Legends to Supplementary Data

Supplementary Figure 1. Protein band patterns and summary protein identification data in NM fractions of colorectal tissue types. A. Coomassie stained protein gel with 2 adenoma samples (ADE), 2 MIN and 2 CIN carcinoma samples. B. Venn diagrams that display numbers of overlapping and unique proteins in biological samples of the same colorectal tissue types. C. Cumulative bar graph to display the number of overlapping proteins against the number of biological samples in which the proteins were identified.

Supplementary Table 1. Clinical and histopathological features of the patients and their respective tumours included in this study. Samples are identified by patient number (p#). Abbreviations: MIN, microsatellite instable; MSS, microsatellite stable; CIN, chromosomal instable; N/A, not applicable. † Grade refers to dysplasia in the case of adenomas, and differentiation in the case of carcinomas.

Supplementary Table 2. Reproducibility Study: Identification and spectral counting quantitation of proteins in triplicate CRC nuclear matrix (NM) fractions. The sheet "ALL Proteins" details all proteins identified in any of three independent NM preparations (NM_1, NM_2, NM_3) prepared from aliquots of one pool of CRC tissues. The sheet "889 Proteins In ALL Replicates" details all proteins identified in all three replicates. Quantitation was performed by spectral counting. Raw spectral counts were normalized by multiplying with a correction factor that was calculated by dividing the average sum of spectral counts in all three samples by the sum of spectral counts in the pertinent sample. Protein identification probability was calculated using the Protein Prophet algorithm. In the "889 Proteins In ALL Replicates" sheet, overall reproducibility of the workflow was calculated by calculating the percentual coefficient of variation (CV%) for all reproducibly identified proteins, and taking the average.

Supplementary Table 3. NM-Enriched Proteins: Statistical assessment of proteins enriched in the NM fraction of 6 patient samples. Paired comparisons between chromatin-binding (CB), intermediate filament (IF), and nuclear matrix (NM) fractions were performed with the normalized spectral counts for each protein separately, using the Mantel-Haenszel test. p-Values are given for each paired test. Fold Change refers to the common odd ratio. The final column enumerates how many of the 6 samples show an enrichment for the pertinent protein in the NM fraction relative to the corresponding CB and IF fractions. Proteins enriched in the NM of 5 out of 6, or 6 out of 6 samples were considered to be "NM-enriched", and only these proteins are shown here. Spectral counts were normalized as described in Supplementary Table 2. Abbreviations: p2, p3, p4, p5, p7, p8 : patient sample IDs; ADE, adenoma; CRC, colorectal carcinoma; MIN, microsatellite instable; CIN, chromosomal instable.

Supplementary Table 4. Identification and spectral counting quantitation of proteins in triplicate, unfractionated CRC nuclear lysates. The table details all proteins identified in any of three independent CRC sample lysates (CRC_1, CRC_2, CRC_3). Quantitation was performed by spectral counting. Protein identification probability was calculated using the Protein Prophet algorithm.
Supplementary Table 5. Differential proteins in NM fractions of adenoma versus CRC tissues.

Identification and spectral counting quantitation of proteins was performed for NM fractions isolated from two adenoma (ADE) and four colorectal carcinoma (CRC) patient samples. Two of the CRC samples were associated with microsatellite instability (MIN), and two were associated with chromosomal instability (CIN). Significantly differential proteins in the NM of ADE versus CRC samples were selected with a statistical cutoff of p<0.05, and then manually curated on the basis of the quantitative information contained in spectral counts, following criteria detailed in the text. Cancer association was deduced from the "Disorder" section of the respective protein entries in the GeneCards web site (http://www.genecards.org), subcellular localization was deduced from all relevant levels in the respective protein entries in the UniProt Knowledge Base (http://www.uniprot.org/uniprot), and significant changes (p<0.1) in mRNA levels were taken from previous data (available at Gene Expression Omnibus (GEO), http://www.ncbi.nlm.nih.gov/geo, accession number GSE8067). The final column enumerates how many of the 6 samples show an enrichment for the pertinent protein in the NM fraction relative to the corresponding CB and IF fractions.

Further abbreviations: CSK, cytoskeleton; CYT, cytoplasm; ECM, extracellular matrix; EXC, extracellular; IF, intermediate filaments; membr., membrane; MITO, mitochondria; NE, nuclear envelope; NO, nucleolus; NPC, nuclear pore complex; NUCL, nucleus; PM, plasma membrane. Up arrow indicates significant higher level of mRNA in CRC relative to ADE; down arrow indicates significant lower level.

Supplementary Table 6. Differential proteins in NM fractions of adenoma versus CIN (CRC) tissues.

Identification and spectral counting quantitation of proteins was performed for NM fractions isolated from two adenoma (ADE) patient samples and two colorectal carcinoma (CRC) patient samples with chromosomal instability (CIN). Significantly differential proteins in the NM of ADE versus CIN samples were selected with a statistical cutoff of p<0.05, and then manually curated on the basis of the quantitative information contained in spectral counts, following criteria detailed in the text. Cancer association was deduced from the "Disorder" section of the respective protein entries in the GeneCards web site (http://www.genecards.org), subcellular localization was deduced from all relevant levels in the respective protein entries in the UniProt Knowledge Base (http://www.uniprot.org/uniprot), and significant changes (p<0.1) in mRNA levels were taken from previous data (available at Gene Expression Omnibus (GEO), http://www.ncbi.nlm.nih.gov/geo, accession number GSE8067).

Further abbreviations: CSK, cytoskeleton; CYT, cytoplasm; ECM, extracellular matrix; EXC, extracellular; IF, intermediate filaments; membr., membrane; MITO, mitochondria; NE, nuclear envelope; NO, nucleolus; NPC, nuclear pore complex; NUCL, nucleus; PM, plasma membrane. Up arrow indicates significant higher level of mRNA in CRC relative to ADE; down arrow indicates significant lower level.

Supplementary Table 7. Differential proteins in NM fractions of adenoma versus MIN (CRC) tissues.

Identification and spectral counting quantitation of proteins was performed for NM fractions isolated from two adenoma (ADE) patient samples and two colorectal carcinoma (CRC) patient samples with microsatellite instability (MIN). Significantly differential proteins in the NM of ADE versus MIN samples were selected with a statistical cutoff of p<0.05, and then manually curated on the basis of the quantitative information contained in spectral counts, following criteria detailed in the text. Significant
changes (p<0.1) in mRNA levels were taken from previous data (available at Gene Expression Omnibus (GEO), http://www.ncbi.nlm.nih.gov/geo, accession number GSE8067). **Up arrow** indicates significant higher level of mRNA in CRC relative to ADE; **down arrow** indicates significant lower level.