Proteomic and glycoproteomic approaches based on nLC-ESI MS/MS for exploring cancer matrisome: explorative study

Lisa Pagani[1], Clizia Chinello[1], Allia Mahajneh[1], Francesca Clerici[1], Federica Facchinetti[2] , Giuliana Pollaci[2] , Luca Roz[2] and Fulvio Magni[1]

[1] *Department of Medicine and Surgery,* ***University of Milano-Bicocca****, Clinical Proteomics and Metabolomics Unit, Vedano al Lambro, Italy.* [2] *Department of Research,* ***Fondazione IRCCS Istituto Nazionale dei Tumori****, Tumor Genomics Unit, Milan, Italy*

e-mail to contact: l.pagani31@campus.unimib.it

ACKNOWLEDGMENTS

The research leading to these results has received funding from the FAR 2014–2018, from Fondazione Gigi & Pupa Ferrari Onlus and Regione Lombardia POR FESR 2014-2020 (2020-CO-iMMUNITY). Call HUB Ricerca ed Innovazione: ImmunHUB. **Italian Ministry of Health (Grant No. RF-2016-02362946).** 2019 Gilead Fellowship Program.

REFERENCES: [1]Socovich AM, Naba A. *Semin Cell Dev Biol*. 2019;89:157-166. [2]Taha IN, Naba A. *Essays Biochem*. 2019;63(3):417-432. [3]Raghunathan R et al. *Mol Cell Proteomics*. 2019;18(11):2138-2148. [4]Deb B et al. *J Clin Med*. 2019;8(9):1303. [5]Krasny L et al. *Biochem J*. 2016;473(21):3979-3995. [6] Santorelli L et al. *Cancers (Basel)*. 2020;12(1):239. [7]Zhang Y et al. *Anal Chem*. 2008;80(9):3144-3158.

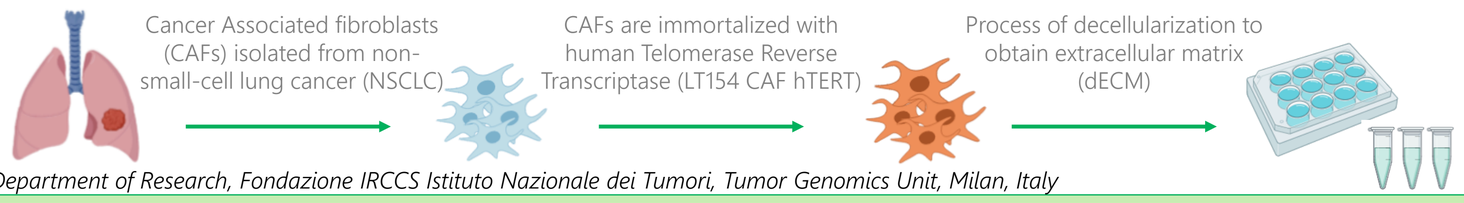
BACKGROUND

All the proteins which compose the extracellular matrix (ECM) of biological samples are defined as **matrisome** [1]. It plays several important roles: from structural to physiological ones and its interaction with other proteins and growth factors is mediated by glycosylation. It's not surprising that, alterations/dysregulation of ECM proteome and its level of glycosylation, have shown a correlation with the development of some human diseases [2-4]. However, since many ECM proteins are large, highly glycosylated and frequently insoluble (due to the presence of covalent cross-links), proteomic characterization of matrisome has been challenging [5].

AIM OF THE STUDY: to develop a method which could be useful to investigate proteomic and glycoproteomic features of cancer matrisome.

WORKFLOW

1st STEP SAMPLE PREPARATION

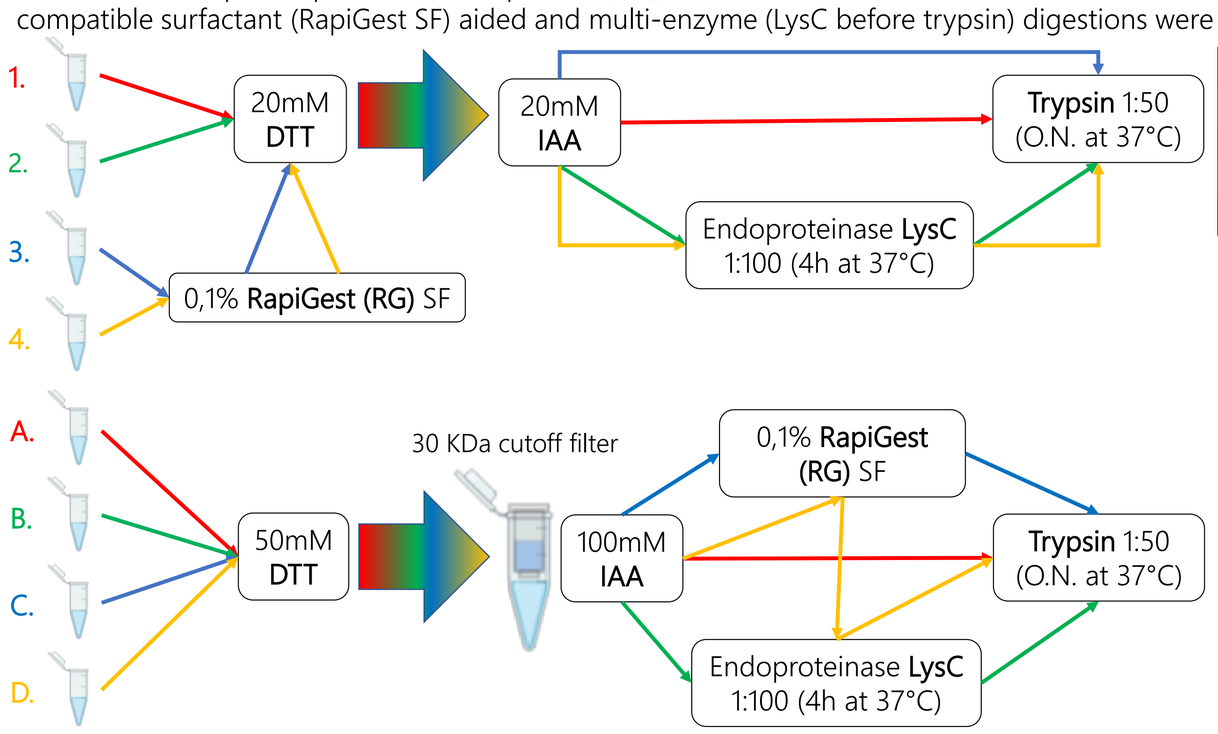


Department of Research, Fondazione IRCCS Istituto Nazionale dei Tumori, Tumor Genomics Unit, Milan, Italy

2nd STEP PROTEIN DIGESTION

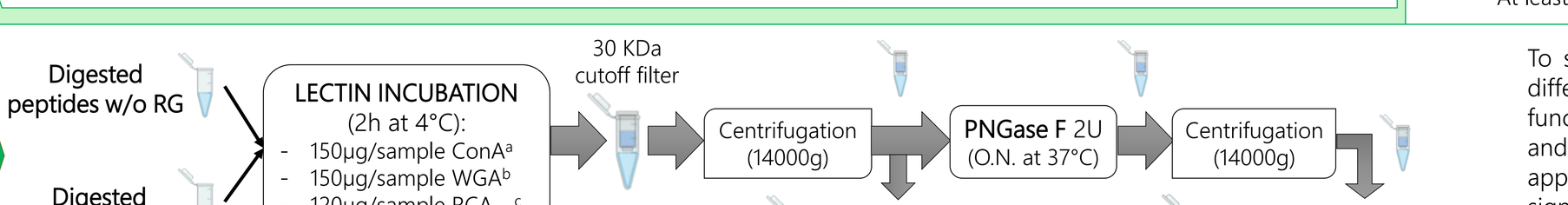
3rd STEP GLYCOPEPTIDES ENRICHMENT

Different methods to enzymatically digest matrisome samples have been compared. In particular, four samples were treated with Filter Aided Sample Preparation (FASP protocol) [6], while other four ones underwent in-solution digestion (InSol.). Also MS-compatible surfactant (RapiGest SF) aided and multi-enzyme (LysC before trypsin) digestions were evaluated.



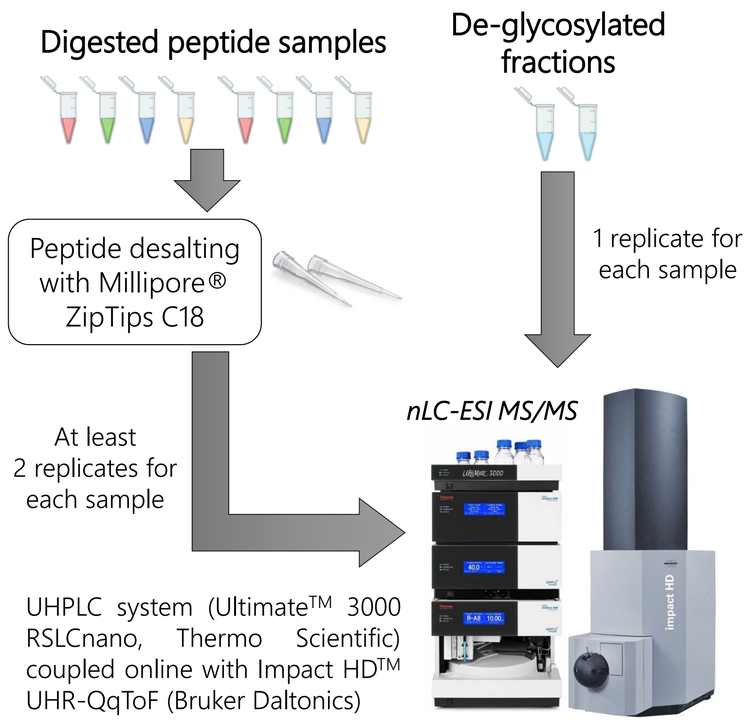
InSol.	RG	LysC	Trypsin
1.			☺
2.		☺	☺
3.	☺		☺
4.	☺	☺	☺

FASP	RG	LysC	Trypsin
A.			☺
B.		☺	☺
C.	☺		☺
D.	☺	☺	☺



a. Concanavalin A; b. Wheat Germ Agglutinin; c. Ricinus Communis Agglutinin 120

4th STEP MASS SPECTROMETRY ANALYSIS



5th STEP PROTEIN IDENTIFICATION

Peaks Studio X+ vers. 10.5
TIC normalization, FDR ≤ 1%
At least 1 significative peptide

To study matrisome glycopeptides, initially two different methods have been used: lectins-functionalized tips (TopTips) [7] (not reported) and N-Glyco-FASP [6] . However, with the first approach it was not possible to identify a significative number of peptides, so we focused our analysis on **N-Glyco-FASP** (workflow on the left).

RESULTS and DISCUSSION

WHICH DIGESTION METHOD IS BETTER?

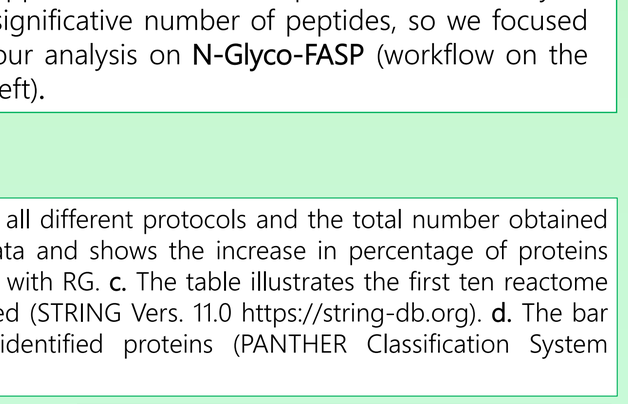
- ✓ Data analysis with **Peaks** platform of digested samples
- ✓ Under the same conditions, **in-solution digestion** allows to identify more proteins than FASP protocol (with a percentage increase from 20% to 100% and an average of 47%) (*image 1a.)*
- ✓ Samples treated with **RapiGest** show a greater identification power in both methods *(image 1b.)*
- ✓ **LysC** seems to not benefit protein identification (a decrease of the identification power has been observed in some cases) *(image 1a.)*
- ✓ **Functional annotation of matrisome proteome** *(image 1c and d.)*
- ✓ Applying the **optimized protocol** (InSol. Digestion + RG) **to an increased ug of proteins** allowed to identify up to 1400 proteins for each sample (data not reported)
- ✓ **CONCLUSION**: in-solution digestion with RapiGest seems to be the more efficient approach to identify matrisome proteins

Image 1. **a.** The table shows the number of proteins identified in all different protocols and the total number obtained combining all the methods. **b.** The graph summarizes table data and shows the increase in percentage of proteins identified between an approach w/o RG and the same approach with RG. **c.** The table illustrates the first ten reactome pathways in which the genes of all identified proteins are involved (STRING Vers. 11.0 https://string-db.org). **d.** The bar chart shows the subcellular position of the genes of all identified proteins (PANTHER Classification System http://www.pantherdb.org/).

a.	Only Trypsin	RG + Trypsin	LysC + Trypsin	RG + LysC + Trypsin	TOTAL
InSol.	218	335	193	338	518
FASP	182	268	133	169	

b.

c. REACTOME PATHWAYS	N° OF IDENTIFIED PROTEINS
Metabolism of proteins	136
Immune System	132
Metabolism	121
Developmental Biology	105
Innate Immune System	94
Signal Transduction	94
Disease	91
Axon guidance	85
Metabolism of RNA	83
Post-translational protein modification	74



d.

IDENTIFICATION OF N-GLYCOPEPTIDES

- ✓ Data analysis with **Peaks** platform of de-glycosylated fractions
- ✓ **N-Glyco-FASP** approach: enrichment of 7% in N-glycoproteins
- ✓ Among all proteins with at least one N-glycosylation site, 56% were found only in de-glycosylated fractions that were digested with RapiGest, 13% in samples without RapiGest and 31% in both *(image 2. to see the full list)*
- ✓ **CONCLUSION**: the use of MS compatible detergents seems to benefit in the glycol-enrichment, BUT an optimization of the method is still required

Image 2. **Summary table** of N-glycosylated proteins (excluding keratins) and their features.

N-glycosylated proteins	Deamidation site (mass shif of 0,98 Da)	Start	End	w/o RG	with RG
VIME_HUMAN	R.DGQVI**N**(+.98)ETSQHHDDLE	451	466	☑	☑
BIP_HUMAN	K.**N**(+.98)GRVEIIANDQGNR.I	47	60		☑
	K.**N**(+.98)GRVEIIA**N**(+.98)DQGNR.I	47	60		☑
PDIA1_HUMAN	R.TVIDY**N**(+.98)GER.T	453	461		☑
CALM2_HUMAN	K.DG**N**(+.98)GYISAAELR.H	96	107	☑	☑
CALM1_HUMAN	K.DG**N**(+.98)GYISAAELR.H	96	107	☑	☑
CALM3_HUMAN	K.DG**N**(+.98)GYISAAELR.H	96	107	☑	☑
TYB10_HUMAN	K.**N**(+.98)TLPTKETIEQEK.R	27	39		☑
YBOX3_HUMAN	K.GAEAA**N**(+.98)VTGPDGVPVEGSR.Y	151	169	☑	
CSPG2_HUMAN	K.ETTVLVAQNG**N**(+.98)IK.I	80	92		☑
PTK7_HUMAN	R.DGTPLSDGQS**N**(+.98)HTVSSK.E	165	181		☑
HMGN2_HUMAN	K.EGNNPAE**N**(+.98)GDAKTDQAQKAEGAGDAK	65	90		☑
ALBU_HUMAN	K.YIC(+57.02)E**N**(+.98)QDSISSK.L	287	298		☑
ITB1_HUMAN	K.C(+57.02)HEG**N**(+.98)GTFEC(+57.02)GAC(+57.02)R.C	477	490		☑
AHNK_HUMAN	K.GDINIEGPSM(+15.99)NIEGPDL**N**(+.98)VEGPEGGLK.G	3501	3527	☑	