | VEGFB | TP53 | SST    | ARX    |
|-------|------|-------|------|--------|--------|
| Day 80| 1346 | 1151  | 909  | 28053  | 4952   |
| Day 87| 773  | 587   | 779  | 26107  | 12220  |
| Day 117| 875  | 890   | 601  | 19481  | 2769   |

|       | GLUC | MUC15 | MUC6 | MUC20 | NEUROD1 |
|-------|------|-------|------|-------|----------|
| Day 80| 58203| 12    | 38018| 987   | 107      |
| Day 87| 22557| 22    | 6182 | 1328  | 160      |
| Day 117| 50267| 18    | 16306| 725   | 148      |

|       | PROM1 | INS   | VEGFC | CDKN1A | MKI67    |
|-------|-------|-------|-------|--------|----------|
| Day 80| 3994  | 2642  | 931   | 231    | 1839     |
| Day 87| 2936  | 3062  | 841   | 145    | 7565     |
| Day 117| 2140  | 3412  | 733   | 99     | 1402     |

|       | MUTYH | PDX1  | MUC16 | EPCAM  | MUC1    |
|-------|-------|-------|-------|--------|---------|
| Day 80| 125   | 452   | 60    | 5200   | 136     |
| Day 87| 221   | 1030  | 129   | 3445   | 208     |
| Day 117| 86    | 344   | 11    | 3097   | 75      |

|       | NEUROG3 | MUC13 | CDKN2A | SOD1 | TOP2A |
|-------|---------|-------|--------|------|------|
| Day 80| 104     | 103   | 334    | 2094 | 984  |
| Day 87| 147     | 31    | 204    | 1124 | 2067 |
| Day 117| 101   | 26    | 146    | 1516 | 515  |
Supplementary Table 2: When the boundaries for endocrine islets are approximated by squares, spatial correlation $\gamma_{ij}$ (mean $\pm$ std) at age 80, 87 and 117 days after fertilization.

| Correlation Intensity | Day 80     | Day 87     | Day 117    |
|-----------------------|------------|------------|------------|
| Typical               |            |            |            |
| SST $\leftrightarrow$ INS | 1.00 $\pm$ 0.02 | 1.00 $\pm$ 0.01 | 1.000 $\pm$ 0.005 |
| INS $\leftrightarrow$ MUC6 | 1.00 $\pm$ 0.04 | 1.00 $\pm$ 0.03 | 0.95 $\pm$ 0.01 |
| INS $\leftrightarrow$ ARX | 1.00 $\pm$ 0.11 | 1.00 $\pm$ 0.03 | 0.89 $\pm$ 0.04 |
| ARX $\leftrightarrow$ MUC6 | 1.00 $\pm$ 0.03 | 1.04 $\pm$ 0.02 | 1.00 $\pm$ 0.03 |
| Strongest             |            |            |            |
| EPCAM $\leftrightarrow$ PROM1 | 1.22 $\pm$ 0.08 | 1.25 $\pm$ 0.07 | 1.25 $\pm$ 0.06 |
| MUC6 $\leftrightarrow$ EPCAM | 1.16 $\pm$ 0.03 | 1.14 $\pm$ 0.06 | 1.13 $\pm$ 0.03 |
| MUC6 $\leftrightarrow$ PROM1 | 1.13 $\pm$ 0.03 | 1.13 $\pm$ 0.08 | 1.17 $\pm$ 0.04 |

Supplementary Table 3: Spatial correlation $\gamma_{ij}$ (mean $\pm$ std) for synthetic dataset I, fitted with three methods — the preliminary statistics baseline, the pairwise Strauss process model and our proposed statistical model. The true spatial correlation $\hat{\gamma}_{ij}$ equals to 1 when gene $i$ and $j$ are independent and is larger than 1 when there is a clustering effect. We randomly generate dataset I for 20 times and average the results.

| Correlation Type | Baseline     | Pairwise Model | Our Model |
|-----------------|--------------|----------------|-----------|
| Clustering ($\hat{\gamma}_{ij} > 1$) | A $\leftrightarrow$ B | 1.37 $\pm$ 0.01 | 1.30 $\pm$ 0.01 | 1.30 $\pm$ 0.02 |
|                | A $\leftrightarrow$ C | 1.38 $\pm$ 0.01 | 1.33 $\pm$ 0.01 | 1.33 $\pm$ 0.02 |
| Independent ($\hat{\gamma}_{ij} = 1$) | B $\leftrightarrow$ C | **1.21 $\pm$ 0.02** | **1.13 $\pm$ 0.02** | **1.00 $\pm$ 0.02** |
Supplementary Table 4: Spatial correlation $\gamma_{ij}$ (mean ± std) for synthetic dataset II with three methods. The true spatial correlation $\hat{\gamma}_{ij}$ equals to 1 when gene $i$ and $j$ are independent and is larger than 1 when there is a clustering effect. We randomly generate dataset II for 20 times and average the results.

| Correlation Type | Baseline | Pairwise Model | Our Model |
|------------------|----------|----------------|-----------|
| Clustering ($\hat{\gamma}_{ij} > 1$) |          |                |           |
| A ↔ B            | 1.37 ± 0.01 | 1.30 ± 0.01 | 1.31 ± 0.02 |
| A ↔ C            | 1.38 ± 0.01 | 1.33 ± 0.01 | 1.34 ± 0.02 |
| D ↔ E            | 1.38 ± 0.02 | 1.32 ± 0.02 | 1.33 ± 0.02 |
| Independent ($\hat{\gamma}_{ij} = 1$) |          |                |           |
| A ↔ D            | 1.049 ± 0.006 | 1.008 ± 0.006 | 1.008 ± 0.006 |
| A ↔ E            | 1.04 ± 0.02  | 1.00 ± 0.02  | 1.00 ± 0.01  |
| B ↔ C            | 1.21 ± 0.02  | 1.13 ± 0.02  | 1.00 ± 0.02  |
| B ↔ D            | 1.03 ± 0.01  | 1.00 ± 0.01  | 1.00 ± 0.01  |
| B ↔ E            | 1.01 ± 0.03  | 1.00 ± 0.02  | 1.00 ± 0.02  |
| C ↔ D            | 1.02 ± 0.01  | 1.00 ± 0.01  | 1.00 ± 0.01  |
| C ↔ E            | 0.99 ± 0.03  | 0.97 ± 0.02  | 1.00 ± 0.03  |

Supplementary Table 5: Total number of reads inside endocrine regions at age 80, 87 and 117 days post fertilization.

|                  | OGG1  | VEGFB | TP53 | SST | ARX | GLUC  | MUC15 | MUC6  | MUC20 | NEUROD1 |
|------------------|-------|-------|------|-----|-----|-------|-------|-------|-------|---------|
| Day 80           | 316   | 183   | 145  | 14098 | 906 |       |       |       |       |         |
| Day 87           | 142   | 121   | 130  | 13006 | 1937 |       |       |       |       |         |
| Day 117          | 387   | 262   | 146  | 11611 | 697 |       |       |       |       |         |
| PROM1            |       |       |      |      |     |       |       |       |       |         |
| Day 80           | 41274 | 2     | 6514 | 208 | 34 |       |       |       |       |         |
| Day 87           | 14448 | 4     | 1372 | 280 | 65 |       |       |       |       |         |
| Day 117          | 37171 | 4     | 4665 | 169 | 75 |       |       |       |       |         |
| MUTYH            |       |       |      |      |     |       |       |       |       |         |
| Day 80           | 31    | 123   | 19   | 1252 | 18 |       |       |       |       |         |
| Day 87           | 43    | 265   | 19   | 998  | 27 |       |       |       |       |         |
| Day 117          | 23    | 129   | 5    | 1056 | 18 |       |       |       |       |         |
| NEUROG3          |       |       |      |      |     |       |       |       |       |         |
| Day 80           | 24    | 10    | 69   | 418 | 188 |       |       |       |       |         |
| Day 87           | 36    | 7     | 38   | 208 | 349 |       |       |       |       |         |
| Day 117          | 40    | 1     | 37   | 394 | 118 |       |       |       |       |         |
Supplementary Table 6: Spatial correlation $\gamma_{ij}$ (mean ± std) at age 80, 87 and 117 days after fertilization.

|                  | Day 80      | Day 87      | Day 117     |
|------------------|-------------|-------------|-------------|
| SST ↔ INS        | 1.00 ± 0.02 | 1.00 ± 0.01 | 1.00 ± 0.01 |
| INS ↔ GLUC       | 1.000 ± 0.004 | 1.00 ± 0.02 | 1.011 ± 0.002 |
| GLUC ↔ SST       | 1.005 ± 0.001 | 1.003 ± 0.002 | 1.008 ± 0.001 |
| EPCAM ↔ PROM1    | 1.26 ± 0.08 | 1.26 ± 0.07 | 1.33 ± 0.09 |
| MUC6 ↔ EPCAM     | 1.15 ± 0.03 | 1.17 ± 0.08 | 1.12 ± 0.02 |
| MUC6 ↔ PROM1     | 1.09 ± 0.02 | 1.13 ± 0.09 | 1.19 ± 0.04 |
| SST ↔ MUC6       | 1.000 ± 0.005 | 1.00 ± 0.01 | 0.99 ± 0.01 |
| SST ↔ PROM1      | 0.95 ± 0.02 | 0.99 ± 0.01 | 1.00 ± 0.03 |
| SST ↔ EPCAM      | 1.00 ± 0.02 | 1.00 ± 0.01 | 1.02 ± 0.01 |
| GLUC ↔ MUC6      | 0.996 ± 0.002 | 1.000 ± 0.008 | 1.000 ± 0.003 |
| GLUC ↔ PROM1     | 1.005 ± 0.003 | 0.990 ± 0.007 | 1.000 ± 0.005 |
| GLUC ↔ EPCAM     | 0.98 ± 0.01 | 1.000 ± 0.004 | 1.000 ± 0.004 |
| INS ↔ MUC6       | 1.00 ± 0.04 | 1.00 ± 0.03 | 0.94 ± 0.03 |
| INS ↔ PROM1      | 1.00 ± 0.08 | 0.91 ± 0.05 | 1.12 ± 0.06 |
| INS ↔ EPCAM      | 1.2 ± 0.2 | 1.00 ± 0.04 | 1.17 ± 0.07 |
| ARX ↔ MUC6       | 1.00 ± 0.03 | 1.07 ± 0.03 | 1.00 ± 0.04 |
| ARX ↔ PROM1      | 1.2 ± 0.1 | 1.00 ± 0.04 | 1.1 ± 0.1 |
| ARX ↔ EPCAM      | 1.4 ± 0.3 | 1.00 ± 0.04 | 1.2 ± 0.1 |
| ARX ↔ SST        | 1.00 ± 0.02 | 1.000 ± 0.005 | 1.00 ± 0.01 |
| ARX ↔ GLUC       | 1.000 ± 0.004 | 1.000 ± 0.004 | 1.01 ± 0.01 |
| ARX ↔ INS        | 1.0 ± 0.2 | 1.00 ± 0.04 | 0.89 ± 0.05 |
Supplementary Table 7: cDNA primers. Here m = 2-O-Me base, + = LNA base.

| Gene     | Reference number | Probe name | Probe sequence (5′→3′)                  |
|----------|------------------|------------|----------------------------------------|
| VEGFB    | NM_001243733.1   | VEGFB<sub>1</sub> | cmugmucmugmugmamtcm gumcactgtctt      |
| VEGFB    | NM_001243733.1   | VEGFB<sub>2</sub> | gcactgaggtgaggtggcggagca               |
| VEGFB    | NM_001243733.1   | VEGFB<sub>3</sub> | cagctgggagcaggtggccatg                |
| VEGFB    | NM_001243733.1   | VEGFB<sub>4</sub> | gtgcggcagctgacacactcc                 |
| VEGFB    | NM_001243733.1   | VEGFB<sub>5</sub> | tcaccccgagctggcagacggtg               |
| VEGFC    | NM_005429.4      | VEGFC<sub>1</sub> | gmgcmtmugmamgmggactc陪同tgaat     |
| VEGFC    | NM_005429.4      | VEGFC<sub>2</sub> | gaaactcagggagggaggggg               |
| VEGFC    | NM_005429.4      | VEGFC<sub>3</sub> | ctgttagtgaccagctctccttcg            |
| VEGFC    | NM_005429.4      | VEGFC<sub>4</sub> | ctgttagctctcccgacatcgc               |
| VEGFC    | NM_005429.4      | VEGFC<sub>5</sub> | gctctcgggctgtgacatgcccc            |
| MUC1     | NM_001018016.2   | MUC1<sub>1</sub> | tmugmngmumagmagnmagnmgactgctggccac  |
| MUC1     | NM_001018016.2   | MUC1<sub>2</sub> | tgtagaggtgaggtgagagcage             |
| MUC1     | NM_001018016.2   | MUC1<sub>3</sub> | agtagtcggtggtggatctcctgtc          |
| MUC1     | NM_001018016.2   | MUC1<sub>4</sub> | ttttagttacgtaactgtgtgtc        |
| MUC1     | NM_001018016.2   | MUC1<sub>5</sub> | actgcagagcagccaaagcagtag           |
| MUC6     | NM_005961.2      | MUC6<sub>1</sub> | gmgtnnggntggmngcmctgmggtgcttgg     |
| MUC6     | NM_005961.2      | MUC6<sub>2</sub> | gacactggagggagtactggacagaggg       |
| MUC6     | NM_005961.2      | MUC6<sub>3</sub> | ttcgaggcctgtggctcctggg             |
| MUC6     | NM_005961.2      | MUC6<sub>4</sub> | gttgggacacgggcaaacggttg          |
| MUC6     | NM_005961.2      | MUC6<sub>5</sub> | ggaactgttgatggcacgctctggct       |
| MUC6     | NM_005961.2      | MUC6<sub>6</sub> | gcctgtgctgctggcctgtggtggtg     |
| MUC6     | NM_005961.2      | MUC6<sub>7</sub> | gtagctgctgtcgtgaacggtggggg       |
| MUC13    | NM_033049.3      | MUC13<sub>1</sub> | tmacngmcmagmcmagmamcmagmctggcttgg    |
| MUC13    | NM_033049.3      | MUC13<sub>2</sub> | tgggtatatctgtcagctgtaggg            |
| MUC13    | NM_033049.3      | MUC13<sub>3</sub> | ggtctccaataaagcgggtccatg          |
| MUC13    | NM_033049.3      | MUC13<sub>4</sub> | atccgatgtaaacattgtggcag           |
| MUC13    | NM_033049.3      | MUC13<sub>5</sub> | aaatattgaggtcagctgaattttgtg      |
| MUC15    | NM_001135091.1   | MUC15<sub>1</sub> | cmacngammacngmngmuatgggctgcttgct   |
| MUC15    | NM_001135091.1   | MUC15<sub>2</sub> | tttaaaaacttctgtcaatgtcctg          |
| MUC15    | NM_001135091.1   | MUC15<sub>3</sub> | ttgatgctccaagaaatgtcctg          |
| MUC15    | NM_001135091.1   | MUC15<sub>4</sub> | gttttatggggtatgttcagacaaag       |
| MUC15    | NM_001135091.1   | MUC15<sub>5</sub> | gaattattgtggttaatatttaaggtagg     |
| Gene   | Accession | Sample | Sequence                          |
|--------|-----------|--------|-----------------------------------|
| MUC16  | NM_024690.2 | MUC16_1 | anmutmccmcamgtntnamagmg getcattcttg |
| MUC16  | NM_024690.2 | MUC16_2 | catccatgacttagatgtagagatac       |
| MUC16  | NM_024690.2 | MUC16_3 | ggaaccatttgagatgtgttagtg         |
| MUC16  | NM_024690.2 | MUC16_4 | catttgcagcatcatttgtaggaattg       |
| MUC16  | NM_024690.2 | MUC16_5 | gagggagtgttgatgtgtctaatg          |
| MUC16  | NM_024690.2 | MUC16_6 | aatcenaagatcagaggagagatg          |
| MUC16  | NM_024690.2 | MUC16_7 | agtgtacagctgtgaccacaccttc         |
| MUC16  | NM_024690.2 | MUC16_8 | tagctcagagcagccagcttatttc         |
| MUC16  | NM_024690.2 | MUC16_9 | actagaactagtgaccagagagctc         |
| MUC16  | NM_024690.2 | MUC16_10| cagaactagtagctgtgaagcttac         |
| MUC16  | NM_024690.2 | MUC16_11| ccagcttttggtggctattgctag           |
| MUC16  | NM_024690.2 | MUC16_12| aggtcagttctgagcagcagactag          |
| MUC16  | NM_024690.2 | MUC16_13| ccagcttttggtggctattgctag           |
| MUC20  | NM_152673.3 | MUC20_1 | gmganamgmgagmaggmgcgmg tgggagtggag |
| MUC20  | NM_152673.3 | MUC20_2 | ttcccaggagcgtggccgcgtgc           |
| MUC20  | NM_152673.3 | MUC20_3 | tggagtggtttttctctctctcag           |
| MUC20  | NM_152673.3 | MUC20_4 | tttcctctgtgagcttagcagagc          |
| MUC20  | NM_152673.3 | MUC20_5 | cagcnaaggaagttggactggccac         |
| SOD1   | NM_000454.4 | SOD1_1 | ccmcmccmccamcmcmctcmctag          |
| SOD1   | NM_000454.4 | SOD1_2 | atagagatccttgcgtgacctg            |
| SOD1   | NM_000454.4 | SOD1_3 | ctgagagtctttctactctcc            |
| SOD1   | NM_000454.4 | SOD1_4 | caccacacagcacaagctccacagc         |
| MUTYH  | NM_001048171.1 | MUTYH_1 | tmaacmgagmtmccmcttmctggctctg gttccctgg |
| MUTYH  | NM_001048171.1 | MUTYH_2 | tggcctgactgtgcttcagatg           |
| MUTYH  | NM_001048171.1 | MUTYH_3 | tcactctctccgtcctctccat            |
| MUTYH  | NM_001048171.1 | MUTYH_4 | aatagtacccagccggacctggg            |
| MUTYH  | NM_001048171.1 | MUTYH_5 | gctgggaacaggtctgtgctggg           |
| MUTYH  | NM_001048171.1 | MUTYH_6 | gttgccagcttcagttgtggag          |
| OGG1   | NM_002542.5 | OGG1_1 | emacmcacmcmcmcmcmagmugmu gcagaatttg |
| OGG1   | NM_002542.5 | OGG1_2 | aacactctagtgaagagtacttgcgc         |
| OGG1   | NM_002542.5 | OGG1_3 | tggcaacagcctcaccatgcgcag          |
| OGG1   | NM_002542.5 | OGG1_4 | aggccagagcatcagggccttggg          |
| OGG1   | NM_002542.5 | OGG1_5 | cccgaaaaattttcctttctcctg           |
| TP53   | NM_000546.5 | TP53_1 | tmmutmtmcmgmegmgagmmsgmg tagactgacc |
| TP53   | NM_000546.5 | TP53_2 | aaagttgttttcaggaagtagtttc         |
| TP53   | NM_000546.5 | TP53_3 | gtgtgaatcaacccacactgcac            |
| TP53   | NM_000546.5 | TP53_4 | agttttttgtcatcacaatatctc          |
| Gene        | Accession | Accession | Sequence                        |
|------------|-----------|-----------|---------------------------------|
| TP53       | NM_000546.5 | TP53_5   | cctcaagctgttccgtccagtag         |
| CDKN2A     | NM_000077.4 | CDKN2A_1 | amggunmutmucmemagmangmctetctggte |
| CDKN2A     | NM_000077.4 | CDKN2A_2 | ccgggcccggcggctggccagcaa        |
| CDKN2A     | NM_000077.4 | CDKN2A_3 | gtgccccacatcatgagctggacc       |
| CDKN2A     | NM_000077.4 | CDKN2A_4 | gcaccacacgcttgccaggaacgc       |
| CDKN2A     | NM_000077.4 | CDKN2A_5 | aggttacctctgccatgegatggc        |
| CDKN1A     | NM_000389.4 | CDKN1A_1 | gmgmgamgamtagmcmagmcmgcggcgtttggag |
| CDKN1A     | NM_000389.4 | CDKN1A_2 | gcggatcagactggcggctcgactg      |
| CDKN1A     | NM_000389.4 | CDKN1A_3 | taggaggtgccccagcaaggggcccc     |
| CDKN1A     | NM_000389.4 | CDKN1A_4 | tgacgagcagcagaggtacagacg       |
| TOP2A      | NM_001067.3 | TOP2A_1  | cmattmtmummumtmtgmmgaattacatgcg |
| TOP2A      | NM_001067.3 | TOP2A_2  | ataccttttcatatttcatctacatactac |
| TOP2A      | NM_001067.3 | TOP2A_3  | ttagctgctcttttttagctgtgtgtg    |
| TOP2A      | NM_001067.3 | TOP2A_4  | gggcaacccctttctcgcctgctgc      |
| TOP2A      | NM_001067.3 | TOP2A_5  | gttgaaaaagctctcttagaatct         |
| MKI67      | NM_001145966.1 | MKI67_1 | amcmttmtmummutmummtmcmmgatgttggag |
| MKI67      | NM_001145966.1 | MKI67_2 | aacgataaatagtatttacatact        |
| MKI67      | NM_001145966.1 | MKI67_3 | gtgatttttcagcaagtgcgttc         |
| MKI67      | NM_001145966.1 | MKI67_4 | cccttatgtctcaaaaggtcttttctc    |
| MKI67      | NM_001145966.1 | MKI67_5 | gatttccatataatgcttttaagtc      |
| MKI67      | NM_001145966.1 | MKI67_6 | ttagcctgctgtttagcctgtggc       |
| MKI67      | NM_001145966.1 | MKI67_7 | ctatagtgttgcccttctcctcagg       |
| MKI67      | NM_001145966.1 | MKI67_8 | cttgtgcttttgggtgctcagggc        |
| MKI67      | NM_001145966.1 | MKI67_9 | ttccactgtggttctttttagtggg       |
| MKI67      | NM_001145966.1 | MKI67_10 | tggagcgcggcttattctctctattt     |
| EPCAM      | NM_002354.2 | EPCAM_1  | T+TA+CG+GC+CA+GC+TTGTAAGTTTTTCA |
| EPCAM      | NM_002354.2 | EPCAM_2  | ccaagtgtttgagccattctttcttg     |
| EPCAM      | NM_002354.2 | EPCAM_3  | gtttctactcgtgcagcagggta        |
| EPCAM      | NM_002354.2 | EPCAM_4  | aagaattttgaacgactaataatagtg    |
| EPCAM      | NM_002354.2 | EPCAM_5  | tcagggtgtttttttcactaataat      |
| EPCAM      | NM_002354.2 | EPCAM_6  | ttatatattgttgctgtggttctcc    |
| PROM1      | NM_001145848.1 | PROM1_1 | T+GA+TT+TG+CG+ACAAAACCATAGAAAGA |
| PROM1      | NM_001145848.1 | PROM1_2 | attttgtgaggcaattccataatttc   |
| PROM1 | NM_001145848.1 | PROM1_3 | aatccctgttcatactagtagacaatc |
|-------|----------------|----------|-------------------------------|
| PROM1 | NM_001145848.1 | PROM1_4  | gctctttagctctttgatgcctg      |
| PROM1 | NM_001145848.1 | PROM1_5  | gtgtcatattcagctgtagtaagagc   |
| PROM1 | NM_001145848.1 | PROM1_6  | tggataataacaccttttgatacc     |
| PROM1 | NM_001145848.1 | PROM1_7  | taatccaactcaacactagaggaag    |
| PROM1 | NM_001145848.1 | PROM1_8  | atttaatataagttccagaggaag     |
| PROM1 | NM_001145848.1 | PROM1_9  | cagaagaatattaagatttaccttct   |
| PROM1 | NM_001145848.1 | PROM1_10 | ttgtagagttctgagcnnatcc       |
| PROM1 | NM_001145848.1 | PROM1_11 | gtcgataatgtagttacagaggaag    |
| PROM1 | NM_001145848.1 | PROM1_12 | gatctttatgataacattatattc     |
| SST  | NM_001048.3    | SST_1    | C+AG+AC+AG+CA+GC +TC+TGCCAAGAAGTA |
| SST  | NM_001048.3    | SST_2    | acageccccaggacaggagatggac     |
| SST  | NM_001048.3    | SST_3    | cactgacagcactgagcagatc       |
| SST  | NM_001048.3    | SST_4    | cttaacaggattgaaaatctttcag    |
| ARX  | NM_139058.2    | ARX_1    | GG+CCGCG+GT+CG +ACGGGCTGTCAGG |
| ARX  | NM_139058.2    | ARX_2    | cagtaggaggagcacaagttgag      |
| ARX  | NM_139058.2    | ARX_3    | cagggcgcgcactgtggggtgcag     |
| ARX  | NM_139058.2    | ARX_4    | gtctacctgcgccgctgtgactc      |
| ARX  | NM_139058.2    | ARX_5    | cgtgcgctctccagagctctcctc     |
| ARX  | NM_139058.2    | ARX_6    | cgccgtcttcgctctagctgctc      |
| ARX  | NM_139058.2    | ARX_7    | ctctcctgctgctgagtggaggc      |
| ARX  | NM_139058.2    | ARX_8    | cggaggcggaggttagctgggagag    |
| ARX  | NM_139058.2    | ARX_9    | tttagaacacctctctgcggctgtctg  |
| PDX1 | NM_000209.3    | PDX1_1   | T+GT+TC+CT+CC +GGCTCCGCAGCCTA |
| PDX1 | NM_000209.3    | PDX1_2   | gtaggcacgcctcgcacacgg         |
| PDX1 | NM_000209.3    | PDX1_3   | tttccacctcatgagggtctttg       |
| PDX1 | NM_000209.3    | PDX1_4   | gggcgccggcgggacgcaccagag     |
| PDX1 | NM_000209.3    | PDX1_5   | gtaggcacgcctacgagggagcgag    |
| NEUROD1 | NM_002500.4 | NEUROD1_1 | amutmutmagmggmaggmacmg gaaagaacctta |
| NEUROD1 | NM_002500.4 | NEUROD1_2 | gctctgtgcctctctgtctgagaac   |
| NEUROD1 | NM_002500.4 | NEUROD1_3 | tcggtgatggttgctgccg      |
| NEUROD1 | NM_002500.4 | NEUROD1_4 | taagcccttgcaaacgtctgagac   |
| NEUROD1 | NM_002500.4 | NEUROD1_5 | gaagggctccacgctgctgcttag    |
| NEUROD1 | NM_002500.4 | NEUROD1_6 | tgatcctgcctttggccctgcc  |
| NEUROG3 | NM_020999.3 | NEUROG3_1 | gmngmnmgtnngmnacmgagagggctttaa |
| NEUROG3 | NM_020999.3 | NEUROG3_2 | cgcgtcgtctctacgggctgtctg   |
| NEUROG3 | NM_020999.3 | NEUROG3_3 | cattgattgctgcggctgctgtg   |
| NEUROG3 | NM_020999.3 | NEUROG3_4 | cgcgtcgtcgggctgcggctgctg  |
|   | Description | Accession | Description | Accession |
|---|-------------|-----------|-------------|-----------|
| NEUROG3 | NM_020999.3 | NEUROG3_5 | ccttcacagaaatctgaaage |
| INS | NM_000207.2 | INS | G+CA+CC+AG+GGC+CCC+CGCCCAGCTCCA |
| GCG | NM_002054.4 | GCG | G+TC+TC+TC+AA+AT+TC+ATCGTGACGTTT |
| Gene   | Reference number | Probe name | Probe sequence (5’ –>3’) |
|--------|------------------|------------|--------------------------|
| VEGFB  | NM_001243733.1   | VEGFB_1    | acctaaaaaaagacacT        |
|        |                  |            | GTGTCTATTTTAGGGAT        |
|        |                  |            | CCCCTAGATGTAACGCT        |
|        |                  |            | ATCGTCTTTGCTGTTAT        |
|        |                  |            | GATCGTCCagccagttg        |
|        |                  |            | gaatgcag                 |
| VEGFB  | NM_001243733.1   | VEGFB_2    | gcgcctgtgctgctggT        |
|        |                  |            | GTGTCTATTTTAGGGAT        |
|        |                  |            | CCCCTAGATGTAACGCT        |
|        |                  |            | ATCGTCTTTGCTGTTAT        |
|        |                  |            | GATCGTCCatgagctccttgctc |
| VEGFB  | NM_001243733.1   | VEGFB_3    | gactgtggagtcatggT         |
|        |                  |            | GTGTCTATTTTAGGGAT        |
|        |                  |            | CCCCTAGATGTAACGCT        |
|        |                  |            | ATCGTCTTTGCTGTTAT        |
|        |                  |            | GATCGTCCaggtggcggtgccct |
| VEGFB  | NM_001243733.1   | VEGFB_4    | ctgcagatgctggagtT         |
|        |                  |            | GTGTCTATTTTAGGGAT        |
|        |                  |            | CCCCTAGATGTAACGCT        |
|        |                  |            | ATCGTCTTTGCTGTTAT        |
|        |                  |            | GATCGTCCgtggtggctgctgcc |
| VEGFB  | NM_001243733.1   | VEGFB_5    | gacacctgcaacctggT         |
|        |                  |            | GTGTCTATTTTAGGGAT        |
|        |                  |            | CCCCTAGATGTAACGCT        |
|        |                  |            | ATCGTCTTTGCTGTTAT        |
|        |                  |            | GATCGTCCgttagagctcaacctc |
| VEGFC  | NM_005429.4      | VEGFC_1    | ttatggaattacagtC          |
|        |                  |            | CTCAATGCTGCTGCTGCTT       |
|        |                  |            | ACCCTAGATGTAACGCT        |
|        |                  |            | ATCGTCTTTAGATGGTTCC       |
|        |                  |            | GCTATTGTgctacctca         |
|        |                  |            | gcaagacg                 |
| VEGFC  | NM_005429.4      | VEGFC_2    | ctcggagaggcccgccgC        |
|        |                  |            | CTCAATGCTGCTGCTGCTT       |
|        |                  |            | ACCCTAGATGTAACGCT        |
|        |                  |            | ATCGTCTTTAGATGGTTCC       |
|        |                  |            | GCTATTGTgctgctgcct        |
|        |                  |            | cccgggct                |
| Gene   | Accession  | Gene   | Accession  | Sequence  |
|--------|------------|--------|------------|-----------|
| VEGFC  | NM_005429.4| VEGFC  | NM_005429.4| agtgtcagctaaaggaacC  
|        |            |        |            | CTCAATGCTGCTGCTGT  
|        |            |        |            | ACCCTAGATGTAACGCT  
|        |            |        |            | ATCGTCTAGATGTACC  
|        |            |        |            | GCTATTGTatattggaa  
|        |            |        |            | aatgtaca  
|        |            | VEGFC  | NM_005429.4| ttagtgttctcggatC  
|        |            |        |            | CTCAATGCTGCTGCTGT  
|        |            |        |            | ACCCTAGATGTAACGCT  
|        |            |        |            | ATCGTCTAGATGTACC  
|        |            |        |            | GCTATTGTctctggtca  
|        |            |        |            | ggaagatt  
|        |            | VEGFC  | NM_005429.4| gctgttacagacggccaC  
|        |            |        |            | CTCAATGCTGCTGCTGT  
|        |            |        |            | ACCCTAGATGTAACGCT  
|        |            |        |            | ATCGTCTAGATGTACC  
|        |            |        |            | GCTATTGTccaccacca  
|        |            |        |            | aacatgca  
|        |            | MUC1   | NM_001018016.2| gtttctgcaggtaatggA  
|        |            |        |            | GTCGGAAGTACTACTCT  
|        |            |        |            | CTTCTACGATTTTACCA  
|        |            |        |            | GTTGCCCTAGATGTACC  
|        |            |        |            | GCTATTGTgtagcccct  
|        |            |        |            | atgagaag  
|        |            | MUC1   | NM_001018016.2| ctttcttcctgctgctgA  
|        |            |        |            | GTCGGAAGTACTACTCT  
|        |            |        |            | CTTCTACGATTTTACCA  
|        |            |        |            | GTTGCCCTAGATGTACC  
|        |            |        |            | GCTATTGTaccgggcac  
|        |            |        |            | ccagctctc  
|        |            | MUC1   | NM_001018016.2| tttaattctctctggagA  
|        |            |        |            | GTCGGAAGTACTACTCT  
|        |            |        |            | CTTCTACGATTTTACCA  
|        |            |        |            | GTTGCCCTAGATGTACC  
|        |            |        |            | GCTATTGTctactgaga  
|        |            |        |            | agaatgct  
|        |            | MUC1   | NM_001018016.2| gttccacgacgtggagacA  
|        |            |        |            | GTCGGAAGTACTACTCT  
|        |            |        |            | CTTCTACGATTTTACCA  
|        |            |        |            | GTTGCCCTAGATGTACC  
|        |            |        |            | GCTATTGTgagaaggtac  
|        |            |        |            | ccaatcat  

22
| MUC1 | NM_001018016.2 | MUC1_5 |
|------|---------------|--------|
|      | gccattgtctatctcatA |       |
|      | GTCGGAAGTACTACTCT |       |
|      | CTTCTACGATTTTAACCA |       |
|      | GTTGCCCTAGATGTTCC |       |
|      | GCTATTGTgttgttgtgcttg |       |
|      | ttgcgtgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctg
| MUC6 | NM_005961.2 | MUC6_7 | gggaccagccaaaccaA GTACGCATGAATCCGTA GTAGTAGCCGTCGACTAT CTTGCTGCTCCACTGTGA CTAGATTGtgacagtga caccagt |
|------|-------------|--------|-----------------|
| MUC13 | NM_033049.3 | MUC13_1 | acatctctgtcaacagA GTCGGAAGTACTACTCT CTTCTAGATGTAACGCT ATCGTTTCTGTCTGTGGA GTTAGTCCtaattctta ctgttaagC |
| MUC13 | NM_033049.3 | MUC13_2 | gegaactagtgtctcactgA GTCGGAAGTACTACTCT CTTCTAGATGTAACGCT ATCGTTTCTGTCTGTGGA GTTAGTCCcagaagaca cagaaac |
| MUC13 | NM_033049.3 | MUC13_3 | tcagggctcctcccactggA GTCGGAAGTACTACTCT CTTCTAGATGTAACGCT ATCGTTTCTGTCTGTGGA GTTAGTCCcagaagaca atcaaatC |
| MUC13 | NM_033049.3 | MUC13_4 | actcgggtgaactgctcA GTCGGAAGTACTACTCT CTTCTAGATGTAACGCT ATCGTTTCTGTCTGTGGA GTTAGTCCattatagcct gtaaccag |
| MUC13 | NM_033049.3 | MUC13_5 | gactgtaaggcataattA GTCGGAAGTACTACTCT CTTCTAGATGTAACGCT ATCGTTTCTGTCTGTGGA GTTAGTCCggactaca gttggactc |
| MUC13 | NM_033049.3 | MUC13_6 | gacaggcagatgcacaaA GTCGGAAGTACTACTCT CTTCTAGATGTAACGCT ATCGTTTCTGTCTGTGGA GTTAGTCCggataacgg cctcaga |
| MUC15 | NM_001135091.1 | MUC15_1 | tgtgctttagataaatgA GTACGCATGAATCCGTA GTTCTACGATTTTACCA GTTGCCCTAGATGTTCG GCTATTGGTgacagaat gaacagt |
|-------|---------------|---------|--------------------|
| MUC15 | NM_001135091.1 | MUC15_2 | ataaacacaacacagaaA GTACGCATGAATCCGTA GTTCTACGATTTTACCA GTTGCCCTAGATGTTCG GCTATTGGTgaaaaagaa atcaagac |
| MUC15 | NM_001135091.1 | MUC15_3 | agtaactcatcagcagaA GTACGCATGAATCCGTA GTTCTACGATTTTACCA GTTGCCCTAGATGTTCG GCTATTGGTgaataacag atttctcc |
| MUC15 | NM_001135091.1 | MUC15_4 | ttcaccttggcttttgttA GTACGCATGAATCCGTA GTTCTACGATTTTACCA GTTGCCCTAGATGTTCG GCTATTGGTctttctt ctgcagaaac |
| MUC15 | NM_001135091.1 | MUC15_5 | actctacagctaccttA GTACGCATGAATCCGTA GTTCTACGATTTTACCA GTTGCCCTAGATGTTCG GCTATTGGTctttctt cagaaaaac |
| MUC16 | NM_024690.2 | MUC16_1 | ggtattctcacaacagaC CTCAATGCTGCTGCTGT ACCGTGCGCCTGGTAGC AATTACCTAGATGTTCG GCTATTGGTcagtgtcct tgtggatg |
| MUC16 | NM_024690.2 | MUC16_2 | tcaatgtcagtctctccC CTCAATGCTGCTGCTGT ACCGTGCGCCTGGTAGC AATTACCTAGATGTTCG GCTATTGGTggccaaag aaagtgte |
| MUC16 | NM_024690.2 | MUC16_3 | tcagaacatctacagC 
CTCAATGCTGCTGCTGT 
ACCGTGCGCCTGGTAGC 
AATTACCTAGATGTTCC 
GCTATTGTCtcactctcc 
cagtcagc |
|-------|-------------|---------|-----------------|
| MUC16 | NM_024690.2 | MUC16_4 | atgtcaccgacacctcC 
CTCAATGCTGCTGCTGT 
ACCGTGCGCCTGGTAGC 
AATTACCTAGATGTTCC 
GCTATTGTagcactgaaa 
ageccaca |
| MUC16 | NM_024690.2 | MUC16_5 | agtacctttacattagaC 
CTCAATGCTGCTGCTGT 
ACCGTGCGCCTGGTAGC 
AATTACCTAGATGTTCC 
GCTATTGTctggtgcta 
catcagag |
| MUC16 | NM_024690.2 | MUC16_6 | agtacacagcaaccteC 
CTCAATGCTGCTGCTGT 
ACCGTGCGCCTGGTAGC 
AATTACCTAGATGTTCC 
GCTATTGTcetteccag 
aaactatg |
| MUC16 | NM_024690.2 | MUC16_7 | gacatctccgtgaagtC 
CTCAATGCTGCTGCTGT 
ACCGTGCGCCTGGTAGC 
AATTACCTAGATGTTCC 
GCTATTGTagtcactacaa 
tgtcacta |
| MUC16 | NM_024690.2 | MUC16_8 | acctcagttgaataactC 
CTCAATGCTGCTGCTGT 
ACCGTGCGCCTGGTAGC 
AATTACCTAGATGTTCC 
GCTATTGTctccaante 
tgatgcc |
| MUC16 | NM_024690.2 | MUC16_9 | ctaggagtgtggaaccteC 
CTCAATGCTGCTGCTGT 
ACCGTGCGCCTGGTAGC 
AATTACCTAGATGTTCC 
GCTATTGTctgtttcace 
tgaggtta |
| Gene   | Accession   | 10th | Sequence 1 |
|--------|-------------|------|-----------|
| MUC16  | NM_024690.2 | MUC16_10 | gcaccagagatggagacC  
|        |             |       | CTCAATGCTGCTGCTGT  
|        |             |       | ACCGTGCCTGCTGAGC  
|        |             |       | AATTACCTAGATGTTCTC  
|        |             |       | GCTATTGTcaacaactct  
|        |             |       | caaccttgtt          |
| MUC16  | NM_024690.2 | MUC16_11 | gagctgagcaactgcAC  
|        |             |       | CTCAATGCTGCTGCTGCTGT  
|        |             |       | ACCGTGCCTGCTGAGC  
|        |             |       | AATTACCTAGATGTTCTC  
|        |             |       | GCTATTGTgagageggc  
|        |             |       | tgtactgg            |
| MUC16  | NM_024690.2 | MUC16_12 | aagtgcctgcctctgtaAC  
|        |             |       | CTCAATGCTGCTGCTGCTGT  
|        |             |       | ACCGTGCCTGCTGAGC  
|        |             |       | AATTACCTAGATGTTCTC  
|        |             |       | GCTATTGTceatgttca  
|        |             |       | agaacaacc            |
| MUC16  | NM_024690.2 | MUC16_13 | ttcctgcctctctgtaC  
|        |             |       | CTCAATGCTGCTGCTGCTGT  
|        |             |       | ACCGTGCCTGCTGAGC  
|        |             |       | AATTACCTAGATGTTCTC  
|        |             |       | GCTATTGTceaaacccag  
|        |             |       | ccaccaca              |
| MUC20  | NM_152673.3 | MUC20_1 | ctcacccgggaactcCA  
|        |             |       | GTACGCATGAATCCGTA  
|        |             |       | GTAGTAGCCCTGACTAT  
|        |             |       | CGTCTCTACTAGATGTTCT  
|        |             |       | GCTATTGTaaacagtga  
|        |             |       | tgtgcagcg              |
| MUC20  | NM_152673.3 | MUC20_2 | gctatgacactagacccA  
|        |             |       | GTACGCATGAATCCGTA  
|        |             |       | GTAGTAGCCCTGACTAT  
|        |             |       | CGTCTCTACTAGATGTTCT  
|        |             |       | GCTATTGTagacacag  
|        |             |       | aagtgcce               |
| MUC20  | NM_152673.3 | MUC20_3 | accgatgacagctcgaA  
|        |             |       | GTACGCATGAATCCGTA  
|        |             |       | GTAGTAGCCCTGACTAT  
|        |             |       | CGTCTCTACTAGATGTTCT  
|        |             |       | GCTATTGTagcttgaca  
|        |             |       | ecccttge               |
| Genes  | Accession  | Genes   | Accession   | Sequence  |
|--------|------------|---------|-------------|-----------|
| MUC20  | NM_152673.3| MUC20_4 |             | tccgatccaccagctctA GTAGC CATGAATCCGTA GTAGTCAGCCGTGACTAT CGTCTCCTAGATGTTCC GCTATTGTtgaagacct cgtccacc |
| MUC20  | NM_152673.3| MUC20_5 |             | gctgggtcagcagtggA GTAGC CATGAATCCGTA GTAGTCAGCCGTGACTAT CGTCTCCTAGATGTTCC GCTATTGTetccggtct ccatagag |
| SOD1   | NM_000454.4| SOD1_1  |             | agcagaaggaaatatA GTCGGAAGTACTACTCT CTA GTAGCCGTGACTAT CGTCTCCTAGATGTTCC GCTATTGTgggca tcatcaattcgc |
| SOD1   | NM_000454.4| SOD1_2  |             | cgttaccagtgcaggtcA GTCGGAAGTACTACTCT CTA GTAGCCGTGACTAT CGTCTCCTAGATGTTCC GCTATTGTggagataat acacgagtcagcag |
| SOD1   | NM_000454.4| SOD1_3  |             | tgttgtcatagtgaagattA GTCGGAAGTACTACTCT CTA GTAGCCGTGACTAT CGTCTCCTAGATGTTCC GCTATTGTaaagatggt gtgccggt |
| SOD1   | NM_000454.4| SOD1_4  |             | gacaggaaacgtggaaA GTCGGAAGTACTACTCT CTA GTAGCCGTGACTAT CGTCTCCTAGATGTTCC GCTATTGTaataagagaa agtacaaaggtg |
| MUTYH  | NM_001048171.1| MUTYH_1 |             | gttttccgtgttatcaC CTAATGCTGCTGTGCTGT ACCGTCGCGCTCGTGATGC AATTATCTGCTCTGTTGA GTTAGTCCcaccgcca tgaanaag |
| MUTYH   | NM_001048171.1 | MUTYH_2       | gggaggcagaagcatgcC  
|         |               |               | CTCAATGCTGCTGCTGT  
|         |               |               | ACCGTGCCTGCTGCTGT  
|         |               |               | AATTATCTGCTGCTGCTG  
|         |               |               | GTTAGTCCaggcagcc  
|         |               |               | gccaggaa  
| MUTYH   | NM_001048171.1 | MUTYH_3       | gggaggcctacatggagC  
|         |               |               | CTCAATGCTGCTGCTGCTGT  
|         |               |               | ACCGTGCCTGCTGCTGCTG  
|         |               |               | AATTATCTGCTGCTGCTGCTG  
|         |               |               | GTTAGTCCgttaagcc  
|         |               |               | aagagaaa  
| MUTYH   | NM_001048171.1 | MUTYH_4       | gggaggccatggagC  
|         |               |               | CTCAATGCTGCTGCTGCTG  
|         |               |               | ACCGTGCCTGCTGCTGCTG  
|         |               |               | AATTATCTGCTGCTGCTGCTG  
|         |               |               | GTTAGTCCcctgccgcc  
|         |               |               | ccctggag  
| MUTYH   | NM_001048171.1 | MUTYH_5       | gttggaggagtgtgccgC  
|         |               |               | CTCAATGCTGCTGCTGCTG  
|         |               |               | ACCGTGCCTGCTGCTGCTG  
|         |               |               | AATTATCTGCTGCTGCTGCTG  
|         |               |               | GTTAGTCCgttcctgg  
|         |               |               | tccagacc  
|         |               |               |  
| MUTYH   | NM_001048171.1 | MUTYH_6       | gttggaggagtggcttcC  
|         |               |               | CTCAATGCTGCTGCTGCTG  
|         |               |               | ACCGTGCCTGCTGCTGCTG  
|         |               |               | AATTATCTGCTGCTGCTGCTG  
|         |               |               | GTTAGTCCgtctgggccg  
|         |               |               | gtcttgac  
| OGG1    | NM_002542.5   | OGG1_1        | tccgtggaggagccaaT  
|         |               |               | GTGTCTATTTTAGTGAT  
|         |               |               | CCTCTACGATTTTACCA  
|         |               |               | GTTGGCCTCCACTGTTA  
|         |               |               | CTAGATTTGctctctgg  
|         |               |               | acaatctt  
| OGG1    | NM_002542.5   | OGG1_2        | gttggaggccgtgcaAT  
|         |               |               | GTGTCTATTTTAGTGAT  
|         |               |               | CCTCTACGATTTTACCA  
|         |               |               | GTTGGCCTCCACTGTTA  
|         |               |               | CTAGATTTGagccacac  
|         |               |               | cccagaca  

29
| Gene   | Accession   | Sample | Sequence  |
|--------|-------------|--------|-----------|
| OGG1   | NM_002542.5 | OGG1_3 | gcccgeatcaetggcaT  
|        |             |        | GTGTCTATTTAGTGGAAT  
|        |             |        | CCTCTACGATTTTACCA  
|        |             |        | GTTGCGCTCCACTGTGA  
|        |             |        | CTAGATTGtctccaaac  
|        |             |        | aaccaacat  |
| OGG1   | NM_002542.5 | OGG1_4 | atatgaggagcccaacT  
|        |             |        | GTGTCTATTTAGTGGAAT  
|        |             |        | CCTCTACGATTTTACCA  
|        |             |        | GTTGCGCTCCACTGTGA  
|        |             |        | CTAGATTGcagtaacga  
|        |             |        | gagctctc  |
| OGG1   | NM_002542.5 | OGG1_5 | cccagaccaacagaT  
|        |             |        | GTGTCTATTTAGTGGAAT  
|        |             |        | CCTCTACGATTTTACCA  
|        |             |        | GTTGCGCTCCACTGTGA  
|        |             |        | CTAGATTGggcgaaggg  
|        |             |        | acceggecc  |
| TP53   | NM_000546.5 | TP53_1 | cacctgaagtcacaT  
|        |             |        | GTGTCTATTTAGTGGAAT  
|        |             |        | CCCCTAGATGTAACGCT  
|        |             |        | ATCGTGCTCCACTGTGA  
|        |             |        | CTAGATTGgcaggtctc  
|        |             |        | actccaggc  |
| TP53   | NM_000546.5 | TP53_2 | ttcagacatggaacT  
|        |             |        | GTGTCTATTTAGTGGAAT  
|        |             |        | CCCCTAGATGTAACGCT  
|        |             |        | ATCGTGCTCCACTGTGA  
|        |             |        | CTAGATTGgtgagtcag  
|        |             |        | gaaacatt  |
| TP53   | NM_000546.5 | TP53_3 | gacctgccctgtgacT  
|        |             |        | GTGTCTATTTAGTGGAAT  
|        |             |        | CCCCTAGATGTAACGCT  
|        |             |        | ATCGTGCTCCACTGTGA  
|        |             |        | CTAGATTGttttgcaca  
|        |             |        | cttggccaa  |
| TP53   | NM_000546.5 | TP53_4 | aaatttcgctggtgagT  
|        |             |        | GTGTCTATTTAGTGGAAT  
|        |             |        | CCCCTAGATGTAACGCT  
|        |             |        | ATCGTGCTCCACTGTGA  
|        |             |        | CTAGATTGtttatccga  
|        |             |        | gtgggaagg  |
| Gene   | Accession | Exon | Sequence                        |
|--------|-----------|------|---------------------------------|
| TP53   | NM_000546.5 | 5    | ctcagttgtaatctacT GTGTCCTATTAGTGGAT CCCCTAGATGAACGCT ATCGTGCTCCACTGTTA CTAGATTGactactaca ctggaaga |
| TP53   | NM_000546.5 | 6    | ggatggagaatatteaT GTGTCCTATTAGTGGAT CCCCTAGATGAACGCT ATCGTGCTCCACTGTTA CTAGATTGcecaaaagaa aaaccact |
| CDKN2A | NM_000077.4 | 1    | acatccccgattgaaagA GTCGGAAGTACTACTCT CTAGTAGCCGGTGACTAT CGTCTGCTCCACTGTTA CTAGATTGcgcggaagg tccctcag |
| CDKN2A | NM_000077.4 | 2    | cggctgactggctggccA GTCGGAAGTACTACTCT CTAGTAGCCGGTGACTAT CGTCTGCTCCACTGTTA CTAGATTGagcagcatg gagcctt |
| CDKN2A | NM_000077.4 | 3    | ggtccggagccgatcgaA GTCGGAAGTACTACTCT CTAGTAGCCGGTGACTAT CGTCTGCTCCACTGTTA CTAGATTGacgcacgca atagttac |
| CDKN2A | NM_000077.4 | 4    | gctgcgcggagggtttA GTCGGAAGTACTACTCT CTAGTAGCCGGTGACTAT CGTCTGCTCCACTGTTA CTAGATTGcgcgacgca tcgaacgac |
| CDKN2A | NM_000077.4 | 5    | ctggagagtgggcecaA GTCGGAAGTACTACTCT CTAGTAGCCGGTGACTAT CGTCTGCTCCACTGTTA CTAGATTGtcgtgcgcg ggacgtgg |
| Gene    | Accession | Description | Sequence 1                                      | Sequence 2                                      | Sequence 3                                      | Sequence 4                                      |
|---------|-----------|-------------|------------------------------------------------|------------------------------------------------|------------------------------------------------|------------------------------------------------|
| CDKN1A  | NM_000389.4 | CDKN1A_1    | gacagatttacactCTCAATGCTGCTGCTGTACAGTAGCCCTGAGTATCGTCTCCTAGATGTTCCGCTATTGTCggeggegcat | accacgtctaccaactCTCAATGCTGCTGCTGTACAGTAGCCCTGAGTATCGTCTCCTAGATGTTCCGCTATTGTCggeggegcat | accacgtctaccaactCTCAATGCTGCTGCTGTACAGTAGCCCTGAGTATCGTCTCCTAGATGTTCCGCTATTGTCggeggegcat | accacgtctaccaactCTCAATGCTGCTGCTGTACAGTAGCCCTGAGTATCGTCTCCTAGATGTTCCGCTATTGTCggeggegcat |
| CDKN1A  | NM_000389.4 | CDKN1A_2    | tggacagcgacgacgacgACCTCAATGCTGCTGCTGTACAGTAGCCCTGAGTATCGTCTCCTAGATGTTCCGCTATTGTCcgcctcttgcatgcgccccag | accacgtctaccaactCTCAATGCTGCTGCTGTACAGTAGCCCTGAGTATCGTCTCCTAGATGTTCCGCTATTGTCcgcctcttgcatgcgccccag | accacgtctaccaactCTCAATGCTGCTGCTGTACAGTAGCCCTGAGTATCGTCTCCTAGATGTTCCGCTATTGTCcgcctcttgcatgcgccccag | accacgtctaccaactCTCAATGCTGCTGCTGTACAGTAGCCCTGAGTATCGTCTCCTAGATGTTCCGCTATTGTCcgcctcttgcatgcgccccag |
| CDKN1A  | NM_000389.4 | CDKN1A_3    | gctctgggccgtgctgcgCTCAATGCTGCTGCTGTACAGTAGCCCTGAGTATCGTCTCCTAGATGTTCCGCTATTGTCcactggaggttgacttc | accacgtctaccaactCTCAATGCTGCTGCTGTACAGTAGCCCTGAGTATCGTCTCCTAGATGTTCCGCTATTGTCcactggaggttgacttc | accacgtctaccaactCTCAATGCTGCTGCTGTACAGTAGCCCTGAGTATCGTCTCCTAGATGTTCCGCTATTGTCcactggaggttgacttc | accacgtctaccaactCTCAATGCTGCTGCTGTACAGTAGCCCTGAGTATCGTCTCCTAGATGTTCCGCTATTGTCcactggaggttgacttc |
| CDKN1A  | NM_000389.4 | CDKN1A_4    | gacgtcgtcactgtcttgCTCAATGCTGCTGCTGTACAGTAGCCCTGAGTATCGTCTCCTAGATGTTCCGCTATTGTCacagaggaagaccatgta | accacgtctaccaactCTCAATGCTGCTGCTGTACAGTAGCCCTGAGTATCGTCTCCTAGATGTTCCGCTATTGTCacagaggaagaccatgta | accacgtctaccaactCTCAATGCTGCTGCTGTACAGTAGCCCTGAGTATCGTCTCCTAGATGTTCCGCTATTGTCacagaggaagaccatgta | accacgtctaccaactCTCAATGCTGCTGCTGTACAGTAGCCCTGAGTATCGTCTCCTAGATGTTCCGCTATTGTCacagaggaagaccatgta |
| TOP2A   | NM_001067.3 | TOP2A_1     | agctttttgtcttcggctgtcTGTCGGGAGTACTACTCTCTCTGGGCGCCTGCTGTAGCATTAGCCCTTACTGTAATCTAGATTCaatgtctctaattataggtatgA | gacgttctttgttctgctcTGTCGGGAGTACTACTCTCTCTGGGCGCCTGCTGTAGCATTAGCCCTTACTGTAATCTAGATTCaatgtctctaattataggtatgA | gacgttctttgttctgctcTGTCGGGAGTACTACTCTCTCTGGGCGCCTGCTGTAGCATTAGCCCTTACTGTAATCTAGATTCaatgtctctaattataggtatgA | gacgttctttgttctgctcTGTCGGGAGTACTACTCTCTCTGGGCGCCTGCTGTAGCATTAGCCCTTACTGTAATCTAGATTCaatgtctctaattataggtatgA |
| TOP2A   | NM_001067.3 | TOP2A_2     | aattttataaattttatgATGTCGGGAGTACTACTCTCTCTGGGCGCCTGCTGTAGCATTAGCCCTTACTGTAATCTAGATTCaatgtctctaattataggtatgA | gacgttctttgttctgctcTGTCGGGAGTACTACTCTCTCTGGGCGCCTGCTGTAGCATTAGCCCTTACTGTAATCTAGATTCaatgtctctaattataggtatgA | gacgttctttgttctgctcTGTCGGGAGTACTACTCTCTCTGGGCGCCTGCTGTAGCATTAGCCCTTACTGTAATCTAGATTCaatgtctctaattataggtatgA | gacgttctttgttctgctcTGTCGGGAGTACTACTCTCTCTGGGCGCCTGCTGTAGCATTAGCCCTTACTGTAATCTAGATTCaatgtctctaattataggtatgA |
| TOP2A   | NM_001067.3 | TOP2A_3     | aaagcttttttaattggATGTCGGGAGTACTACTCTCTCTGGGCGCCTGCTGTAGCATTAGCCCTTACTGTAATCTAGATTCaatgtctctaattataggtatgA | gacgttctttgttctgctcTGTCGGGAGTACTACTCTCTCTGGGCGCCTGCTGTAGCATTAGCCCTTACTGTAATCTAGATTCaatgtctctaattataggtatgA | gacgttctttgttctgctcTGTCGGGAGTACTACTCTCTCTGGGCGCCTGCTGTAGCATTAGCCCTTACTGTAATCTAGATTCaatgtctctaattataggtatgA | gacgttctttgttctgctcTGTCGGGAGTACTACTCTCTCTGGGCGCCTGCTGTAGCATTAGCCCTTACTGTAATCTAGATTCaatgtctctaattataggtatgA |
| Gene | Accession | Segment | Sequence |
|------|-----------|---------|----------|
| TOP2A | NM_001067.3 | TOP2A_4 | tgggtgaagttaaggeA GTCGGAAGTACTACTCT CTCGTGCAGCCTGTTAGC AATTAGCTCCACTGTGTA CTAGATTGtagaaagca tactaaacr |
| TOP2A | NM_001067.3 | TOP2A_5 | aaagtttggaaccagA GTCGGAAGTACTACTCT CTCGTGCAGCCTGTTAGC AATTAGCTCCACTGTGTA CTAGATTGggaagactca aatattacr |
| TOP2A | NM_001067.3 | TOP2A_6 | ttcaaacgggaatgacaaA GTCGGAAGTACTACTCT CTCGTGCAGCCTGTTAGC AATTAGCTCCACTGTGTA CTAGATTGggttaagta ttacttgcc |
| TOP2A | NM_001067.3 | TOP2A_7 | acgggtttgatatctA GTCGGAAGTACTACTCT CTCGTGCAGCCTGTTAGC AATTAGCTCCACTGTGTA CTAGATTGgttaaggag taatgact |
| TOP2A | NM_001067.3 | TOP2A_8 | gaactagaaggcctaaA GTCGGAAGTACTACTCT CTCGTGCAGCCTGTTAGC AATTAGCTCCACTGTGTA CTAGATTGctcaagaga atggtgtg |
| MKI67 | NM_001145966.1 | MKI67_1 | atcaaggaacagccctaaA GTCGGAAGTACTACTCT CTCGTGCAGCCTGTTAGC AATTAGCTCCACTGTGTA CTAGATTGggttaagta agaaaact |
| MKI67 | NM_001145966.1 | MKI67_2 | ctaaaacatggagatgaGTCGGAAGTACTACTCT CTCGTGCAGCCTGTTAGC AATTAGCTCCACTGTGTA CTAGATTGggttaagtagctagtaaccg |

33
| MKI67 | NM_001145966.1 | MKI67_3 |
|-------|----------------|--------|
|       | gcaagatgtagtaaaactC | CTCAATGCTGCTGCTGCTGT |
|       |                 | ACCGTGCGCCTGGTAGC |
|       |                 | AATTACTTGCTGCTGTAT |
|       |                 | GATCGTCCtgttgtgcaaa |
|       |                 | aatcatgg |

| MKI67 | NM_001145966.1 | MKI67_4 |
|-------|----------------|--------|
|       | atgaaaggaatagaagC | CTCAATGCTGCTGCTGCTGT |
|       |                 | ACCGTGCGCCTGGTAGC |
|       |                 | AATTACTTGCTGCTGTAT |
|       |                 | GATCGTCCaaaggagag |
|       |                 | aaggagag |

| MKI67 | NM_001145966.1 | MKI67_5 |
|-------|----------------|--------|
|       | ggtgatgagaagacatC | CTCAATGCTGCTGCTGCTGT |
|       |                 | ACCGTGCGCCTGGTAGC |
|       |                 | AATTACTTGCTGCTGTAT |
|       |                 | GATCGTCCcgccecaac |
|       |                 | cagcagga |

| MKI67 | NM_001145966.1 | MKI67_6 |
|-------|----------------|--------|
|       | cccacaaaaaccacagC | CTCAATGCTGCTGCTGCTGT |
|       |                 | ACCGTGCGCCTGGTAGC |
|       |                 | AATTACTTGCTGCTGTAT |
|       |                 | GATCGTCCaatetac |
|       |                 | aaccagac |

| MKI67 | NM_001145966.1 | MKI67_7 |
|-------|----------------|--------|
|       | tcgccaaagacactaaC | CTCAATGCTGCTGCTGCTGT |
|       |                 | ACCGTGCGCCTGGTAGC |
|       |                 | AATTACTTGCTGCTGTAT |
|       |                 | GATCGTCCaactggga |
|       |                 | tggagaggg |

| MKI67 | NM_001145966.1 | MKI67_8 |
|-------|----------------|--------|
|       | tcgccagggcaacactC | CTCAATGCTGCTGCTGCTGT |
|       |                 | ACCGTGCGCCTGGTAGC |
|       |                 | AATTACTTGCTGCTGTAT |
|       |                 | GATCGTCCtcagggaaac |
|       |                 | gaacacca |

| MKI67 | NM_001145966.1 | MKI67_9 |
|-------|----------------|--------|
|       | tttaacagcgctccaaC | CTCAATGCTGCTGCTGCTGT |
|       |                 | ACCGTGCGCCTGGTAGC |
|       |                 | AATTACTTGCTGCTGTAT |
|       |                 | GATCGTCCatggtgctg |
| Gene   | Reference ID | Symbol | Sequence 1 | Sequence 2 |
|--------|--------------|--------|------------|------------|
| MKI67  | NM_001145966.1 | MKI67_10 | tcggctcctgaaataaAC | CTCAATGCTGCTGCTGCTGACCTGACGCTGATAATCTTTGCTGCTGATGATCGTCACACAAAAATcacaagac |
| EPCAM  | NM_002354.2   | EPCAM_1 | AAATGTGTCGTGAAAACATGTCGGAAAGTACTACTCCTCTCACTAGATTATACCAGTTGCGCTCCACTGTAGTACTGATTGTTGCCAGCA | GCTCAGGAAG |
| EPCAM  | NM_002354.2   | EPCAM_2 | ggtagaagggagaattAGTCGGAAGTACTACTCTCTTCCTACAGTATTTACCACTTTGCGCTCCACTGTAGTACTGATTGTTGCCAGCA |aatgttttaaatgttatctacttc |
| EPCAM  | NM_002354.2   | EPCAM_3 | actgaataaactgtaeeAGTCGGAAGTACTACTCTCTTCCTACAGATTTTACCACTTTGCGCTCCACTGTAGTACTGATTGTTGCCAGCA |acaaggacactgtaaatgttatcactaatcg |
| EPCAM  | NM_002354.2   | EPCAM_4 | gaaaaattgtattgttgtAGTCGGAAGTACTACTCTCTTCCTACAGATTTTACCACTTTGCGCTCCACTGTAGTACTGATTGTTGCCAGCA |tattgtatatgtatatgtaaaaatggatatgat |
| EPCAM  | NM_002354.2   | EPCAM_5 | tttaattttataagtgAAGTCGGAAGTACTACTCTCTTCCTACAGATTTTACCACTTTGCGCTCCACTGTAGTACTGATTGTTGCCAGCA |ggtcacaacagtggatcctgttgtaaatgag |
| EPCAM  | NM_002354.2   | EPCAM_6 | gtagcagaggaactcaAGTCGGAAGTACTACTCTCTTCCTACAGATTTTACCACTTTGCGCTCCACTGTAGTACTGATTGTTGCCAGCA |atgagtgatcctggttatgtaaatgag |

35
| PROM1 | NM_001145848.1 | PROM1_1 | CATTGGCATCTTCTATGG CCTCAATGCTGCTGCTG TACTCTACGATTACCC AGTTGCGCTCCACTGTT ACTAGATTTGACTTGTA TAATAATAAG |
|-------|----------------|---------|-----------------------------|-------------------------------------------------------------|
| PROM1 | NM_001145848.1 | PROM1_2 | tcctaaggcttggaattC CTCAATGCTGCTGCTG GTACTCTACGATTTTACCA GTTGCGCTCCACTGTGA CTAAGATTGccttcatec acagagt | |
| PROM1 | NM_001145848.1 | PROM1_3 | tggattatgacaagattgC CTCAATGCTGCTGCTG GTACTCTACGATTTTACCA GTTGCGCTCCACTGTGA CTAAGATTGcttcatggac cc | |
| PROM1 | NM_001145848.1 | PROM1_4 | gtccatggcaacagcgaC CTCAATGCTGCTGCTG GTACTCTACGATTTTACCA GTTGCGCTCCACTGTGA CTAAGATTGgcatatgaa actccaaat | |
| PROM1 | NM_001145848.1 | PROM1_5 | cagcatcagattgtctcC CTCAATGCTGCTGCTG GTACTCTACGATTTTACCA GTTGCGCTCCACTGTGA CTAAGATTGtcaagtgaa acctgcaaa | |
| PROM1 | NM_001145848.1 | PROM1_6 | tgtctgtcaggttctcC CTCAATGCTGCTGCTG GTACTCTACGATTTTACCA GTTGCGCTCCACTGTGA CTAAGATTGcagcttgttacctgcaaa accagac | |
| PROM1 | NM_001145848.1 | PROM1_7 | caacggagggcgtctccC CTCAATGCTGCTGCTG GTACTCTACGATTTTACCA GTTGCGCTCCACTGTGA CTAAGATTGcaggctgttacctc | |
| PROM1   | NM_001145848.1 | PROM1_8          | ctgggaactatcctctCC  
|         |               |                  | CTCAATGCTGCTGCTGT    
|         |               |                  | ACTCTACGATTTTACCA    
|         |               |                  | GTTGGCGCTCCACTGTTA   
|         |               |                  | CTAGATTGtacttaacctacta 
|         |               |                  | aataattgaaga          
| PROM1   | NM_001145848.1 | PROM1_9          | attggaactatgtaaggC   
|         |               |                  | CTCAATGCTGCTGCTGT    
|         |               |                  | ACTCTACGATTTTACCA    
|         |               |                  | GTTGGCGCTCCACTGTTA   
|         |               |                  | CTAGATTGggaagcataagcagtga 
| PROM1   | NM_001145848.1 | PROM1_10         | agcttctctggatttttgC  
|         |               |                  | CTCAATGCTGCTGCTGT    
|         |               |                  | ACTCTACGATTTTACCA    
|         |               |                  | GTTGGCGCTCCACTGTTA   
|         |               |                  | CTAGATTGagagtaactaggattct 
| PROM1   | NM_001145848.1 | PROM1_11         | tgttgtgtgtatttttcC   
|         |               |                  | CTCAATGCTGCTGCTGT    
|         |               |                  | ACTCTACGATTTTACCA    
|         |               |                  | GTTGGCGCTCCACTGTTA   
|         |               |                  | CTAGATTGgccaaccgetc 
|         |               |                  | ctagataac            
| PROM1   | NM_001145848.1 | PROM1_12         | aaatatggaatgtgttaC   
|         |               |                  | CTCAATGCTGCTGCTGT    
|         |               |                  | ACTCTACGATTTTACCA    
|         |               |                  | GTTGGCGCTCCACTGTTA   
|         |               |                  | CTAGATTGgaaactatacccagttga 
| SST     | NM_001048.3    | SST_1            | GAACTGGCCAAGTACTTC   
|         |               |                  | TGTGTCTATTTTAGTGGGA  
|         |               |                  | TCCAGTAGCCCGTGACTA   
|         |               |                  | TCCTCTTGTGCTGTGA     
|         |               |                  | TGATCGTGCTCCGCGCC    
|         |               |                  | GGGGAAGCAG            
| SST     | NM_001048.3    | SST_2            | gctggctgegttgtccatT  
|         |               |                  | GTGTCATATTTAGTGGAAT  
|         |               |                  | CCAGTAGCCGCTGACTAT   
|         |               |                  | CGTCTCTTGTGCGTGTA    
|         |               |                  | GATCGTGCTCgctgcctc   
|         |               |                  | cagtgege             

37
| Gene | Accession | Description | Sequence 1 | Sequence 2 |
|------|-----------|-------------|------------|------------|
| SST  | NM_001048.3 | SST_3       | cctggaacctgaagacgatcT  |
|      |            |             | GTGTCTATTTTAGTGGAT     |
|      |            |             | CCAATAGCCGTGACTAT      |
|      |            |             | CGTCTCTTTGTGCTGTAT     |
|      |            |             | GATCGTCCacgaaggagag    |
|      |            |             |aatgatgc               |
| SST  | NM_001048.3 | SST_4       | aagattttctggaagacgatcT |
|      |            |             | GTGTCTATTTTAGTGGAT     |
|      |            |             | CCAATAGCCGTGACTAT      |
|      |            |             | CGTCTCTTTGTGCTGTAT     |
|      |            |             | GATCGTCCacgcaagacgag  |
|      |            |             |tggctgc                |
| ARX  | NM_139058.2 | ARX_1       | GCTTTTTTCACAATGGAGT    |
|      |            |             | TGTCTATTTTAGTGGAT      |
|      |            |             | CGTCTCTTTGTGCTGTAT     |
|      |            |             | GATCGTCCacgcaagacgag  |
|      |            |             |tggctgc                |
| ARX  | NM_139058.2 | ARX_2       | ccaaaagtaaatctccaaagcT |
|      |            |             | GTGTCTATTTTAGTGGAT     |
|      |            |             | CCGTGCCGCTGTGAGC       |
|      |            |             | AATTACTTTGTGCTGTAT     |
|      |            |             | GATCGTGCCcagcneagcag  |
|      |            |             |cccgagag                |
| ARX  | NM_139058.2 | ARX_3       | ggccgagctcagcacaactgcT |
|      |            |             | GTGTCTATTTTAGTGGAT     |
|      |            |             | CCGTGCCGCTGTGAGC       |
|      |            |             | AATTACTTTGTGCTGTAT     |
|      |            |             | GATCGTGCCcagcneagcag  |
|      |            |             |ccgttcca                |
| ARX  | NM_139058.2 | ARX_4       | ggacagctcaagagctcaagcT |
|      |            |             | GTGTCTATTTTAGTGGAT     |
|      |            |             | CCGTGCCGCTGTGAGC       |
|      |            |             | AATTACTTTGTGCTGTAT     |
|      |            |             | GATCGTGCCcagcneagcag  |
|      |            |             |ccgctcgc                |
| ARX  | NM_139058.2 | ARX_5       | tggaggaaggagagagagagT  |
|      |            |             | GTGTCTATTTTAGTGGAT     |
|      |            |             | CCGTGCCGCTGTGAGC       |
|      |            |             | AATTACTTTGTGCTGTAT     |
|      |            |             | GATCGTGCCcagcneagcag  |
|      |            |             |gaagangag               |
| Gene | Accession | Gene | Accession | Sequence |
|------|-----------|------|-----------|----------|
| ARX  | NM_139058.2 | ARX_6 | | gctgacccggaagagcT GGTGCATTTTAGTGGAT CCGTGCACCCTGGTACG AATTACTTGCTGTAT GATCGTCACGcaaggag gagctgtct |
| ARX  | NM_139058.2 | ARX_7 | | ccagaacgctegggaT GGTGCATTTTAGTGGAT CCGTGCACCCTGGTACG AATTACTTGCTGTAT GATCGTCACGcaagctcag gtcggttc |
| ARX  | NM_139058.2 | ARX_8 | | cgcgcgcgcgcgttcT GGTGCATTTTAGTGGAT CCGTGCACCCTGGTACG AATTACTTGCTGTAT GATCGTCACGcgcgcgct gcgcgcgcg |
| ARX  | NM_139058.2 | ARX_9 | | cgggacacccggaagggT GGTGCATTTTAGTGGAT CCGTGCACCCTGGTACG AATTACTTGCTGTAT GATCGTCACGcgcgcgctca acatcctg |
| PDX1 | NM_000209.3 | PDX1_1 | | GCGGCGCCCTACGCTGGG GCCGCAAATCTGCTGCTG TACCGTCACCCTGGTACG GATTTAGCTCACAAGTTT ACTAGATTGGGAAGGCG CAGTGGCGG |
| PDX1 | NM_000209.3 | PDX1_2 | | ccgggacacccggaagggC CTCAATGCTGCTGCTG TACCGTCACCCTGGTACG GATTTAGCTCACAAGTTT CTAGATTGgcgcgcgcgc gcgcgcgcgc |
| PDX1 | NM_000209.3 | PDX1_3 | | aagatcctggccacaaaaC CTCAATGCTGCTGCTG TACCGTCACCCTGGTACG GATTTAGCTCACAAGTTT CTAGATTGgcgcgcgcgc gcgcgcgcgc |
| Gene  | Accession  | Position | Sequence                          |
|-------|------------|----------|-----------------------------------|
| PDX1  | NM_000209.3| PDX1_4   | tcggcgaggagttctgC                 |
|       |            |          | CTCAATGCTGCTGCTGT                 |
|       |            |          | ACCGTGCGCCTGTTGTC                 |
|       |            |          | AATTAGCTCCACTGTTGTA               |
|       |            |          | CTAGATTGGaggtgcgtgagctggcC        |
|       |            |          | cccggcacc                      |
| PDX1  | NM_000209.3| PDX1_5   | ccttcaacgtgcggtcgcC              |
|       |            |          | CTCAATGCTGCTGCTGT                 |
|       |            |          | ACCGTGCGCCTGTTGTC                 |
|       |            |          | AATTAGCTCCACTGTTGTA               |
|       |            |          | CTAGATTGGttagccgcggtgcgctggcC    |
|       |            |          | cgcaccac                      |
| NEUROD1 | NM_002500.4| NEUROD1_1 | AATTCAATTTCTACTTTAAA            |
|       |            |          | AGTACGCAAT GAATCCGTA             |
|       |            |          | AGTCGTGCGCCTGTTGTA               |
|       |            |          | GAATTATCTGTCTGTTGTA              |
|       |            |          | AGTTAGTCCAAAT TGGCA              |
|       |            |          | CAATTTGAGC                      |
| NEUROD1 | NM_002500.4| NEUROD1_2 | cgaggtgctctgtgctctgctc          |
|       |            |          | GTACGCATGAATCCGTA                |
|       |            |          | GTCGTGCGCCTGTTGTA                |
|       |            |          | AAATTATCTGTCTGTTGTA              |
|       |            |          | AGTTAGTCCcttcaacgcgggtcggcC     |
|       |            |          | tggacaga                      |
| NEUROD1 | NM_002500.4| NEUROD1_3 | gaaggctacagccgggA               |
|       |            |          | GTACGCATGAATCCGTA                |
|       |            |          | GTCGTGCGCCTGTTGTA                |
|       |            |          | AAATTATCTGTCTGTTGTA              |
|       |            |          | AGTTAGTCCtttaaatttg              |
|       |            |          | agacgcgctcagctctctcagctctcagctc|
| NEUROD1 | NM_002500.4| NEUROD1_4 | cctgtgtctctgtgctctgtc          |
|       |            |          | GTACGCATGAATCCGTA                |
|       |            |          | GTCGTGCGCCTGTTGTA                |
|       |            |          | AAATTATCTGTCTGTTGTA              |
|       |            |          | AGTTAGTCCgggtcgcgggtcgc          |
|       |            |          | gggcagggtcgcgggtcgcgggtcgc       |
| NEUROD1 | NM_002500.4| NEUROD1_5 | tccggcgcagctcagctacA          |
|       |            |          | GTACGCATGAATCCGTA                |
|       |            |          | GTCGTGCGCCTGTTGTA                |
|       |            |          | AAATTATCTGTCTGTTGTA              |
|       |            |          | AGTTAGTCCgttccacgcgggtcggcC     |
|       |            |          | gtttaagc                      |
| Gene   | Accession | Exon | Sequence                        |
|--------|-----------|------|---------------------------------|
| NEUROD1| NM_002500.4| 6    | tgcccgacactggcagA               |
|        |           |      | GTCGCAATGAAATCCCTA             |
|        |           |      | GTCGTGGGCTTGGTACG              |
|        |           |      | AATTATCTGTCTGTTGA              |
|        |           |      | GTTAGCCTtttaacctg              |
|        |           |      | cactatcc                      |
| NEUROG3| NM_020999.3| 1    | TGCGCTCATTTTAGGgCCT             |
|        |           |      | AGTCCGAAGTACTACTCT             |
|        |           |      | TCTCCTAGATGTAACCGC             |
|        |           |      | TATCGCTTTGGCTTGTA              |
|        |           |      | TGATCGTCGGGAAGTGggG            |
|        |           |      | CATTGCAAAAG                   |
| NEUROG3| NM_020999.3| 2    | cgccccactgtccaagtggA          |
|        |           |      | GTCGGAAGTACTACTCTCT            |
|        |           |      | ATCTCTAGATGTAACCGCT            |
|        |           |      | ATTCGTCTTGTGCTGTAT             |
|        |           |      | GATCGTCGCgccctacaacc           |
|        |           |      | ctcggggtg                     |
| NEUROG3| NM_020999.3| 3    | cgaagaagggcccaacgaA            |
|        |           |      | GTCGGAAGTACTACTCTCT            |
|        |           |      | CTCTCTAGATGTAACCGCT            |
|        |           |      | ATTCGTCTTGTGCTGTAT             |
|        |           |      | GATCGTCGCgccctacaacc           |
|        |           |      | ctcggggtg                     |
| NEUROG3| NM_020999.3| 4    | agcttgtacgcgtggaA             |
|        |           |      | GTCGGAAGTACTACTCTCT            |
|        |           |      | ATCTCTAGATGTAACCGCT            |
|        |           |      | ATTCGTCTTGTGCTGTAT             |
|        |           |      | GATCGTCGCgccctacaacc           |
|        |           |      | ctcggggtg                     |
| NEUROG3| NM_020999.3| 5    | ccagcgactgctgcttA            |
|        |           |      | GTCGGAAGTACTACTCTCT            |
|        |           |      | ATCTCTAGATGTAACCGCT            |
|        |           |      | ATTCGTCTTGTGCTGTAT             |
|        |           |      | GATCGTCGCgccctacaacc           |
|        |           |      | ctcggggtg                     |
| INS    | NM_000207.2|      | AGGTGGGGGAGGTGGGAGC           |
|        |           |      | CTCAATGCTGCTGCTGCTGT          |
|        |           |      | ACTCTAGATGTTTACCA             |
|        |           |      | GTTGCCCTAGATGTGCC             |
|        |           |      | GCTATTTGTCCGGGAGGTC           |
|        |           |      | AGAGGACCTGC                   |
| GCG | NM_002054.4 | GCG |
|-----|-------------|-----|
|     |             | GAATAACATTGCCAAACG |
|     |             | TGTGTCTATTTAGTGGA  |
|     |             | TCCGTGCCTGGTAG     |
|     |             | CAATTAGCTCCACTGTT   |
|     |             | ACTAGATTGGAAATACCA  |
|     |             | AGAGGAACAG         |
Supplementary Table 9: Detection Probes

| Hybridization Cycle | Probe name | Probe sequence (5’→3’) |
|---------------------|------------|------------------------|
| 1                   | DO_1.1     | AGUCGGAAGUAUCTACTCUCT_FITC |
| 1                   | DO_1.2     | CCUCAATGCUGCTGCTGUAC_Cy3 |
| 1                   | DO_1.3     | TGUGTCTATUTAGTGGAUCC_Cy5 |
| 1                   | DO_1.4     | AGUACGCAUGAAUCCGUAGT_TR  |
| 2                   | DO_2.1     | CGUGCAGCCUGGTAGCAAUTA_FITC |
| 2                   | DO_2.2     | AGUAGCCGUGACTATCGUGT_Cy3 |
| 2                   | DO_2.3     | TCUACGATUTTACCAGTG_Cy5   |
| 2                   | DO_2.4     | CCUAGATGUACGCUAUCGT_TR   |
| 3                   | DO_3.1     | CCUAGATGTUCCGCTAFUGT_FITC |
| 3                   | DO_3.2     | GCUCACGTGTGACGTAUTG_Cy3  |
| 3                   | DO_3.3     | CTUGTCGTGUATGATCGUCC_Cy5 |
| 3                   | DO_3.4     | TCUGTCTGUTGAGUTAGUCC_TR  |
Supplementary References

[1] Mignardi, M. et al. Oligonucleotide gap-fill ligation for mutation detection and sequencing in situ. *Nucleic acids research* **43**, e151–e151 (2015).

[2] Ke, R. et al. In situ sequencing for rna analysis in preserved tissue and cells. *Nature methods* **10**, 857 (2013).

[3] Jones, M. C. Simple boundary correction for kernel density estimation. *Statistics and Computing* **3**, 135–146 (1993).

[4] Svensson, V., Teichmann, S. A. & Stegle, O. Spatialde: identification of spatially variable genes. *Nature methods* **15**, 343 (2018).