



Supplementary Figure 1. Diagrammatic representation of the Bioinformatics analysis performed. Filtering is performed based on the signal intensity and on the criterion whether this signal is above a certain level. In our analysis, filtering was performed using the equation: $S < B_L + 1.5 \cdot \sigma_{B_L}$, where S is the measured signal intensity, B_L is the local background measured and σ_{B_L} is the standard deviation of the local background. Background correction was performed by subtracting the median global background from the median local background from the signal intensity. A threshold of 2 was set as cut-off, meaning that spot intensity should be twice as much as that of the background. Microarray data were cross-normalized, using a quantile algorithm, in order to account for the bias that is included due to experimentation, different platforms and different sampling. In order to compare all the available microarray datasets among them, we used the following methodology:

a) First, we searched for differences, comparing all control samples, which we considered as one group, against all tumor samples, which we considered as another group, using a two-tailed two-sample T-test. Since those groups contained samples varying from nationality to tumor grade, we entailed all bias by comparing them as unified groups. It would be expected that this analysis would give genes that are common, due to the common denominator, which is the fact that all samples derive from urinary bladder cancerous tissue.

b) Second, we separated samples into groups (11 in total) and each group was compared against all control samples, using a two-tailed two-sample T-test. We expected that this would give less common genes among all groups simultaneously.

c) We compared samples individually for significant genes among each experiment, using a two-tailed z-test, which would be referred to as “intra-experimental”. This type of comparison had a particularity. Since genes i.e. gene ratios were compared to the mean of the genes of the same experiment, then differentially expressed genes would signify the difference that each tissue has against the normal tissues. This means that the common genes among them would be those genes that are common due to the tumor tissue.

d) We compared samples individually for significant genes i.e. gene ratios, among experimental setups, using a two-tailed z-test, which we would refer to as “inter-experimental”. In other words, we searched for genes that are different from one sample to the next and not against the control samples. Interestingly, the significant genes derived from those genes that were not differentially expressed. In other words, they derived from those that remained the same across all samples. We would expect those that were common to all samples simultaneously, if such exist, as those that are universally common to all tumors, regardless of tissue origin or experimentation.

In order to find the differentially expressed genes, we used two methods. Genes were considered to be significantly differentially expressed if they obtained a p-value < 0.05 . Comparisons were made both among experiments as well as within experiments. Set manipulation was then used in order to discover further subsets that would characterize, if possible, all tumor samples. For further analyses we used the genes that were differentially expressed among tumor samples, on a need-to-use basis. In the case where sample groups were compared, the mean of each gene was taken against the mean of all control samples. In the case of individual comparisons of samples, gene ratios were formed against the mean of all control samples.