SUPPLEMENTARY MATERIAL

Algorithm for Allelic imbalance scoring

The algorithm to score AI was based on a Hidden Markov Model (HMM). The HMM has two hidden (unobserved) states, Loss and No-Loss, that is assigned to all SNP positions in all tumors. The HMM jumps between the two states with probabilities that depend on the state of the previous position. For a given position a symbol signifying the SNP call in blood and tumor tissue is emitted. The probability of emitting a given symbol depends on the two states, as illustrated in the table below

Blood-Tumor	Hetero-Hetero	Hetero-Homo	Hetero-No Call	No Call-Hetero	All Others
Loss	0	1-p-r	p	0	r
No-loss	1-q-r	0	q	q	r-q

In the table Hetero is short for the SNP is heterozygous, Homo for homozygous, and No Call for a no call. There are few observed cases where a SNP is homozygous AA in blood and BB in tumor, or homozygous in blood and heterozygous. Based on this observation we assume all experimental errors, SNPs with poor quality and erroneous scoring of a SNP with Affy software give rise to no calls. Thus Hetero-Hetero can only occur in the No-Loss state, and Hetero-Homo only in the Loss state. The probability of emitting Hetero-No Call and No Call-Hetero in the No-Loss state is the same, because No Call is assumed due to an experimental error. Also Homo-Homo has the same probability of being emitted in the two states because it is impossible to tell whether an allele has been lost or not when Homo in blood. The probability of Hetero in blood is independent of whether the HMM is in the Loss or No-Loss state, thus the probabilities in columns 5 and 6. We used the EM-algorithm to estimate the parameters in the table and the most likely combination of Loss and No-loss for each tumor. Parameter estimation was made jointly for all tumors in the stable groups and jointly for the tumors in the unstable group. If the number of Hetero-Homo's (less than 25) is small the regions of loss is underestimated, i.e., some Hetero-No Call is incorrectly classified as No-loss. A refinement of the model could be made by taking the distance between markers into account. However, most markers are fairly equally spaced and the present model captures the basic dependencies between markers.

Examples on how the algorithm used to make the call of AI and non-AI is defining AI based on the presence of homo – or heterozygosity in blood and tissue.

	Reading direction →					Distance<x	Distance=x
Blood	Homo	Homo	Hetero	Homo	Hetero	Hetero	Hetero
Tumor	Homo	Homo	Homo	Homo	Homo	No Call	Hetero
Score	Non-AI	Non-AI	AI	AI	AI	AI	Non-AI

Legend to supplementary figure 1 (Expression level in bladder cancer of the genes UNC5B and UNC5C):
The histograms show level of expression of UNC5B and UNC5C based on analysis using customized Affymetrix expression arrays. Data for UNC5B are from different stages of bladder

cancer, while data for UNC5C are from normal urothelium and stage Ta grade I and Ta grade III bladder tumors. The left axis indicates abitrary units of expression as generated by the Affymetrix software. “n” indicates number of samples analyzed.
UNC5B expression was significantly different in stage T1 (*T-test, p<0.05) and in stage T2-4 (**p<0.01) compared to stage Ta. UNC5C expression was significantly different in Ta grade I and in Ta grade III (**p<0.01) compared to normal bladder tissue and significantly different in Ta grade III (#p<0.05) compared to Ta grade I.
UNC5B and UNC5C expression data were generated previously, in-house, using either customized Affymetrix GeneChip EOS Hu03 (the UNC5B expression data, Lars Dyrskjøt, Manuscript in prep.) or Affymetrix GeneChip U133A Array (the UNC5C expression data, Mads Aaboe, unpublished results). Purification of total RNA, preparation of cRNA from cDNA, and hybridization and scanning were performed as described in (1). After scanning, the data were normalized using the Robust Multi-array Analysis (RMA) normalization approach in the Bioconductor Affy package to the R project for statistical computing.

Legend to supplementary figure 2 (Supplementary data-T1 chr13 and chr14):
T1 tumors that were followed by muscle invasive disease: 1058-10, 172-3, 365-1, 501-1, 839-1, 1013-1, 1017-1, 1033-1, 1276-1 (N=9).
T1 tumors that were not followed by muscle invasive disease: 112-12, 747-5, 825-5, 865-2, 140-16, 154-10, 166-14 (N=7).
1058-10, 172-3, 839-1, 1276-1 have an area of AI in common on chr. 13 between 47.09-47.21 MB; slightly displaced (52.9-55.6 MB) 1058-10, 172-3, 1013-1 and 1033-1 have an area of AI in common.
747-5, 140-16, 154-10, 166-14 have an area of common AI at the terminal end of 15q.

Reference List

1. Dyrskjot L, Thykjaer T, Kruhoffer M, Jensen JL, Marcussen N, Hamilton-Dutoit S, Wolf H, Orntoft TF. Identifying distinct classes of bladder carcinoma using microarrays. *Nat.Genet.* 2003;33:90-6.