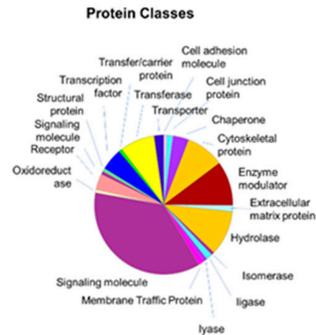


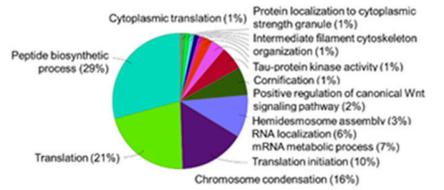
# **Supplementary Materials: Nucleolin-SLeA glycoforms as E-selectin ligands and potentially targetable biomarkers at the cell surface of gastric cancer cells**

Elisabete Fernandes, Rui Freitas, Dylan Ferreira, Janine Soares, Rita Azevedo, Cristiana Gaitero<sup>a</sup>, Andreia Peixoto, Sofia Cotton, Marta Relvas-Santos, Luis Pedro Afonso, Carlos Palmeira, Maria José Oliveira, Rita Ferreira, André M. N. Silva, Lúcio Lara Santos, José Alexandre Ferreira

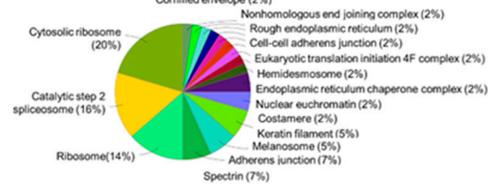
N87



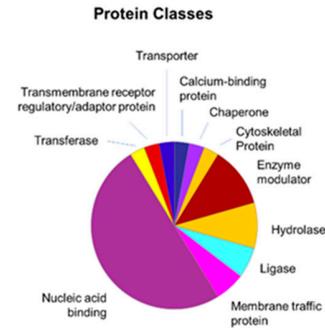
**Main Biological Functions Translated by Protein-Protein interaction Networks**



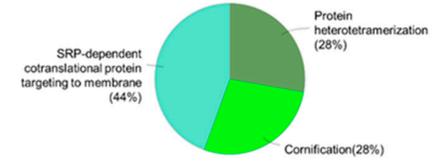
**Main Biological Functions Translated by Protein-Protein interaction Networks**



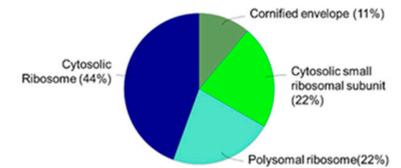
OCUM-1



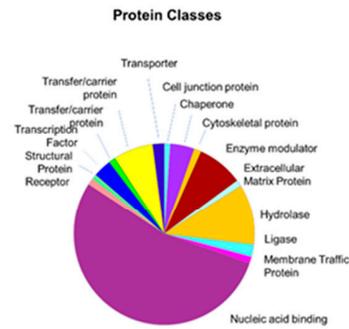
**Main Biological Functions Translated by Protein-Protein interaction Networks**



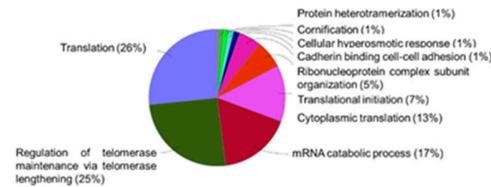
**Main Cellular Locations Translated by Protein-Protein interaction Networks**



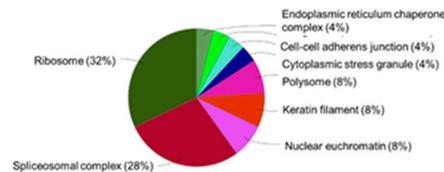
KATO III



**Main Biological Functions Translated by Protein-Protein interaction Networks**



**Main Cellular Locations Translated by Protein-Protein interaction Networks**

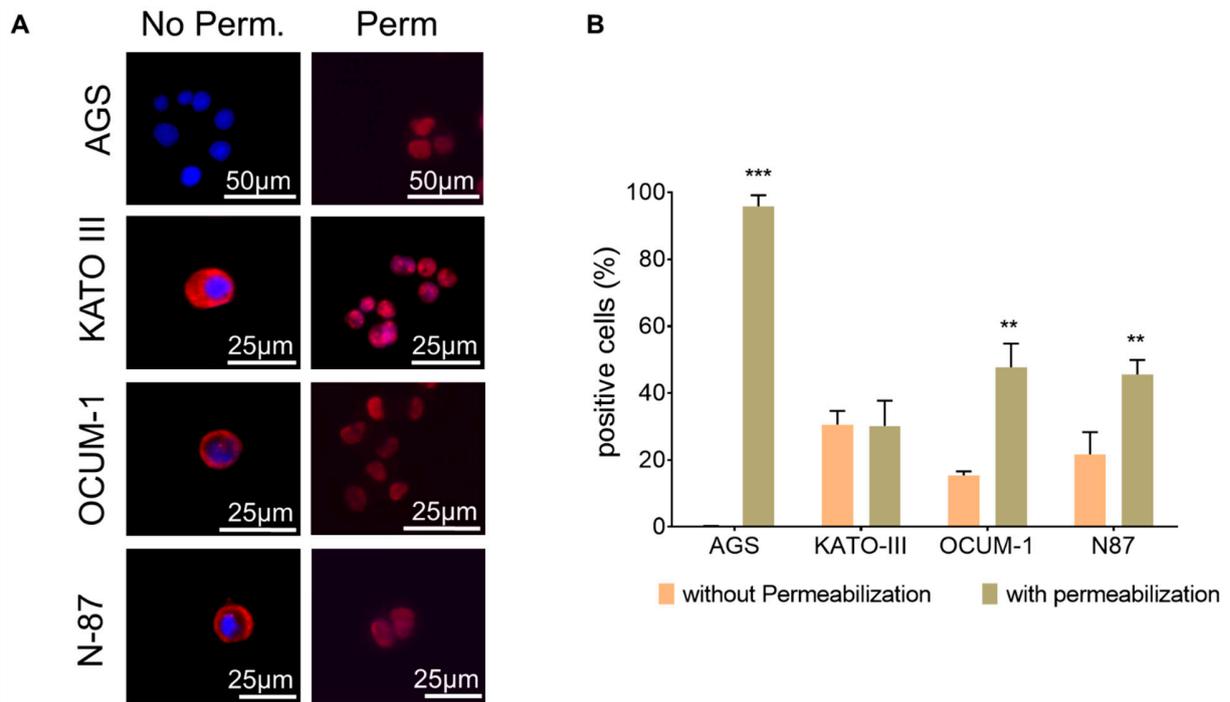


**Figure S1.** Main protein classes and associated biological functions translated by protein-protein interactions and main subcellular localization for glycoproteins expressing O-glycosylated glycans with SLe<sup>A</sup> and showing affinity for E-selectin for N87, OCUM-1 and Kato III cell lines. A wide number of different protein classes are represented for each cell line (20 for N87, 11 for OCUM-1, 14 for KATO III), denoting significant differences

between the cell models. Nevertheless, the main represented protein classes included signaling and nucleic acid binding for N87 and OCUM-1 and N87 cell lines, respectively. Strikingly, many of these glycoproteins (20-60%) were primarily associated with protein translation and biosynthesis in the ribosomes, spliceosome complexes and even at the nuclear euchromatin. Proteins typically found in the cytoplasm and the cytoskeleton were also detected (10-20%). On the other hand, only 10-15% of them could be associated to the plasma membrane, cell junctions and the extracellular space, known to typically undergo O-glycosylation across the secretory pathways.

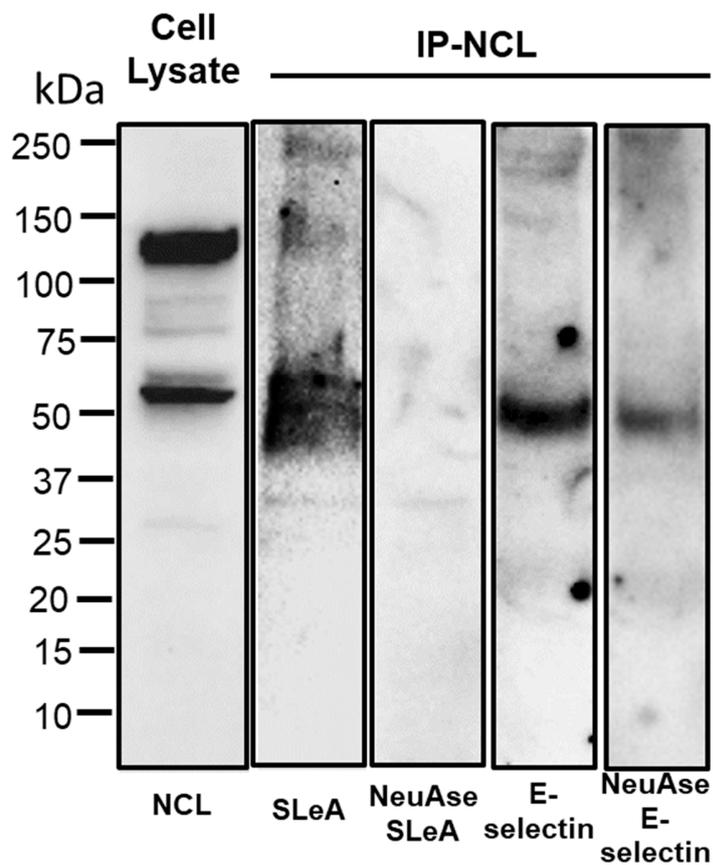
<sup>1</sup>MVKLAKAGKNQGDPKKMAPPPKEVEED**S**EDEEM**S**EDEEDDSSGEEVIPQ  
 KKGKAAA**TS**AKKVVV**ST**KKVAVA**T**PAKKA**V****T**PGKAAA**T**PAK**T**  
 V**T**PAKAV**TT**PGKGA**T**PGKALVA**T**PGKGAIPAKGAKNGKNAKED**S**  
 DEEEDDDSEDEEDDEDEDEDEDEIEPAAMKAAAAAPASEDEDEDEDEDEDD  
 DDDEEDDSEEEAME**TT**PAKGKKA**V**VPVKAKNVAEDEDEEEDDEDEDDDD  
 DEDEDDDDDEDEEEEEEEEEEPVKEAPGKRKKEMAKQKA**PEAKKQK**VEG**T**  
 EP**T**TAFNLFVGNLNF**N**KSAP**EL**KTGISDVFAKNDLAVVDVRIGMTRKFGYVDF  
 ESAEDLEKALELTGLKVF**NEIKLEKPKGKDSK**KERDARTLLAKNLPYKVTQDEL  
 KEVFEDAAEIRLVSKDGKSGIAYIEFKTEADA**E**KTFEEKQGTEIDGRSISLYYTGE  
 KGQNQDYRGGK**NS**TW**SGES**KT**LVLS****N**LSYSATEETLQEVFEKATFIKVPQN  
 QNGKSGYAFIEFASFEDA**KEALNSCN**KREIEGRAIRLELQGPRG**S**PNAR**S**QP  
 SKTLFVKGLSEDTTEETLKESFDG**SVRARI**VDRETGSSKGF**GFVDF**NSEEDAKA  
 AKEAMEDGEIDGNKVTLDWAKPKGEGGF**GGRGG**G**GG**FG**GG**FG**GG**FG**GG**FG**GG**  
 FG**GR**FG**GG**FG**GG**FR**GG**FG**GG**GGDHKPQGGK**TKFE**<sup>710</sup>

**Figure S2.** Potential *N*- (Blue) and *O*-glycosylated (Red) sites for NCL predicted with NetNGlyc 1.0 and NetOGlyc 4.0 bioinformatics tools and using the protein canonical form as reference. The Figure highlights the higher density of potential *O*-glycosylation sites ( $n=23$ ) in comparison to *N*-glycosylation sites ( $n = 3$ ).

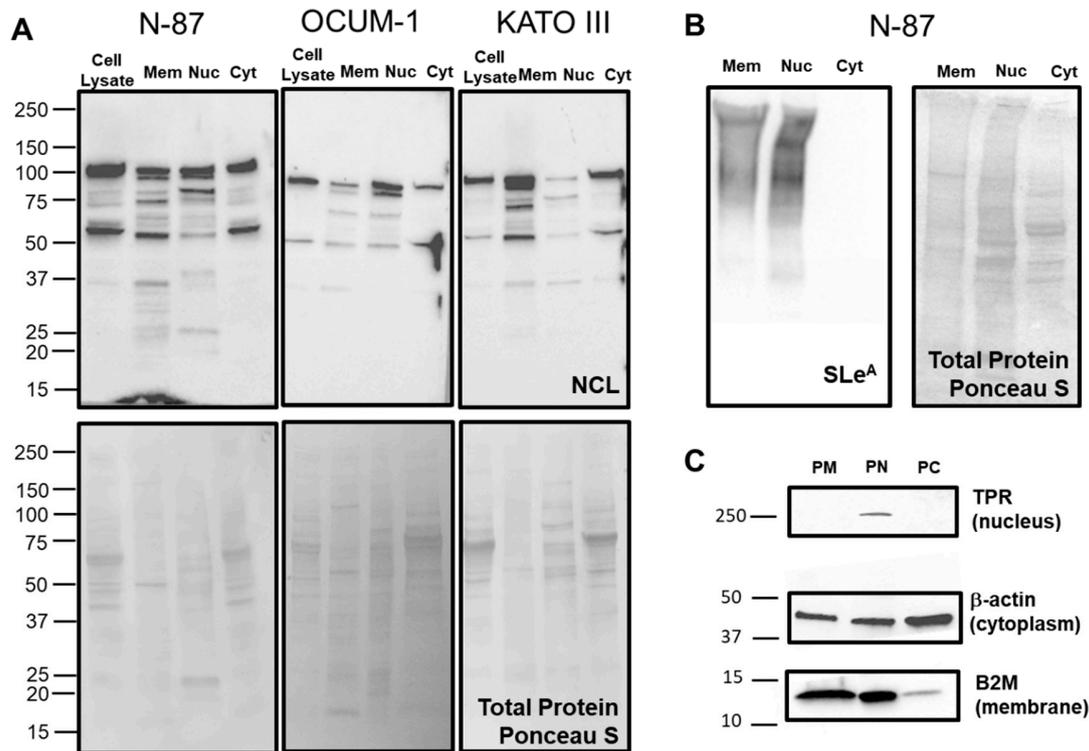


**Figure S3.** (A) Immunofluorescence for NCL, (B) Percentage of NCL positive cells and (C) NCL levels before and after plasma membrane permeabilization. Collectively, this data supports that SLe<sup>A</sup>-negative AGS cells do not express NCL at the cell membrane, as translated by the absence of NCL expression at the cell membrane before cell membrane permeabilization. Notably, almost all AGS cells expressed intracellular NCL. The SLe<sup>A</sup> positive cell lines (KATO-III, OCUM-1, N-87) all showed NCL at the cell membrane before permeabilization. After permeabilization the bulk of the staining was found in the nucleus. Notably, the percentage of NCL positive cells for these models varied between 15-30% and significantly doubled for OCUM-1 and N87 cell lines after permeabilization. These observations suggest that these cell lines may present both intracellular and membrane NCL.

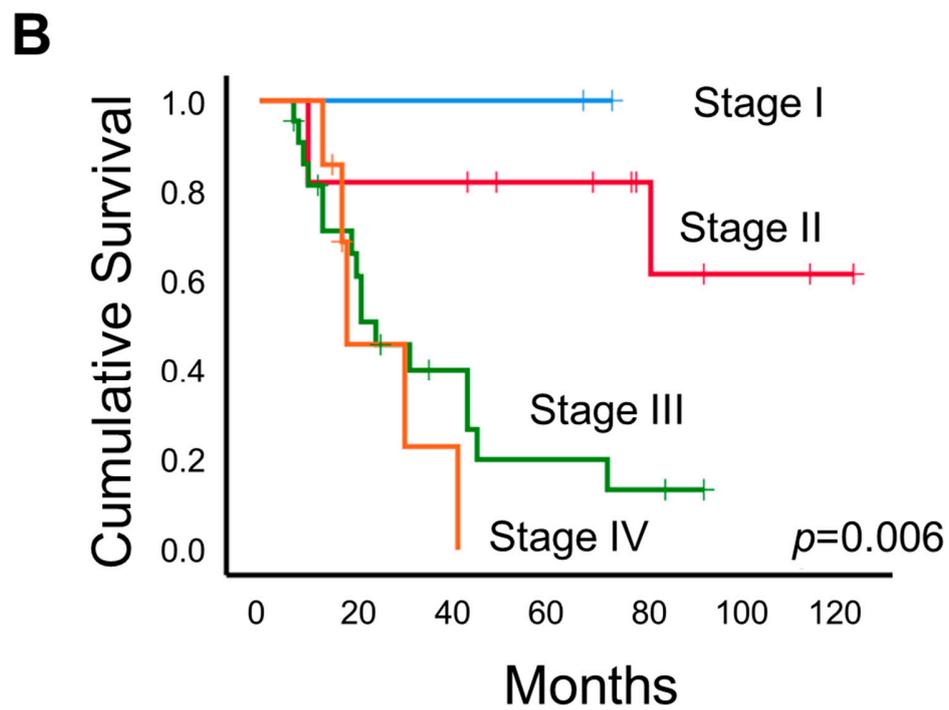
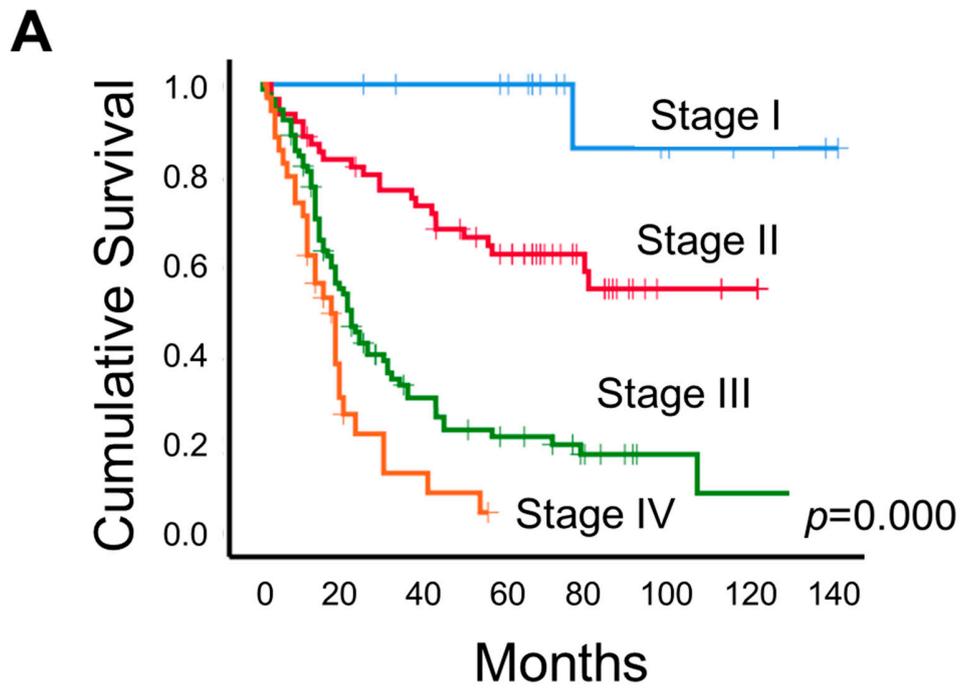
# N87



**Figure S4.** Western blot for NCL in whole cell lysates and SLeA and E-selectin in NCL-IPs. Proteins digested in situ with neuraminidase (NeuAse) where used as negative controls. Collectively these blots support that existence of NCL-SLe<sup>A</sup> expressing glycoproteins at 100 kDa and bellow, with a major band just above 50 kDa. This band and maybe others of less intensity at higher and lower molecular weights, showed affinity for E-selectin mediated by sialylation. The disappearance and/or reduction the signals after NeuAse digestion supports these observations.

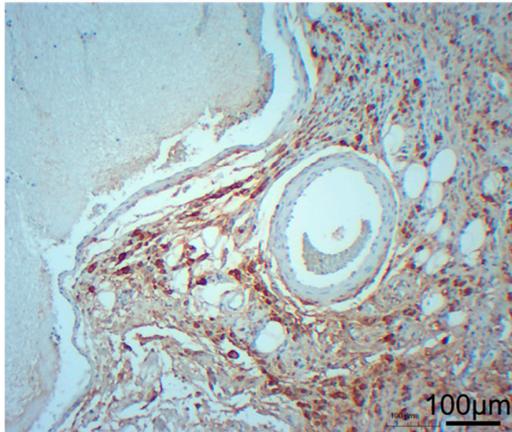


**Figure 5.** (A) NCL expression profiles in the whole cell lysates and different protein extracts enriched for membrane, nuclear and cytoplasmatic proteins. (B) SLe<sup>A</sup> expression in the different fractions. (C) Expression of nuclear (TPR), cytoplasm ( $\beta$ -actin) and B2M (plasma membrane) biomarkers across the different fractions for the N87 cell line. Collectively, the panels show the presence of NCL in all subcellular fractions across the three cell lines, with very similar patterns between different cell lines (panel A). The NCL/Total protein (estimated by Ponceau S) ratio suggested that the membrane protein fraction is more enriched for NCL compared to the other two. In addition, all fractions presented bands at approximately 50 kDa and above 100 kDa, suggesting that these proteoforms could be present in all subcellular locations. Nevertheless, NCL expression patterns varied according to the subcellular fraction, suggesting distinct molecular signatures. Notably, the plasma membrane fraction was significantly enriched for SLe<sup>A</sup> and B2M typically found in at the cell surface (panel B), even though with some contamination with  $\beta$ -actin, suggesting also the presence of cytoplasmatic proteins. However, it did not present TPR, demonstrating little or no contamination with nuclear proteins and reinforcing the membrane origin of some of the observed NCL proteoforms. The nuclear fraction presented high amounts of TPR and, to less extent  $\beta$ -actin. Strikingly, it was also highly contaminated with membrane glycoproteins (SLe<sup>A</sup><sub>high</sub>, B2M<sub>high</sub>), not allowing to fully disclose nuclear-specific signatures. However, the cytoplasmatic fraction suggested little contamination with proteins from other subcellular locations (SLe<sup>A</sup><sub>negative</sub>, B2M<sub>low</sub>, TPR<sub>negative</sub>,  $\beta$ -actin<sub>high</sub>). Taken together, these findings support that NCL at 50 and 100 kDa were not membrane specific. It also strongly suggested the existence of NCL specific proteoforms at the cell membrane, distinct from those of intracellular nature.

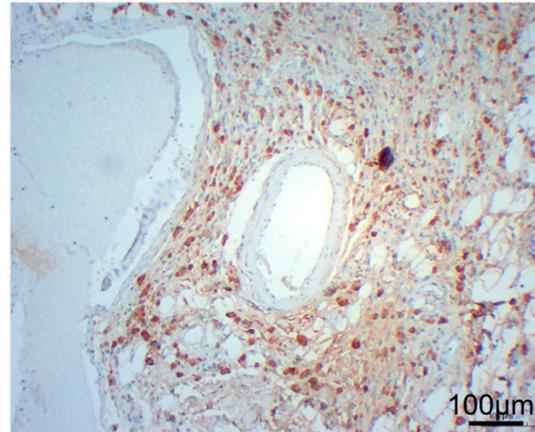


**Figure S6.** Kaplan-Meier survival curves for the patients' cohort used for A) SLe<sup>A</sup> and the patient's subset used for B) NCL and NCL-SLe<sup>A</sup> evaluations. The two datasets reflect a statistically significant decrease in patient's survival with increasing stage of the tumour. The  $p$ -values in survival analyses reflect the Breslow test.

SLe<sup>A</sup> expression in GC



SLe<sup>A</sup> expression in GC  
after PNGase F



**Figure S7.** SLe<sup>A</sup> expression in a gastric tumour section before and after PNGase F digestion. The panels highlight that SLe<sup>A</sup> expression in gastric tumours is not significantly affected by PNGase F digestion. Notably, PNGase F-digested bladder tumours (positive control for *N*-glycans expressing SLe<sup>A</sup>) showed a strong decrease in SLe<sup>A</sup> expression, highlighting that the enzyme is effective in removing *N*-glycans. Moreover, SLe<sup>A</sup> expression decreased after neuraminidase digestion (negative control; not shown). These observations support that the bulk of SLe<sup>A</sup> expression in gastric tumours derives from *O*-glycans.

**Table 1.** Origin and clinicopathological features for the GC cell lines adopted for this study.

Cell line	Cell origin	Gender	Age	Lauren classification	Differentiation	Ethnicity
AGS	Primary tumour	Female	54	Intestinal type	Moderately differentiated adenocarcinoma	Caucasian
KATO III	Pleural effusion	Male	55	Diffuse type	N/A	Asian
MKN-74	Liver metastasis	Male	37	Intestinal type	Moderately differentiated tubular adenocarcinoma	Asian
NCI-N87	Liver metastasis	Male	N/A	Intestinal type	Well differentiated carcinoma	N/A
OCUM-1	Primary tumour	Female	38	Diffuse type	Poorly differentiated adenocarcinoma containing signet ring cells	Asian