UNIVERSITEIT GENT

FACULTY OF PHARMACEUTICAL SCIENCES

Human embryonic stem cells:

From the follow-up of pluripotency to quantitative peptide analysis

Ellen Scheerlinck

Ghent University

Faculty of Pharmaceutical Sciences

Laboratory for Pharmaceutical Biotechnology

Promoter: Prof. Dr. Apr. Dieter Deforce

Promoter: Dr. Apr. Katleen Van Steendam

Promoter: Dr. Maarten Dhaenens

Thesis submitted in the fulfilment of the requirements for the degree of Doctor in Pharmaceutical Sciences.

Proefschrift voorgelegd tot het bekomen van de graad van Doctor in de Farmaceutische Wetenschappen.

Members of the jury

Prof. Dr. Apr. Dieter Deforce (promoter)

Laboratory for Pharmaceutical Biotechnology, Faculty of Pharmaceutical Sciences, UGent

Dr. Apr. Katleen Van Steendam (promoter)

Laboratory for Pharmaceutical Biotechnology, Faculty of Pharmaceutical Sciences, UGent

Dr. Maarten Dhaenens (promoter)

Laboratory for Pharmaceutical Biotechnology, Faculty of Pharmaceutical Sciences, UGent

Prof. Dr. Apr. Christophe Stove

Laboratory of Toxicology, Faculty of Pharmaceutical Sciences, UGent

Prof. Dr. Bart De Spiegeleer

Drug Quality and Registration, Faculty of Pharmaceutical Sciences, UGent

Prof. Dr. Björn Heindryckx

Ghent Fertility and Stem cell Team (G-FaST), Department for Reproductive Medicine, Ghent University Hospital

Prof. Dr. Jean-Paul Noben

Biomedical Research Institute (BIOMED), Faculty of Medicine and Life Sciences, UHasselt

Em. Prof. Dr. Joël Vandekerckhove

Department for Biochemistry, Faculty of Medicine and Health Sciences, UGent

This research was financed by GOA- project no. 01G01112 – Pathways to pluripotency and differentiation in embryos and embryonic stem cells.

“Energy and persistence conquer all things”

Benjamin Franklin

Table of contents

DANKWOORD **1**

LIST OF ABBREVIATIONS **5**

CHAPTER 1: INTRODUCTION **9**

Human embryonic stem cells (hESC) **11**

1. Characterization of hESC 12
   1.1 Transcription factors and cell surface antigens 12
   1.2 Signaling pathways 13
   1.3 Karyotype 15
2. Culture of hESC 15

Techniques used in hESC analysis **20**

1. Daily screening technique for hESC pluripotency 20
2. hESC and Proteomics 23
   2.1 Bottom-up experiment 26
      2.1.1 Sample preparation 26
      2.1.2 Liquid chromatography-tandem mass spectrometry analysis (LC-MS/MS) 29
      2.1.3 Data analysis 35
   2.2 Quantitative proteomics 37
      2.2.1 Relative quantification 38
      2.2.2 Absolute quantification 43

CHAPTER 2: AIMS & OVERVIEW **45**

CHAPTER 3: MONITORING OF PLURIPOTENCY IN HESC **49**

Abstract 51
Introduction 52
Material & Methods 53
Results 56
Discussion 66
Conclusion 69

CHAPTER 4: SAMPLE PREPARATION PROTOCOL FOR PROTEOMICS **71**

Abstract 73
Introduction 74
Material & Methods 76
Results 80
Discussion 84
Conclusion 87
Acknowledgements 87
Supplementary data 87

CHAPTER 5: SILAC OPTIMIZATION IN HESC 89

Abstract 91
Introduction 92
Material & Methods 94
Results & Discussion 100

CHAPTER 6: BROADER INTERNATIONAL CONTEXT 115

CHAPTER 7: CONCLUSION 125

CHAPTER 8: SUMMARY/SAMENVATTING 129

CHAPTER 9: REFERENCES 139

DANKWOORD

Een dankwoord schrijven, het laatste onderdeel van een doctoraat. Een reis vol uitdagingen die nu op zijn einde loopt. Een reis die ik niet zonder deze personen had kunnen ondernemen:

Mijn promotor Dieter: bedankt dat ik mocht beginnen op dit labo en dat jij mij de kans gaf om mijn eigen ideeën te ontwikkelen. Bedankt om mij hierbij bij te sturen wanneer dit nodig was. Ook bedankt voor de vele verwennerij: Nieuwjaarsetentje, team building event, Sinterklaas,…

Bedankt aan de examencommissie om mijn doctoraat grondig te lezen en te beoordelen.

Mijn collega's…toen ik begon, had ik "mijn" bureau nog op het gelijkvloers bij de IT. Bedankt Nicky om mij toen op te vangen en bedankt Filip voor je bureau destijds en de kritische inbreng eerst op mijn IWT en vervolgens op mijn artikels. Christophe, hopelijk vind je je job en de collega's aan de overkant even tof. Yannick, je snelle service bij een IT probleem was een wonder (waren al mijn aannemers voor mijn bouw maar zo efficiënt geweest). Dieter, bedankt voor de gesprekjes over den bouw en hopelijk geniet jij ook van je nieuwe thuis. Sander, jouw IT vaardigheden zijn zeker een aanwinst voor de massa spectrometry collega's. Ik zal onze spelnamiddagen, samen met de anderen, wel erg missen…

Voor de administratieve kant, kon ik altijd rekenen op Inge, Astrid en Nadine. Voor problemen met de massa of LC, was daar onze eigen Sofie! Bedankt Sofie voor alles wat je aan mij geleerd hebt. Ik kon mij geen betere lerares indenken. En met de massa, denk ik natuurlijk aan mijn mede massa spectrometry collega's: Katleen, Maarten, Paulien, Elisabeth en Laura. Bedankt Katleen en Maarten om mij te begeleiden tijdens mijn doctoraat, om mij in te tomen van "oh, dat kan ik ook nog doen". Ik vond het altijd leuk om ideeën en resultaten met jullie te bespreken en de constructieve commentaren van jullie beiden te gebruiken. En kijk…dit heeft toch maar geleid tot 3 artikels. Paulien, bedankt om met mij samen te werken en samen de problemen aan te pakken, eerst met de hESC cultuur samen met Liesbeth en vervolgens met de TripleTOF. Als ik het even niet meer wist, had jij wel een goed idee. Ook veel succes met je doctoraat, maar ik ben er zeker van dat dit prima zal verlopen! Elisabeth, jij bent nu de

volgende in rij om de Synapt te ontdekken. Je eerste artikel is intussen al gesubmit en er komen zeker nog anderen. Laura, ook jij bent al volop bezig met je doctoraat: hESC, massa spectrometry…in een korte tijd heb je veel bijgeleerd.

Voor de hESC cultuur, had ik natuurlijk ook nog veel hulp van onze derde musketier Liesbeth. Weet je nog de problemen die wij hadden toen we begonnen? Maar we zijn er dan toch maar in geslaagd om de hESC cultuur in het labo te implementeren. Veel succes in de VS! Bart, mijn overbuur in de bureau, altijd volop in de weer en ik kan maar hopen dat mijn presentatie voor de openbare verdediging even goed was als de jouwe!

Naast de research was er natuurlijk ook de routine! Bedankt David voor de vele verslagen en mailtjes hieromtrent, al dan niet even serieus. Al zou ik toch nog altijd graag mijn chocolade krijgen…je weet toch..je hebt mij zelf een brooddoos gegeven "liever eten, dan werken", dus ik moet mij daarna gedragen. Ik zal onze conversaties hieromtrent zeker en vast missen. Trees, eerst deel van de research en nu van de routine, blijf altijd je vrolijke zelf. Het was ook altijd een genoegen om van je brownies te proeven ☺. Saskia, Petra, Sabine, Sylvie, Eveline, Evelien en Leen: bedankt voor de gesprekjes op het labo, tijdens de teambuilding en Nieuwjaarsetentje.

Senne en Lieselot, onze vertegenwoordiging op het derde verdiep. Senne, bedankt om samen met Elisabeth als organiserend comité op te treden voor de leuke momenten achter de uurtjes en nog veel succes met de samenwerking met IMEC. Lieselot, mijn mede West-Vlaamse collega, jammer toch dat je niet proteomics gekozen hebt en veel succes met het overnemen van 1 van mijn labo-taakjes.

Een doctoraat is een tijd van komen en een tijd van gaan (voor de meesten dan toch onder ons). Mado, Sandra, Bert, Shahid, Marlies, Pieter, Veerle, Yens en Heleen. Uit het oog, maar zeker niet uit het hart! Bedankt voor de vele mooie momenten te samen.

Ook bedankt aan mijn thesisstudenten voor jullie enthousiasme en inzet.

Mijn dank gaat ook uit naar het G-FaST team. Bedankt voor jullie hulp met het opzetten van de hESC cultuur in ons labo.

Mama en papa, bedankt om mij altijd te steunen in wat ik ook onderneem. Jullie staan altijd voor mij klaar en het is steeds leuk om nog naar mijn oude thuis terug te keren (alhoewel Bergerac ook nog altijd de max is). Broertje, negen jaar ouder dan mij, wat wijst op meer ervaring zowel met de kids als met een job. Ik hoop dat ik binnen negen jaar ook al zo veel bereikt kan hebben als jou. Ondertussen getrouwd, en een nieuwe familie erbij. Bedankt voor al jullie steun en de vele kaarsjes! Voor jullie is dit reeds het tweede doctoraat (naast die van Koen) en ik kan alleen maar hopen dat jullie er evenveel van begrepen hebben als van bij hem! Bedankt ook aan mijn vriendinnen en vrienden om mij te steunen, jullie begrepen er hopelijk iets meer van ☺.

En natuurlijk mijn Koenie…eerst ontmoet tijdens Oldafolk waar ik direct voor je viel (letterlijk weliswaar). Maar toch wist je wel nog dat ik mijn IWT aan het indienen was, waardoor je mij een mailtje kon sturen voor succes. Alhoewel ik het IWT niet gehaald heb, ben jij mij altijd blijven steunen tijdens mijn doctoraat en heb je mijn problemen mogen aanhoren als de massa bv weer niet wilde werken…Ik kan geen woorden vinden om te zeggen wat je voor mij betekent. We hebben ondertussen al zoveel ondernomen: getrouwd, gebouwd en onze lieve Fien gekregen. Fientje, mijn lieve kleine meid (nu mag ik dit tenminste nog zeggen), je maakt het leven zoveel meer waard en laat de tijd nu nog sneller voorbij gaan. Je ontwikkelt je zo snel: toen ik begon met dit dankwoord kon je nog maar net zitten met 1 hand vasthoudend aan je park, maar nu begin je al te kruipen…mijn kleine lachebek...Ik ben blij om jouw mama te mogen zijn.

LIST OF ABBREVIATIONS

2D-LC	two-dimensional liquid chromatography
2D-PAGE	two-dimensional polyacrylamide gel electrophoresis
AA	amino acid
AMRT	accurate mass retention time alignment
AP	acid precipitation
APC	adenomatous polyposis coli protein
APEX	absolute protein expression
AUC	area under the curve
BA	N_{α}-benzoyl-L-arginine
bFGF	basic fibroblast growth factor
BMP4	bone morphogenetic protein 4
BIO	brominedirubin-3'-oxime
CE	collision energy
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
CID	collision induced dissociation
CM	conditioned medium
COFRADIC	combined fractional diagonal chromatography
Da	Dalton
DDA	data-dependent acquisition
DIA	data-independent acquisition
DMEM	Dulbecco's modified Eagle's medium
DNA	deoxyribonucleic acid
DPBS	Dulbecco's PBS
DTT	dithiotreitol
E	L-glutamate

E8	Essential 8
ECM	extra-cellular matrix proteins
EDTA	ethylenediaminetetraacetic acid
eGFP	enhanced green fluorescent protein
emPAI	exponentially modified PAI
ERK	extracellular signal-regulated kinase
ESI-Q-TOF	electrospray ionization-quadrupole-time of flight
ETD	electron transfer dissociation
FASP	filter aided sample preparation
FBS	fetal bovine serum
FC	flow cytometry
FDR	false discovery rate
FM	fluorescence microscopy
G-banding	Giemsa banding
$HDMS^E$	high definition MS^E
hESC	human embryonic stem cells
HILIC	hydrophilic interaction chromatography
HPLC	high performance liquid chromatography
IAM	iodoacetamide
ICAT	isotope coded affinity tag
ICM	inner cell mass
IMS	ion mobility separation
iPSC	induced pluripotent stem cells
IRES	internal ribosomal entry site
iTRAQ	isobaric tag for absolute and relative quantification
ITS-A	insulin-transferrin-sodium selenite-sodium pyruvate supplement
IVF	*in vitro* fertilization

K	L-lysine
KO-SR	KnockOut Serum Replacement
LC	liquid chromatography
LIF	leukemia inhibitory factor
m/z	mass to charge ratio
MALDI	matrix assisted laser desorption ionization
MEF	mouse embryonic fibroblasts
MEK	mitogen activated protein kinase kinase
MHC 1	major histocompatibility complex 1
MMTS	methyl methanethiosulfonate
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MS^E	elevated mass spectrometrys
MudPit	multidimensional protein identification technology
NA	numerical aperture
NEAA	non-essential amino acids
Neo	neomycine resistence gene
NO	nitric oxide
Nor-NOHA	N^{ω}-hydroxy-nor-L-arginine
NSAF	normalized spectral abundance factor
OCT4	octamer binding transcription factor 4
P	L-proline
PAI	protein abundance index
PBS	phosphate buffered saline
PEG	polyethylene glycol
PI3K	phosphatidylinositol-3-kinase
PT	two-phase solvent extraction

R	L-arginine
RA	retinoic acid
RP	reverse phase
SCX	strong cation exchange
SDC	sodium deoxycholate
SDS	sodium dodecyl sulphate
SILAC	stable isotopic labeling of amino acids in cell culture
SILAM	stable isotopic labeling by amino acids in mammals
SOX2	sex determining region Y-related high-mobility group box protein-2
SSEA	stage-specific embryonic antigen
SWATH	sequential window acquisition of all theoretical fragment ion spectra
TCEP	tris(2-carboxyethyl)phosphine
TeABC	triethylammonium bicarbonate
TFA	trifluoroacetic acid
TGFβ	transforming growth factor β
TMT	tandem mass tags
TRA	keratan sulfate antigen
T-wave	traveling wave technology
UCM	unconditioned medium
$UDMS^E$	ultradefinition MS^E
UV	ultraviolet
VN	vitronectin

CHAPTER 1: INTRODUCTION

Human embryonic stem cells (hESC)

Techniques used in hESC analysis

Human embryonic stem cells (hESC)

Human embryonic stem cells are mostly derived from the inner cell mass (ICM) of a 64-cell stage embryo, called a blastocyst. This blastocyst, consisting of trophectoderm (outer layer) and ICM, is generally derived by means of *in vitro* fertilization (IVF) and donated to science after informed consent (1). hESC arise from this ICM after different intermediate states (2).

hESC distinguish themselves from other cells by two properties: (1) self-renewal and (2) pluripotency (1). This capability to self-renew is due to telomerase. This reverse transcriptase enzyme elongates the 3'end of DNA strands in the telomere region by adding TTAGGG preventing chromosomal damage during DNA replication (3, 4). Pluripotency is the ability to differentiate to cells belonging to the three primary germ cell lines: endoderm (respiratory and gastrointestinal tracts), ectoderm (epidermis and nervous tissue) and mesoderm (muscle, bone, blood & cartilage) (4). Upon differentiation, stem cells start to display different grades of differentiation, making it possible to group them further (5, 6) (**Figure 1**).

Less cell type possibilities for differentiation. ↓

Grade of differentiation	Example
Totipotent	embryonic stem cells derived from a zygote/morula stage → differentiation to all cells
Pluripotent	embryonic stem cells derived from a blastocyst → differentiation to all cells of primary germ cell lines
Multipotent	Hematopoëtic stem cells → differentiation to blood cells
Oligopotent	Lymphoid stem cells → differentiation to cells of the lymphoid system
Unipotent	Muscle stem cells → differentiation to cells of the muscle

Figure 1. Overview of the different grades of differentiation in stem cells.

These characteristics make hESC promising for use in different applications in the future: new medicines (toxicity screening), studies regarding embryogenesis and regenerative therapy (organ transplantations) (1). Although some results in the field of regenerative therapy are promising, for example the *in vitro* differentiation to cardiomyocytes (7), many challenges remain. In particular, the large amount of cells needed for differentiation and the problem of immunological rejection due to a different expression of major histocompatibility complex I (MHC I) antigens are major hurdles (8). A solution to the latter could be the generation of induced pluripotent stem cells (iPSC), as developed by Takahashi & Yamanaka (2006) (9). These cells are generated out of the patient's own somatic cells avoiding the graft versus host problem. iPSC are made by transfecting the somatic cells with a vector system in which a virus contains the genes (*OCT4, SOX2, KLF4* and *c-MYC/n-MYC* or *OCT4, SOX2, LIN28* and *NANOG*) to obtain pluripotent cells (10). The efficiency of cellular programming can be enhanced by the use of small molecules for example vitamin C or valproic acid (11). While promising, generation of iPSC gives rise to alterations in the DNA (DNA methylation amongst others), resulting in a higher risk of tumor development (12, 13).

1. Characterization of hESC

1.1 Transcription factors and cell surface antigens

The transcription factor OCT4 (encoded by *OCT4* or synonyms *POU5F1* and *OCT3*, class V of the POU family), discovered in 1991, is expressed in pluripotent stem cells (1). OCT4 can activate and inhibit a wide range of genes related to embryo development: OCT4 activates amongst others the genes Fgf4, Opn, Rex1 and Utf1 (14-16). In hESC, the levels (low or high) of OCT4 in combination with the possible activation of the bone morphogenetic protein 4 (BMP4) pathway regulates a different developmental state. High levels of OCT4 in the presence or absence of BMP4 promotes respectively mesendodermal differentiation and non-differentiation. Low levels of OCT4 in the presence or absence of BMP4 promotes respectively extraembryonic lineage and primitive endoderm differentiation and ectoderm differentiation (17). Two other important pluripotency transcription factors are SRY-related high-

mobility group box protein-2 (SOX2) and NANOG (called after Tir nan Og) (14, 18). In general, it is reported in hESC that a decrease in SOX2 and NANOG expression results in differentiation to cells positive for markers of primitive endoderm and trophectoderm (14, 17, 19). These genes can quantitatively be determined by means of real-time PCR and microarray platforms (20, 21). hESC differentiation on the other hand is characterized by increased expression of GATA4/GATA6 for endodermal differentiation and GATA2/CDX2 for trophectodermal differentiation, amongst others (14).

In addition, cell surface antigens related to hESC pluripotency can be determined by immunostaining in combination with fluorescence microscopy or flow cytometry. Some examples are stage-specific embryonic antigen (SSEA-3 & SSEA-4) and keratan sulfate antigen (TRA-1-60 & TRA-1-81) (4, 22). Of note, high alkaline phosphatase levels are also a marker of undifferentiated hESC (23).

1.2 Signaling pathways

Four main pathways involved in pluripotency can be distinguished: (1) transforming growth factor β (TGFβ)-activin-NODAL pathway, (2) phosphatidylinositol 3-kinase (PI3K) pathway, (3) Ras-Raf-MEK (mitogen activated protein kinase kinase)-ERK (extracellular signal-regulated kinase) pathway or MAPK/ERK pathway and (4) the WNT pathway (14, 24-26) (**Figure 2**). Addition of activin A to a hESC culture activates the TGFβ-activin-NODAL pathway through the signal transducer SMAD2/3 (24). Addition of fibroblast growth factor on the other hand activates the PI3K pathway and the Ras-Raf-MEK-ERK pathway. In short, for the PI3K pathway, phosphatidylinositol 4,5-bisphosphate becomes phosphorylated by means of PI3K during activation. This phosphatidylinositol 3,4,5-triphosphate binds with Akt (also known as protein kinase B). Activation of this pathway results in increased concentrations of OCT4, NANOG and SOX2 and consequently maintenance of pluripotency (26, 27). On the other hand, activation of Ras-Raf-MEK-ERK pathway leads to the activation of Ras (a GTPase), which in turn binds with Raf. This kinase phosphorylates another kinase, MEK, which phosphorylates ERK. ERK translocates to the nucleus leading to the phosphorylation of c-Myc, c-Jun and c-Fos (28, 29). The WNT signaling pathway is activated by 6-brominedirubin-3'-oxime (BIO). This

agent inhibits glycogensynthase kinase-3 which normally promotes the degradation of β-catenin in a complex with Axin and APC (adenomatous polyposis coli protein) by making it a target for the proteasome (29, 30). As a consequence, β-catenin accumulates in the cytoplasm and a portion of this pool translocates to the cell nucleus and interacts with genes important for keeping hESC undifferentiated (25).

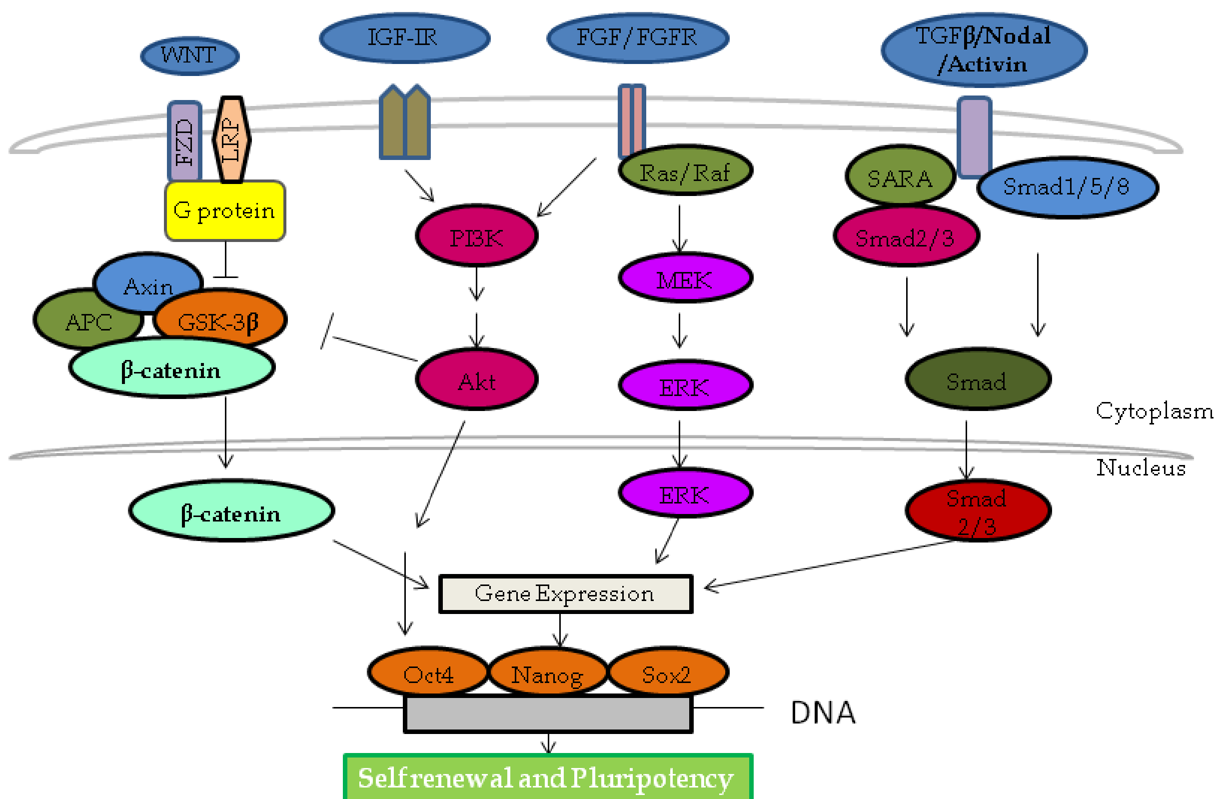


Figure 2. Signaling pathways regulating pluripotency. (Image taken from (31))

Some important differences exist between mouse and human embryonic stem cells. While the leukemia inhibitory factor (LIF)-STAT3 pathway is required for maintaining a pluripotent state only in mouse embryonic stem cells, BMP4 causes differentiation to trophoblast only in hESC (through the signal transducer SMAD1/5/8) (14, 32). This is in contrast to mouse embryonic stem cells, in which the combination BMP4 with LIF helps in maintaining the pluripotent state by suppressing the neural differentiation (14, 32).

1.3 Karyotype

hESC need to be karyotyped because of genomic alterations (gain of chromosomes 12 and 17) that arise during culture. Karyotypical changes can have an impact on self-renewal and proliferation (33). The detection of chromosomal aberrations is routinely performed by Giemsa banding (G-banding) (33). Newer techniques are spectral karyotyping or multicolor-fluorescence *in situ* hybridization and digital karyotyping (34, 35). In spectral karyotyping, the chromosomes are visualized using fluorescence. This fluorescence is obtained by using fluorophore labeled single-stranded DNA (= probe) specific for a region of a chromosome. After hybridization with the chromosome, spectral differences can be observed. The main advantage in comparison with G banding is that no specialized training is needed to observe the differences in a chromosome (34). Digital karyotyping (Single Nucleotide Polymorphism arrays or array comparative genome hybridization) is a quantitative-based analysis of short sequences of DNA. In this method, DNA sequences specific for a region on the chromosome are used on a microarray plate. In short, DNA of the sample and of the control are labeled with a different fluorophore. After hybridization on a microarray plate, analysis can be obtained. A disadvantage of this technique is that it cannot be used to detect aberrations that do not result in copy number changes (for example inversion) (35, 36).

Additionally, epigenetic changes (DNA methylation-histone tail modifications) arise, but these are not as of yet routinely characterized.

2. Culture of hESC

Thomson et al. (1998) were the first to culture hESC on mouse embryonic fibroblasts (MEF) with hESC medium (4). This culture method, named feeder culture, uses replication deficient fibroblasts (treated with mitomycin C/γ-irradiation) for hESC attachment (coating) (4). Attachment is possible because of the expression of different extra-cellular matrix proteins (ECM) at the cellular membrane of the fibroblasts. A few examples are the expression of collagen IV and laminin (37). It is believed that the main interaction between ECM and hESC is due to the binding to specific integrines (for example α6β1)

present at the cell surface of hESC (8, 38). Besides attachment, these cells are also producing factors needed to keep hESC undifferentiated and are for this reason called feeder cells.

Feeder cell culture of hESC comes with several difficulties: high work load because of the culture of feeder cells, lot-variability, undefined culture and the use of animal products that would result in immunological rejections and risk on zoonosis, viruses and prions when later applied in the clinic (1, 39). For this reason, the search for new feeder-free culture models was explored. A feeder-free culture combines a coating for hESC attachment and a specific medium which contains pluripotency promoting factors (37). One of the first coatings available was Matrigel®. Matrigel® is a combination of different ECM proteins (laminin, collagen IV, entactin and heparan sulphate proteoglycan) extracted out of Engelbreth-Holm-Swarm mouse sarcoma cells (1). Although successful in hESC attachment, Matrigel® has different disadvantages such as animal derived, lot-variability and undefined composition. Other defined coatings were thus developed: vitronectin, collagen IV, fibronectin and laminin. The potential of each coating to keep hESC pluripotent depends on the hESC medium it is combined with. The first hESC medium used was conditioned medium (CM) from fibroblasts (mostly derived from MEF called MEF CM). MEF CM is produced by incubating inactivated MEF for 24 hours in hESC medium consisting out of 80 % Knockout-Dulbecco's modified Eagle's medium, 20 % KnockOut Serum Replacement (KO-SR), 1 mM L-glutamine, 0.1 mM β-mercaptoethanol, 1 % non-essential amino acids (NEAA), and 4 ng/ml human basic fibroblast growth factor. In combination with Matrigel® or laminin, hESC were kept pluripotent for at least 130 population doublings in this way (8). MEF CM is however still undefined, lot-variable and of animal origin. Proteomic analysis of this medium to determine pluripotency promoting factors resulted in a list of proteins (37, 40). Chin et al. (2007) tested 6 growth factors identified in MEF CM to keep hESC undifferentiated, but unfortunately hESC were not viable after 3 passages (3, 40).

Meanwhile, molecular studies broadened our knowledge on how pluripotency is maintained. Out of this knowledge, new specific culture conditions were developed. An overview of some hESC culture conditions can be found in **Table 1**. Although some of these culture conditions were indeed successful

in keeping hESC pluripotent, a disadvantage was the presence of undefined factors in the media (HESCO) or coatings (Matrigel®) (39, 41). In 2011, Chen et al. developed a fully defined, xeno-free and albumin-free hESC culture condition consisting of vitronectin (coating) in combination with Essential8 (E8) as medium. E8 medium contains the following components: Dulbecco's modified Eagle's medium/F12 (basal medium) with insulin (19.4 mg/L), fibroblast growth factor 2 (100 µg/L), L-ascorbic acid-2-phosphate magnesium (64 mg/L), sodium selenium (14 µg/L), transferrin (10.7 mg/L), sodium bicarbonate (543 mg/L) and NODAL (100 µg/L) or transforming growth factor beta (2 µg/L) (42). This culture system was further used in this thesis.

hESC cultures need to be passaged every 3-5 days to expand and to avoid differentiation and cell overgrowth. Different dissociative solutions and methods are available to passage the cells. T'Joen et al. (2011) have compared 4 dissociative solutions (TrypLE™ Express, Trypsin-EDTA, Cell Dissociation Buffer and Accutase) with manual and bead-based passaging techniques. The use of Cell Dissociation Buffer scored the best in colony amount, produces stable expanding hESC lines and hESC remained pluripotent (43). Although the full composition is company confidential, cell dissociation buffer contains ethylenediaminetetraacetic acid (EDTA) and is prepared in phosphate buffered saline (PBS) without calcium and magnesium (44). At our lab, a similar buffer is used (0.5 mM EDTA prepared in DulbeccoPBS) as dissociative solution, as recommended by Thermo Scientific.

Table 1. Overview of some hESC culture conditions.

Coating	Medium	Number of passages	Reference
Matrigel	mTeR1	>30 passages	(45, 46)
	MEF-CM	88 passages	(47)
Laminine	MEF-CM	10 passages	(38, 48)
Gelatine	MEF-CM	<1 passage	(8)
		20 passages	(49)
Cellstart	StemPro	30 passages	(46)
Fibronectine	HESCO	8 passages	(39)
	MEF-CM	24 passages	(47)
Vitronectin	mTeSR1	30 passages	(46)
	E8 medium	25 passages	(42)

In contrast to hESC, mESC have a different culture condition. As mentioned earlier, other pathways are necessary to maintain pluripotency and self-renewal in mESC. mESC can be kept in culture on inactivated mouse embryonic fibroblasts or gelatin-coated flasks in combination with LIF (activates LIF/STAT3) and serum as growth factors (50, 51). BMP4 can substitute for serum and a fully defined medium can be obtained in this way (50). Instead of LIF and BMP4, one can also add 2i to the culture medium. 2i is a combination of the small molecule inhibitors named glycogen synthase kinase 3 (CHIR99021) and extracellular regulated kinase (PD0325901). CHIR99021 enhances embryonic stem cell growth capacity and viability, while PD0325901 inhibits the MAPK/ERK pathway, which is responsible for differentiation in mESC (52). This is therefore in contrast to hESC in which activation of the MAPK/ERK pathway by means of bFGF is needed for keeping hESC undifferentiated (27). The addition of LIF to this 2i medium enhances cloning efficiency (50).

Of note, two kinds of pluripotent stem cells can be found in mouse: mESC and mouse epiblast stem cells (mEpiSC). The differences between both cells are summarized in **Table 2** (53)**.** As can be observed, a different pluripotent state exists between mESC and mEpiSC. This primed form reflects a more developed form of pluripotency because not all cells can be formed out of this pluripotent state. For example, Honda et al. (2013) showed that naïve rabbit induced pluripotent stem cells could differentiate towards oligodendrocytes unlike their primed ones (54). It is clear that naïve stem cells have some advantages above their primed ones (single cell clonogenicity and homogeneous cell population) and it seems that in some cases naïve stem cells are better for clinical applications. For example, Jang et al. (2014) observed that neural stem cells from naïve mESC are more similar to bona fide neural stem cells in comparison with mEpiSC (55). For this reason, the derivation of naïve embryonic stem cells can be of high interest for clinical approaches. mEpiSC could be converted to mESC by exposure to the LIF/STAT3 pathway and improved by expression of *KLF4, KLF2, NANOG* or *c-MYC* (56). In addition, the characteristics of hESC show that these cells resembles more the primed state of pluripotency. In analog to the conversion of mEpiSC, primed hESC can be converted to naïve hESC by using the same ectopic forced expression of these markers in combination with LIF and 2i. Of note, forskolin which induces *KLF4* and *KLF2* expression can be used to transiently substitute for the ectopic transgene expression (56). Other ways to form naïve human embryonic stem cells are for example (1) directly from preimplantation embryo's using 2i and bFGF and (2) by conversion of existing primed hESC by preculture in the histone deacetylase inhibitors butyrate and suberoylanilide hydroxamic acid, followed by culture in 2i with bFGF (57).

Table 2 Overview mESC versus mEpiSC. Differences between some of the characteristics between mESC and mEpiSC are presented.

	mESC	**mEpiSC**
Pluripotent state	Naïve	Primed
Morphology	Domed	Flattened
Corresponding *in vivo* tissue	Early epiblast (pre-implantation)	Epiblast (peri-/post-implantation)
Single Cell clonogenicity	Yes	No (trypsin intolerant)
Cell population	Homogeneous	Heterogeneous
Chimerism	High efficiency	Very low efficiency
Female X inactivation	XaXa	XaXi
Genes expressed	*OCT4, NANOG, SOX2, KLF2, KLF4, KLF5, ZPF42, DPPA3, FGF4*	*OCT4, NANOG, SOX2, DMNT3B, FGF5, MEIS1, SOX11*
Global DNA methylation	Hypomethylated	Hypermethylated

Techniques used in hESC analysis

In this dissertation, some techniques that can be used in culturing and studying hESC were refined. More specifically the daily monitoring of the pluripotency of hESC by measurement of the expression of OCT4 and the use of mass spectrometry to analyze the hESC proteome. More in particular, the sample preparation for label-free quantitation and Stable Isotopic Labeling by Amino acids in Cell culture (SILAC) for protein quantification was optimized.

1. Daily screening technique for hESC pluripotency

As described above, hESC can be monitored by either the expression of genes specific for pluripotency (OCT4, NANOG, SOX2) or for differentiation (SSEA-1) (18). Different techniques (real-time PCR, immunostaining, flow cytometry) can be used to that end, but have the disadvantage that the same cells cannot be kept in culture afterwards (destructive techniques).

In this dissertation, a screening method was developed for daily monitoring of pluripotency (**Chapter 3**) with continuous hESC culture. This screening method includes the use of a commercially available WA01 OCT4-enhanced green fluorescent protein (eGFP) Knock-In hESC line (Wicell Research Institute, Madison, WI, USA) coupled to fluorescence microscopy.

eGFP is an auto-fluorescent protein which absorbs light at 489 nm and emits at higher wavelength (509 nm) (Stokes shift) (58, 59). This fluorescence can be measured with fluorescence microscopy (60).

In this case, eGFP is used as a reporter molecule to monitor gene expression. The eGFP-coding sequence is placed under the transcriptional control of the promotor of OCT4, marker of pluripotency, giving a directly visible readout of the gene's expression in a hESC culture. High eGFP intensity means high OCT4 expression, indicating pluripotency. This reporter hESC line was developed by Zwaka & Thomson (2003) by means of homologous recombination (**Figure 3**) (61). In short, a targeting vector (**Figure 3A**) is constructed by insertion of an internal ribosomal entry site (IRES)-eGFP and an IRES-neomycine resistence gene (neo) into the 3' untranslated region of the fifth exon of the human *POU5F1* gene (encoding OCT4). This cassette is flanked with 2 homologous arms. After electroporation, both neomycine resistance and flow cytometry to detect eGFP expression was used to detect cells which contain the construct (**Figure 3B**). Afterwards, differentiation was induced by growing hESC on Matrigel® with unconditioned medium. This validation of the reporter hESC line was examined with flow cytometry in which a loss of eGFP expression was observed after 5 days of differentiation (**Figure 3C**).

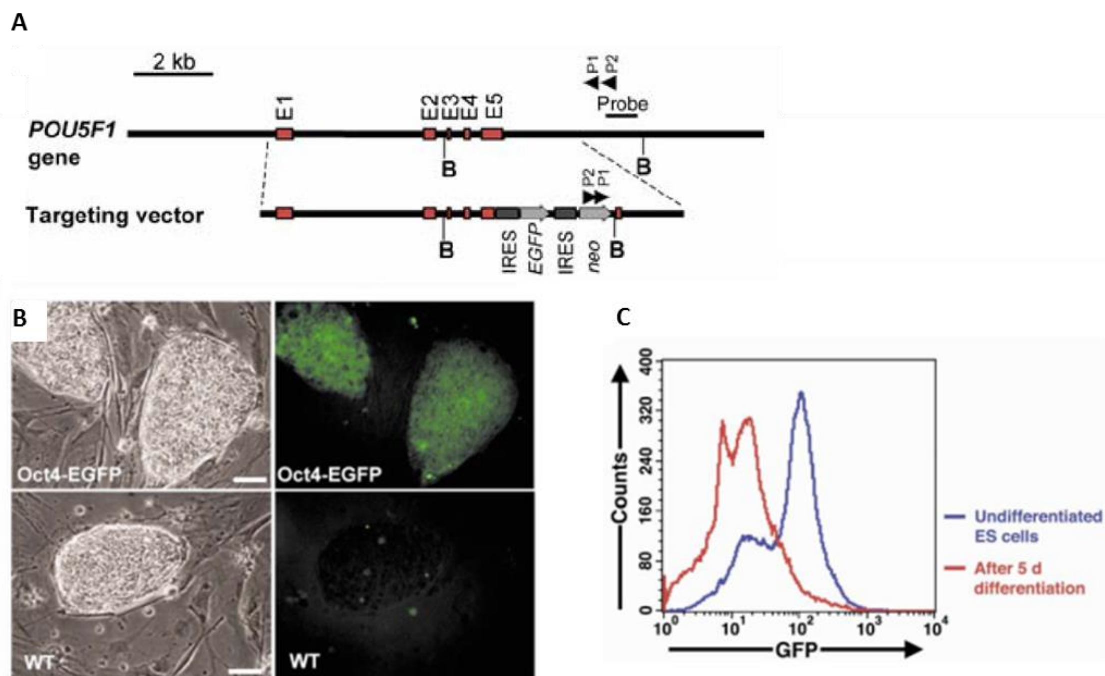


Figure 3. Targeting of an IRES-eGFP-IRES-*neo* cassette into the 3'UTR of the gene *OCT4*. IRES = internal ribosomal entry site, eGFP = enhanced green fluorescent protein, neo = neomycine resistence gene, UTR = untranslated region (A) Partial structure of the human *OCT4* (=*POU5F1*) gene and the gene-targeting vector. E, exon (B) Fluorescence microscopy (right) and phase-contrast microscopy (left) of OCT4 knock-in and wild type colonies. Bar, 25µm. (C) Flow cytometry of OCT4 knock-in undifferentiated (EGPF-positive) hESC (blue) and their differentiated form after 5 days of differentiation (red). (Image taken from (61))

Flow cytometry is a laser-based technology that can be used for cell analysis (cell counting, analysis of specific markers by means of fluorescence, cell characteristics (size and internal cellular complexity). It is important that cells are passed through the laser beam one at a time to avoid incorrect results. This is obtained by hydrodynamic focusing in which the sheath fluid draws the sample fluid into the stream passing a small aperture. The laser beam falls subsequently onto the cell and the light will be scattered. The forward scatter gives information about the size of the cell, while side scatter (light detected 90 ° relative to the laser beam) gives information about the granularity or complexity of the cell. Cells of interest can be selected in the 2D scatter plot by gating of the population. Fluorescence of eGFP can simultaneously be measured by the detector (**Figure 4**) (62). Of note, cells need to be in suspension in

an isotonic solution. hESC, which grow in colonies, are as a consequence treated with 0.25 % trypsin-EDTA to obtain single cells and suspended in PBS before analysis occurs.

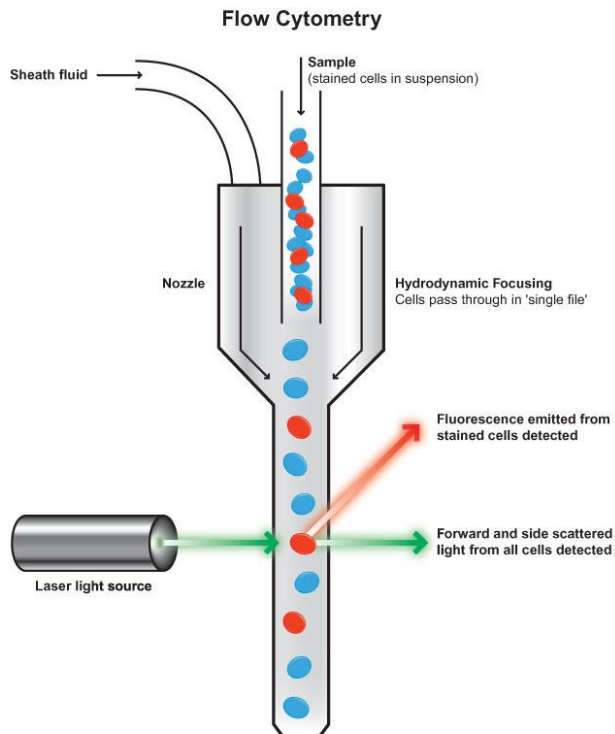


Figure 4. Flow Cytometry. Cells in suspension are focused to pass one by one through the laser by means of hydrodynamic focusing. The light that falls onto the cell is scattered. Forward scatter and side scatter of the light gives information of respectively size and complexity of the cell. By means of staining of a cell or by using a reporter cell line, fluorescence emitted by these cells can also be measured. (Image taken from (63)

Of note, other reporter stem cell lines are also available (hESC lines containing a hREX-GFP construct or NANOG-eGFP construct) (64, 65).

2. hESC and Proteomics

Proteomics determines the gene and cellular function, directly at the protein level (66). The first proteomic dataset of undifferentiated hESC was provided in 2006 (67). Several of these proteins could be hypothesized to characterize stemness-specific proteins such as hepatoma-derived factor and cellular retinoic acid-binding protein 1. During time, a lot of proteomic analysis on hESC was performed in

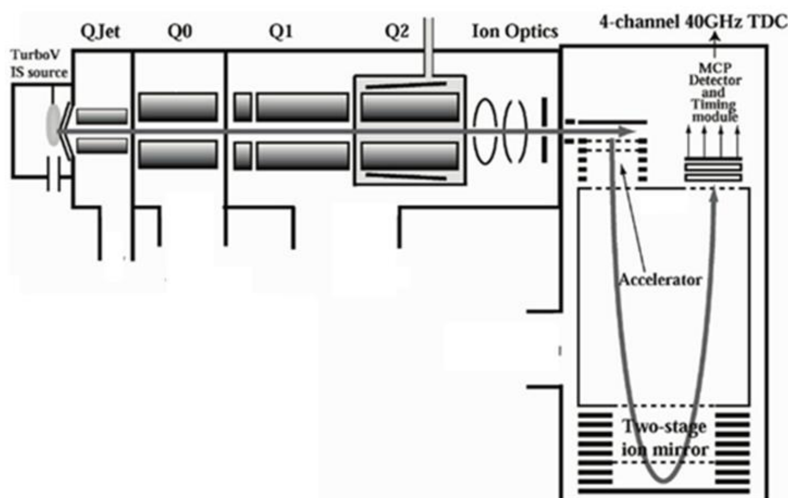
which for example OCT4 and alkaline phosphatase were identified in pluripotent hESC (6). Proteomics is thus a valuable tool to study the biology of hESC and to enlarge our knowledge about these cells.

When complete analysis of the proteome of cells is required, mass-spectrometry is currently the method of choice.

A mass-spectrometer is built out of (1) an ionization source, (2) a mass analyzer (measures the mass-to-charge (*m/z*) ratio of the ionized analytes) and (3) a detector (66). In this thesis two electrospray ionization-quadrupole-Time of Flight (ESI-Q-TOF) mass spectrometers were used: TripleTOF 5600 (Sciex; (68)) and Synapt G2Si (Waters, (69)) (**Figure 5**). Of note, other mass spectrometers are commercially available and more information can be found about some of these mass spectrometers in the following article (70).

An ESI-Q-TOF mass spectrometer uses ESI as ionization source and quadrupole and time-of-flight as mass analyzers. ESI, rather than Matrix Assisted Laser Desorption Ionization (MALDI) is mostly used as ionization source in shotgun approaches, because ESI can be coupled to liquid chromatography (LC), which is needed for the separation of the complex peptide mixture (66). A quadrupole is built out of four cylindrical rods in which a specific mass to charge ratio (*m/z*) can be selected by means of an electric field that is applied to the rods. In this way, a quadrupole can work as an ion guide (no selection of specific m/z, regulating ion in-flux), a filter (selection of specific *m/z*) and as a collision cell (fragmenting the peptide precursor ions by means of a collision gas). In the Synapt G2Si, the quadrupole technology is supplemented by Traveling Wave Technology (T-wave) which makes it possible to perform ion-mobility separation (IMS) (70). This ion-mobility is used as an extra separation step, which separates peaks according to their collisional cross section specified by the charge, size and shape of the peptide (71). Peptide ions with higher charge states experience a higher electric field resulting in a higher drift velocity resulting in a lower drift time. Finally, a time-of-flight makes use of the fact that each ion has its own kinetic energy depending on its *m/z* value (low-mass peptides reach the detector earlier) and determines in this way the ion's *m/z*.

TripleTOF 5600 (Sciex)



Synapt G2-Si (Waters)

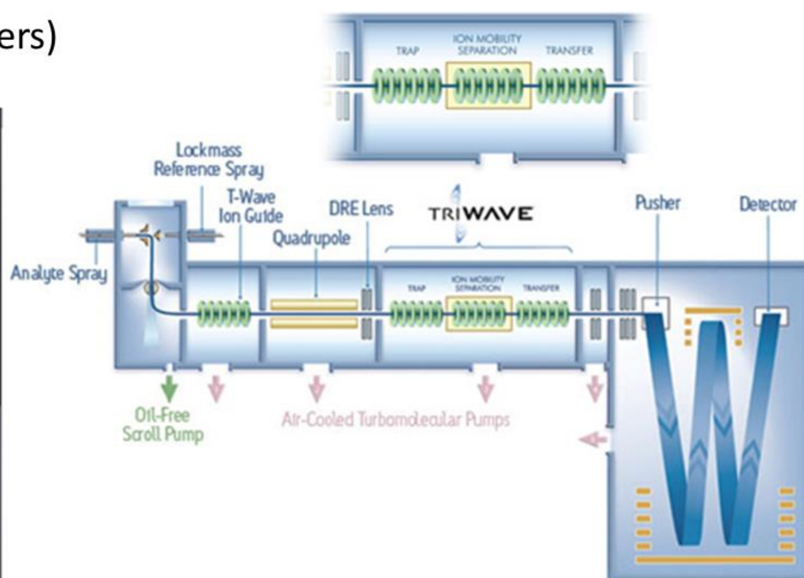


Figure 5. Representation of the mass spectrometers available in the Laboratory of Pharmaceutical Biotechnology: TripleTOF 5600 (Sciex) and Synapt G2Si (Waters).

Several approaches exist to analyze proteins. Complete proteins can be analyzed in top-down approaches, as opposed to analysis of peptides by bottom-up approaches or larger peptide sequences as is done in middle-down proteomics (72). In this thesis, a bottom-up approach (or shotgun approach) was used to analyze a complex proteomic sample. A bottom-up experiment can in general be described in 3

steps: (1) sample preparation (proteins → peptides), (2) LC-tandem mass spectrometry analysis (LC-MS/MS) and (3) data analysis.

In the next paragraph, the several steps of a bottom-up experiment are described in more detail.

2.1 Bottom-up experiment

2.1.1 Sample preparation

The first step of a proteomics experiment can be cell lysis. Different surfactants and physical techniques (sonication, freeze-thaw) are available to lyse cells (73). The choice of surfactant will depend on the goal of the experiment: analysis of a complete cell proteome asks for an ionic detergent which breaks protein-protein interactions (for example sodium dodecyl sulphate (SDS)), while analysis of protein interactions requires a mild non-ionic detergent like Triton X-100, which keeps the protein-protein interactions intact (74). When cells are lysed, nucleic acids and proteases are released. An endonuclease (e.g. benzonase) is added to the cell lysis buffer to hydrolyze these nucleic acids, allowing a reduction of viscosity and improvement of the analysis of transcription factors. Proteins are protected against degradation by the addition of several protease (serine and cysteine proteases) and phosphatase inhibitors (serine and threonine phosphatases). This addition is especially essential when analyzing post-translational modifications (72, 73).

After lysis, proteins are either (1) directly cleaved into peptides (in-solution digest) or (2) fractioned e.g. by their molecular weight or isoelectric point, or by specific characteristics such as the presence of certain post-translational modifications (75). Fractionation is performed if the sample is too complex or if only a certain protein group (e.g. histones) is of interest. Two dimensional polyacrylamide gel electrophoresis (2D-PAGE) is a popular method to fractionate proteins.

After fractionation (if needed) and protein quantitation, proteins are cleaved into peptides by digestion. For an optimal enzyme digest, the protein sequence has to be accessible to the digesting enzyme. The protein must therefore be (1) denatured and (2) disulphide bridges between cysteines need to be broken.

Several denaturants are nowadays available for complete denaturation, each with its own disadvantages (**Table 3**). As can be seen in **Table 3**, denaturants can have an inhibitory effect on the activity of the digesting enzyme (in this case trypsin). This inhibition is monitored by means of the formation of the UV active N_α-benzoyl-L-arginine (BA), the digestion product of BA ethyl ester (76). LC-MS/MS compatibility also needs to be considered. SDS for example can bind to the reverse phase column and acts as an ion-exchanger. In addition, SDS causes ion suppression resulting in non-ionized peptides which cannot be analyzed (77, 78). Finally, addition of a denaturant can induce modifications. Urea degrades to cyanate under high temperature, causing carbamylation of free amines (79). This modification has to be taken into account during data analysis. In this dissertation, the use of sodium deoxycholate (SDC) during cell lysis and in-solution digest was optimized (**Chapter 4**).

Subsequently, disulphide bridges present on cysteine-containing proteins are broken by first reducing the sample, most frequently by means of 10 mM dithiotreitol (DTT) or 5 mM tris(2-carboxyethyl)phosphine) (TCEP). After reduction, the sample is alkylated to prevent the reformation of disulphide bridges. This is performed most often by the addition of 10 mM methyl methanethiosulfonate (MMTS) or 20 mM iodoacetamide (IAM). Of note, IAM can generate non-specific N and O-alkylation (80).

Table 3. Overview of some of the most used denaturants during digestion. SDS = sodium dodecyl sulphate; SDC = sodium deoxycholate. The effect on trypsin activity, compatibility with LC-MS/MS and peptide modifications of each denaturant is mentioned.

Denaturant	% trypsin activity (BA absorbance)	Compatibility LC-MS/MS	Inducing additional modifications	Reference
0.1 %/0.5 % Rapigest	100 %/87 %	Removal necessary (acid)	No	(81-83)
0.2 % ProteaseMAX	100 %	Yes (degradable by trypsin)	No	(81)
1 M/2 M Urea	97 %/83 %	Yes	Carbamylation of lysine and N termini (high °C)	(82)
0.1 %/0.5 % SDS	20 %/1 %	Ion suppression	No	(76)
1 %/2 % SDC	95.4 %/86.4 %	Removal necessary (acid)	No	(76)

Finally, the protease for digestion needs to be chosen. An overview of the different enzymes is shown in **Table 4**. Trypsin cleaves C-terminal to arginine and lysine, making that every peptide is at least 2+ (3+ charge is due to histidine in the sequence (84)). In this way, the peptide is easily ionized and trypsin is for this reason the most used protease. Of note, more protein identification could be obtained when the same sample is digested with more enzymes for example the combination of trypsin and GluC (72).

In addition, sample clean-up is sometimes necessary: high salt concentrations can interfere with the ionization process of ESI, making sample clean-up with for example C_{18} ZipTips mandatory in some cases (85). Another issue is the presence of contaminants or impurities in the sample. These contaminants can be observed in the mass spectrum and are for example polyethylene glycol (PEG) or protein-related contaminants (e.g. keratins). Polyethylene glycol, a polymer built out of repeating units of ($-CH_2-CH_2-O-$), is present in some detergents (e.g. Triton X) or protease inhibitor cocktail tablets. This polymer leads to ion suppression of the sample during ESI and can be observed in the mass spectrum by a repeating train of peaks of 44 Da. Keratins, originating from clothing and human skin, are often present in the background. These proteins can be avoided by working with latex-free gloves in a laminar air flow closet (86, 87).

Table 4. Common Proteases Used for Shotgun Proteomics. (Image taken from (72))

Protease	Cleavage specificity[a]	Common proteomic usage
Trypsin	-K,R-↑-Z- not -K,R-↑-P-	General protein digestion
Endoproteinase Lys-C	-K-↑-Z-	Alternative to trypsin for increased peptide length; multiple protease digestion
Chymotrypsin	-W,F,Y-↑-Z and –L,M,A,D,E-↑-Z- at a slower rate	Multiple protease digestion
Subtilisin	broad specificity to native and denatured proteins	Multiple protease digestion
Elastase	-B-↑-Z-	Multiple protease digestion
Endoproteinase Lys-N	-Z-↑-K-	Increase peptide length; create higher charge state for electron transfer dissociation (ETD, 2.1.2.)
Endoproteinase Glu-C	-E-↑-Z- and 3000 times slower at –D-↑-Z-	Multiple protease digestion
Endoproteinase Arg-C	-R-↑-Z-	Multiple protease digestion
Endoproteinase Asp-N	-Z-↑-D- and Z-↑-cystic acid-but not –Z-↑-C-	Multiple protease digestion
Proteinase K	-X-↑-Y-	Nonspecific digestion of membrane-bound proteins
OmpT	-K,R-↑-K,R-	Increase peptide length for middle-down proteomics

[a] B = uncharged, nonaromatic amino acids (i.e, A,V, L, I, G, S); X = aliphatic, aromatic, or hydrophobic amino acids; and Z = any amino acid.

2.1.2 Liquid chromatography-tandem mass spectrometry analysis (LC-MS/MS)

A cell lysate digest is too complex to analyze it with mass spectrometry alone. Therefore, peptides can be separated by using several techniques such as high performance liquid chromatography (HPLC). One of the most used methods is a HPLC system with a reverse C_{18} column as stationary phase (RP). By using a gradient with increasing organic content through time (=mobile phase), peptides are separated based on their hydrophobic, ionic and polar interactions with the stationary and mobile phase (88). A higher separation grade of a cell lysate, and thus reduced sample complexity, is possible with two-dimensional liquid chromatography (2D-LC) (84). Peptides are in this technique separated in "two dimensions" based on two of their physical properties. Gilar and al. (2005) investigated a selectivity of

common LC modes and identified systems with useful orthogonality. Suitable orthogonality was observed by the combination of Strong Cation Exchange (SCX)-RP, hydrophilic interaction chromatography (HILIC)-RP and high pH – low pH RP-RP (84). Such an approach to reduce sample complexity is called multidimensional protein identification technology (MudPit) (89). Another approach is the isolation of a set of representative peptides. Only these peptides are analyzed to attain reduced sample complexity and increased proteome coverage. An example of this approach is COmbined FRActional DIagonal Chromatography (COFRADIC), which can be applied to isolate methionyl peptides and cysteinyl peptides, amongst others. In this technique, the sample is first chromatographically separated into fractions. Secondly, the sample is chemically or enzymatically modified on peptides containing rare amino acids, resulting in a hydrophilic or hydrophobic shift of the representative peptides. Some of the primary fractions are pooled and acquired. During the second run, the modified peptides have a shift in retention time in comparison with the first run and these peptides are analyzed with mass spectrometry (90).

After separation, the peptides are analyzed with mass spectrometry as described above.

Two kinds of fragment spectrum acquisition modes are possible in mass spectrometry: data-dependent (DDA) and data-independent acquisition (DIA) (91, 92). In data-dependent acquisition (DDA) the mass spectrometer switches between MS (=survey scan) and MS/MS (=fragmentation scan) mode. An MS precursor will be selected for fragmentation when a predefined intensity is reached. Subsequently, the mass spectrometer switches into MS/MS mode in which the precursor ion will be fragmented using collision induced dissociation (CID). By using CID, fragmentation happens at the amide binding of a peptide, creating b and y-ions and allowing peptide identification. Of note, fragmentation can also occur with electron transfer dissociation (ETD). ETD creates c- and z-type ions and preserves most of the post-translational modifications (93). Several problems are reported using DDA (94-96) :

(1) only a fraction of the eluting precursors gets selected, resulting in poor peptide identification repeatability. Michalski et al. (2011) showed by using a LTQ-Orbitrap Velos that out of a HeLa cell

lysate 16 % of all precursor peptide ions were targeted for MS/MS of which 58 % were identified with a false discovery rate of 1 %.

(2) Instrumental scanning speed in MS/MS is too low, resulting in loss of information about the precursor peptide ions that are eluting at that time. In addition, this MS/MS scanning can interrupt the acquisition of the eluting precursor peptide ion at that moment and this leads to inaccurate and non-reproducible quantitation using Area Under the Curve (AUC) quantitation (2.2.1.3.).

(3) cofragmentation of 2 precursor ion peptide precursors in the same isolation window leads to a mixed MS/MS spectrum, resulting in poor identification. This can be solved by using ion mobility - as present in Synapt G2Si - which isolates the two precursor ion peptide precursors before fragmentation.

DIA was developed as a solution to overcome DDA related problems by discarding precursor ion selection altogether. Two types of DIA, both available at our lab, are elevated MS (MS^E, performed on the Synapt G2Si) and Sequential Window Acquisition of all THeoretical fragment ion spectra (SWATH, performed on the TripleTOF 5600) (92). In MS^E, the mass spectrometer alternates between low- and high energy collision conditions without preselection of a precursor ion. Intact precursor ions are recorded in the low energy scan, while in the high energy scan all precursor ions are fragmented by applying a collision energy ramp. This method is also called “parallel fragmentation approach”. The main challenge with MS^E lies in the analysis of the data. In MS^E, the fragment is connected with its precursor ion with the similarity in retention time profile (Accurate Mass Retention Time alignment, AMRT). An iterative depletion process is used in which the annotated peptides and their fragments are subsequently removed out of the data before analysis continues. However, the retention time profile alone proved to be inadequate for accurate identification and quantification (92, 95). For this reason, high definition MS^E ($HDMS^E$) was developed. Herein, traveling wave ion mobility separation (IMS) was merged with the conventional ESI-Q-TOF design. IMS gives an additional dimension of separation, a mobility profile or drift time, to more reliably connect the fragment with its precursor (92). This type of acquisition mode can be run only on a Synapt G2S- or Vion-type of intrument (Waters). This $HDMS^E$

was further optimized to ultradefinition MS^E ($UDMS^E$), in which ion mobility drift time-specific collision energy profiles are used instead of a collision energy ramp. This is represented in **Figure 6A** (97). By using drift time specific profiles instead of an aspecific ramp, Tenzer et al. increased the fragmentation efficiency considerably, resulting in increased peptide identification and proteome coverage (**Figure 6B** (number of proteins) and C (number of peptides) (97).

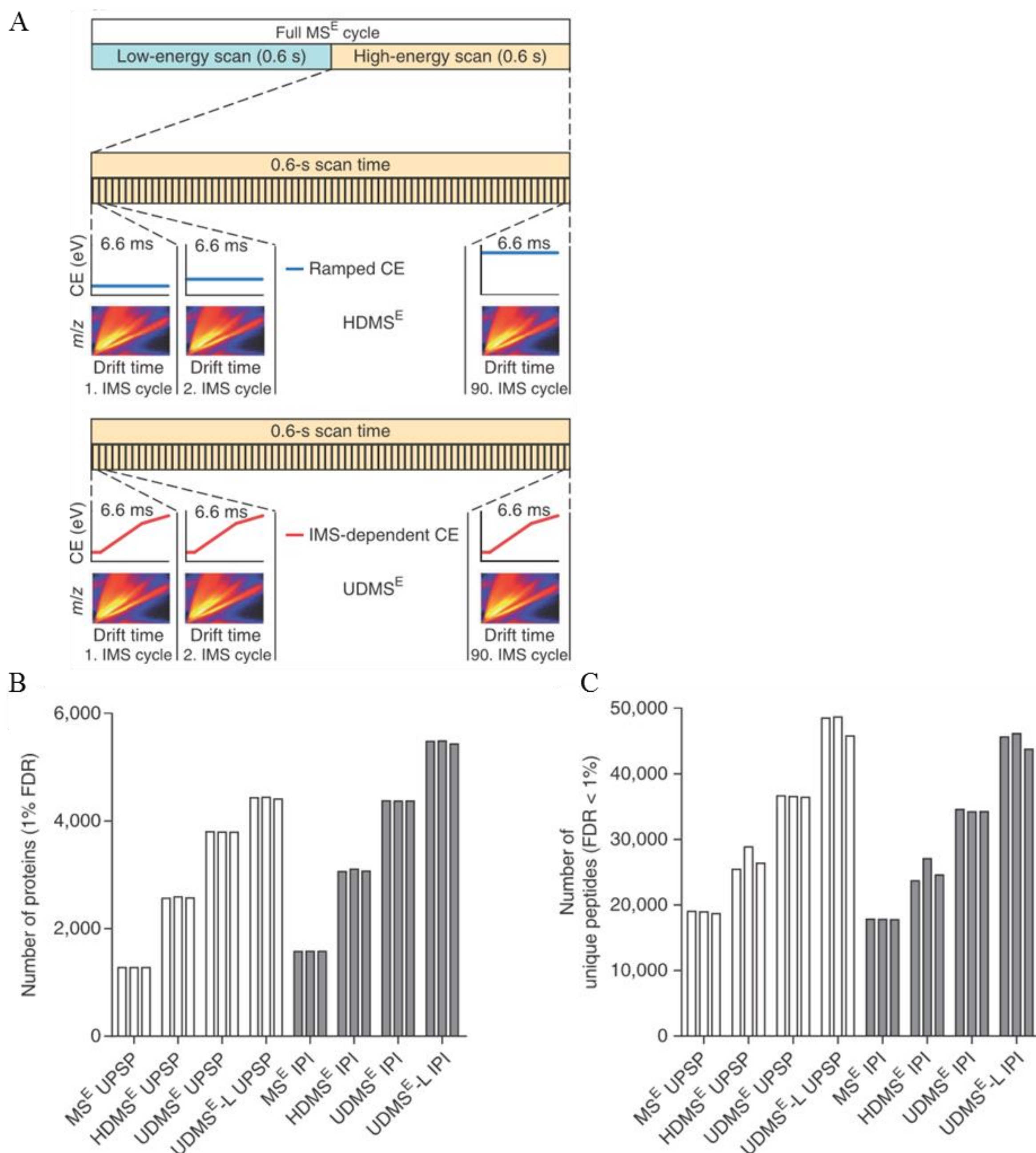


Figure 6. UDMS^E compared to MS^E and HDMS^E. (A) Schematic of HDMS^E with ramped collision energy and UDMS^E with drift time-dependent collision energy profiles in the elevated energy scan. (B, C) Comparison of MS^E, HDMS^E and UDMS^E. HeLa tryptic digest (200 ng) was analyzed in triplicates by MS^E, HDMS^E and UDMS^E using 90-min gradients (UDMS^E-L: 300 ng, 180 min). Proteins and peptides were identified searching against UniProtKB/Swiss-Prot reference proteome (UPSP) and international protein index (IPI) human databases (B) Proteins (FDR <1%; 2 peptides/protein) and (C) peptides (FDR < 1%; ≥6 amino acids). (Image taken from (97))

In contrast to MS^E, SWATH acquisition is acquired on the TripleTOF-type instruments (Sciex). In this acquisition mode, identification of peptides is still done using DDA acquisition, resulting in a library of peptides and their respective retention times. In the subsequent quantitation runs, a *m/z* range (400-1200 Da) is divided into a number of precursor isolation windows (also called swaths or bins, somewhere between 3 and 25 Da in size). The precursor ions selected by this window are fragmented (MS/MS) and data is collected. The mass spectrometer iterates through all windows, before reanalyzing from the start (**Figure 7A**). A full MS scan (without CID) is also acquired with SWATH to obtain precursor ion information. Quantitative information is obtained from MS/MS peak-extracted ion chromatograms of fragment ions of targeted proteins and peptides by the relative quantification of peak areas (**Figure 7B**).

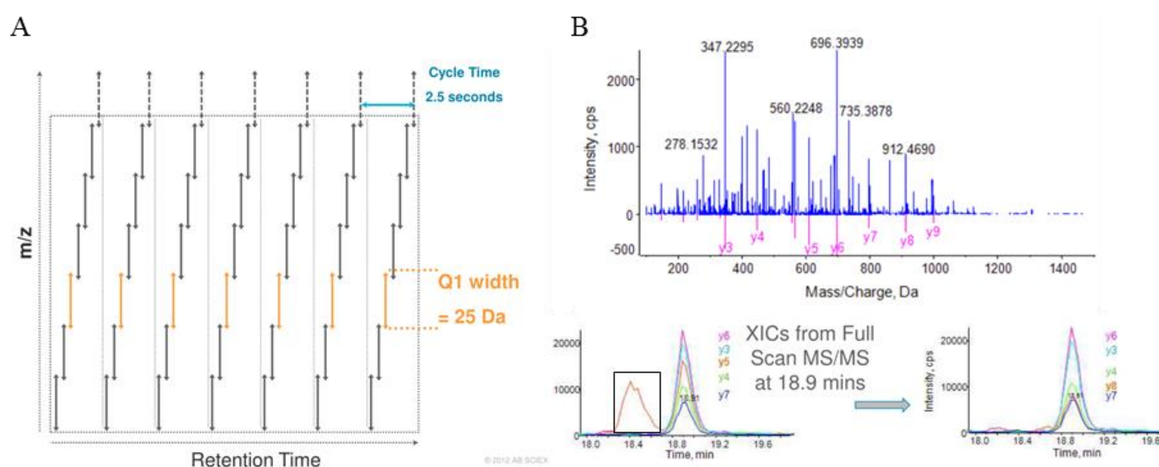


Figure 7. SWATH MS data-independent acquisition and targeted data analysis. (A) A *m/z* range (400 -1200 Da) is divided into precursor isolation windows of around 25 Da width (orange arrows). During the run, the mass spectrometer repeatedly acquires these windows. The cycle time is defined as the time required to return to the acquisition of the same precursor isolation window and is around 2.5 sec. The MS scan, before the beginning of each cycle, is presented as a grey full line. (B) a MS/MS spectrum of a peptide. An extracted ion chromatogram of the fragment ion trances can be plotted for a specific time and can be evaluated. In this case, the y5 fragment has not the same retention time as the other fragments and is not used. (Image adapted from **(98)** and information provided on the website of Sciex)

This is in contrast to MS^E, which obtains information at the MS level (98). Of note, a spectral library of the proteins in the sample is needed for peptide identification. A pre-run of several DDA acquisitions is sometimes needed if no online library is available (94). In addition, co-fragmentation of precursors is inevitable using a 25 Da width bin resulting in an average of 3.4 peptide precursors which are fragmented simultaneously resulting in complex MS/MS spectra (99).

2.1.3 Data analysis

After LC-MS/MS, peptides need to be identified. One way to identify the peptides is by means of sequence database searching. This was performed with the search engine Mascot (Matrix Science) in this dissertation (100). In this technique, RAW data files, obtained by the mass spectrometer, are first converted to peak lists. Subsequently a parameter file (type of enzyme, expected peptide charge, modifications (fixed or variable), MS and MS/MS tolerance) is filled out and a database (SwissProt) is chosen. The database is *in silico* digested taking into account the predefined parameters. Finally, the experimental peak list (containing the masses of precursor ions as well as of peptide fragmentation ions) is compared with the generated peaklist of the *in silico* digested database (100, 101). A probability-based ion score and accompanying "expectancy value" is provided for each peptide, reflecting the probability of this match being a random event. Of note, no peptide identification can occur if (1) the used database doesn't contain the peptide, (2) there are unknown modifications and (3) if the mass tolerance window is too narrow. An error tolerant search can be used to find new modifications (102).

If an MS/MS spectrum is not identified, other approaches such as *de novo* sequencing, sequence tag-based approaches or spectral matching can be considered (101, 103).

Validation of the annotation accuracy can be done by a decoy database search (104). A decoy database consists of randomized or reversed sequences, so no true matches can be expected. The number of matches that is found, is an estimation of the number of false positives amongst the hits in the true database (104). The false discovery rate (FDR) can be calculated out of these data by dividing the number of matches in the decoy database (false positives) by the number of matches in the target database (true positives and false positives). A second way to validate the data is with the use of Percolator (105). Percolator is an algorithm which utilizes a semi-supervised machine learning for improved discrimination between correct and incorrect spectrum identifications. Data matches from a decoy search and high confident data are used as respectively incorrect and correct match examples (105).

MS^E data can be analyzed with the commercially available Progenesis Q1 Software for Proteomics (Nonlinear Dynamics, Waters). This software works as is presented in **Figure 8**: first, retention time alignment is performed by aligning the retention time of all the samples to one sample, the quality control. This quality control (a mixture of all samples of the experiment) can equally be used to check the technical repeatability of the LC-Synapt G2Si, as it is run several times between the randomized samples. A score is given to each alignment. After alignment, peak detection is performed by Progenesis algorithm. Peak filtering is possible in which maximum charge (4+ in our case), minimum and maximum retention time, sensitivity and chromatographic peak width can be adapted. In parallel, a database search performed by the Ion Accounting algorithm has to be performed for peptide identification. An input file is generated by a combination of the Apex3D algorithm (peak detection), a lock-spray mass correction of the data and the Pep3D algorithm is used to group ions belonging to the same peptide. By means of AMRT, precursor ions are assigned to their product ions. A decoy database is generated using a protein FDR of 4 %. Next, the database search is performed by the Ion Accounting algorithm in which an iterative process is used. This iterative process consists out of three passes (99):
(pass 1) Each parent/product ion list is matched against the protein database. The peptides are scored based on their correlation. The process terminates when the FDR is reached.
(pass 2) Only the depleted data are used in this search. The peptides are subjected to possible modifications or non-specific cleavage and are assigned to peptides identified in pass 1.
(pass 3) A product ion is allowed to have a higher intensity than its precursor ion (characterizing in-source fragmentation).

The output results file of the Ion Accounting algorithm is matched up against the Progenesis MS precursors at the end in the Identify Peptides stage of the workflow. Normalization can be performed. One can choose between “Normalize to all proteins” or “Normalize to a set of housekeeping proteins” such as a known amount of spike-in, for example Hi3 protein or β-galactosidase. Quantification can be

performed both relative and absolute (for example by means of the spike-in) based on the peak detection by Progenesis. In addition statistical analysis can be performed (106).

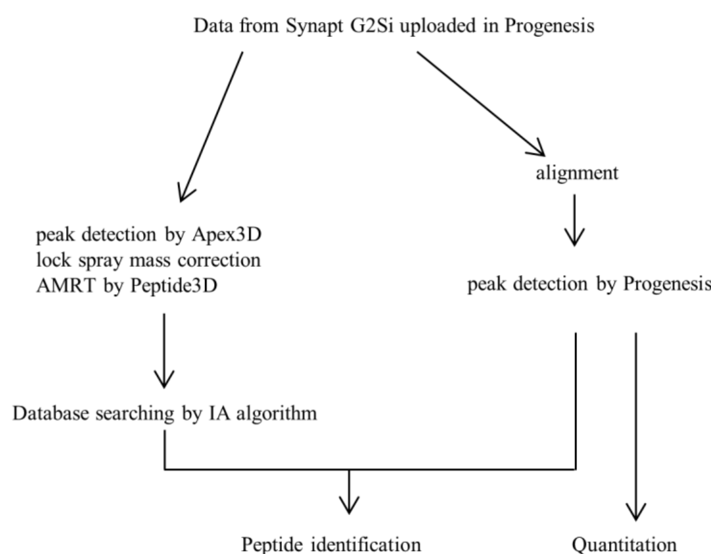


Figure 8. Presentation of the working of Progenesis for MS^E data. Data are aligned and peak detection occurs. For peptide identification, an input file is created for the Ion Accounting (IA) algorithm by Apex 3D, lock-mass correction and Peptide 3D. Subsequently, database searching was performed using the IA algorithm. The output file was subsequently matched to the MS precursors identified in Progenesis. Quantitation can also be determined out of the peak detection made by Progenesis (2.2.1.3).

Besides identification of proteins, quantitative information concerning proteins is often mandatory for biological interpretation. Several MS-based techniques are available each of which contends with its own difficulties.

First, one can chose between relative or absolute quantification. Relative quantification provides a protein ratio between two samples in which the ratio between abundances highlights the differences between two protein profiles. In absolute protein quantification, a protein quantity value is provided (for example “protein copies per cell”) making comparison across datasets possible. Another example of absolute protein quantification can be found in the field of new therapeutic and diagnostic biomarkers: knowing the biomarker concentration in the blood or urine in a patient’s sample makes it possible to define a cut-off value (107).

Several technologies are available in relative and absolute quantification. These are described into detail below.

2.2.1 Relative quantification

Relative quantification technologies can be divided into three groups: metabolic labeling, chemical labeling and label-free quantification. As can be observed in **Figure 9,** the point in the protocol where samples are joined and where consequently technical variation can no longer accumulate differs significantly between approaches. Samples are combined the earliest with metabolic labeling, followed by chemical labeling and no combination of the samples takes place with label-free quantification. The latter approach thus is most prone to technical variability and preparative steps need to be minimized herein.

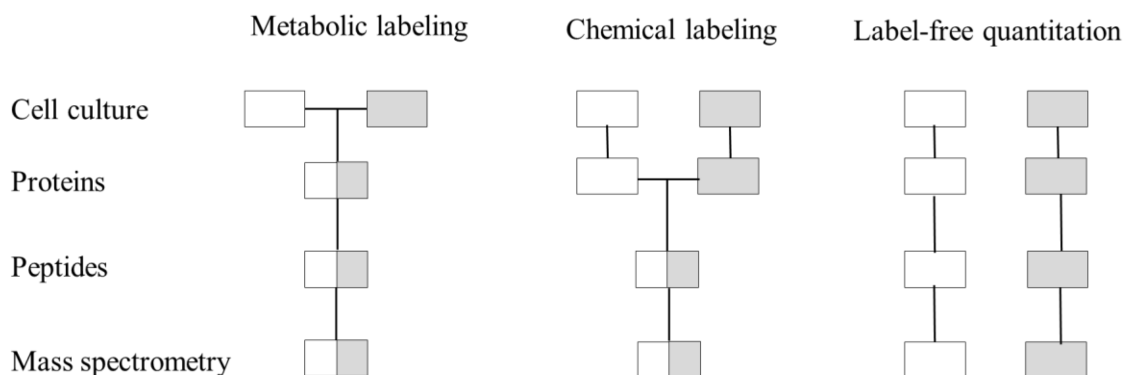


Figure 9. Presentation of the different technologies for relative quantification. This scheme shows the different stages of an experiment (starting from cell culture to mass spectrometry analysis). The 2 boxes (white and grey) represents the two samples, while the horizontal black line denotes the stage in which samples are combined. (Figure adapted from (108))

2.2.1.1 Metabolic labeling

In metabolic labeling, the proteome is labeled during cell culture or *in vivo* by means of a stable-isotope labeled medium. An unlabeled culture/organism is used as a reference. This stable-isotope labeled medium induces a mass shift of the precursor peptide ion, making relative quantification possible (72). Stable Isotopic Labeling by Amino Acids in Mammals (SILAM) labels organisms *in vivo* (109). This is

possible by feeding the organisms with a ^{13}C or ^{15}N-enriched diet (in the form of spirulina) or by using $^{13}C_6$ L-lysine (109, 110).

The technology that labels cell cultures is called Stable Isotopic Labeling by Amino Acids in Cell culture (SILAC) and is developed by the group of M. Mann in 2002 (**Figure 10A**) (111). SILAC makes use of one or two labeled amino acids to label the proteome. These amino acids have to be essential, because otherwise a mix of labeled and non-labeled amino acids will arise in one sample, resulting in incorrect quantification. Essential amino acids in human are histidine, (iso)leucine, lysine, methionine, phenylalanine, threonine, tryptophan and valine (112). Arginine is an essential amino acid in several cell cultures (113). Of note, amino acids are also present in fetal bovine serum (FBS), used as a nutrient source for cells. Hence, this needs to be dialyzed before it can be used in SILAC medium. The first SILAC study used deuterated leucine for labeling (111), but it is known that deuterated peptides induce a chromatographic shift (other retention time) on a reverse phase column making it more difficult to handle the data afterwards (114). In addition, not every peptide contains a leucine, and thus not every peptide is quantifiable. For this reason, the use of $^{13}C_6$ arginine and $^{13}C_6$ lysine is more popular nowadays since trypsin, cleaving C-terminal of those two amino acids, is the most used enzyme for digest (115).

After a complete incorporation of these amino acids (mostly 5 population doublings), the cells of 2 (or to a maximum of 5) samples are combined. After cell lysis and an in-solution digest, samples are ready for LC-MS/MS and can be analyzed afterwards.

Incorporation efficiency and metabolic conversion need to be addressed if SILAC is used for the first time. Incorporation efficiency is the time needed to obtain a full incorporation of labels into the proteome and is determined by checking the presence of light peptides in the mass spectrum (115). Another issue is the possibility of metabolic conversion, such as can be seen when using arginine. This arginine is converted to proline and glutamate in some cell types. This results in extra peaks in the mass spectrum, resulting in incorrect quantification as is presented in Figure 10B. This metabolic interconversion has been reported in HeLa, HEK 293 and hESC (**Chapter 5**) (115, 116).

Disadvantages of SILAC and SILAM are the high cost of the labels and the limited applicability to systems grown in a lab environment, excluding e.g. the use of patient samples. But overall these techniques are considered to provide the most accurate information.

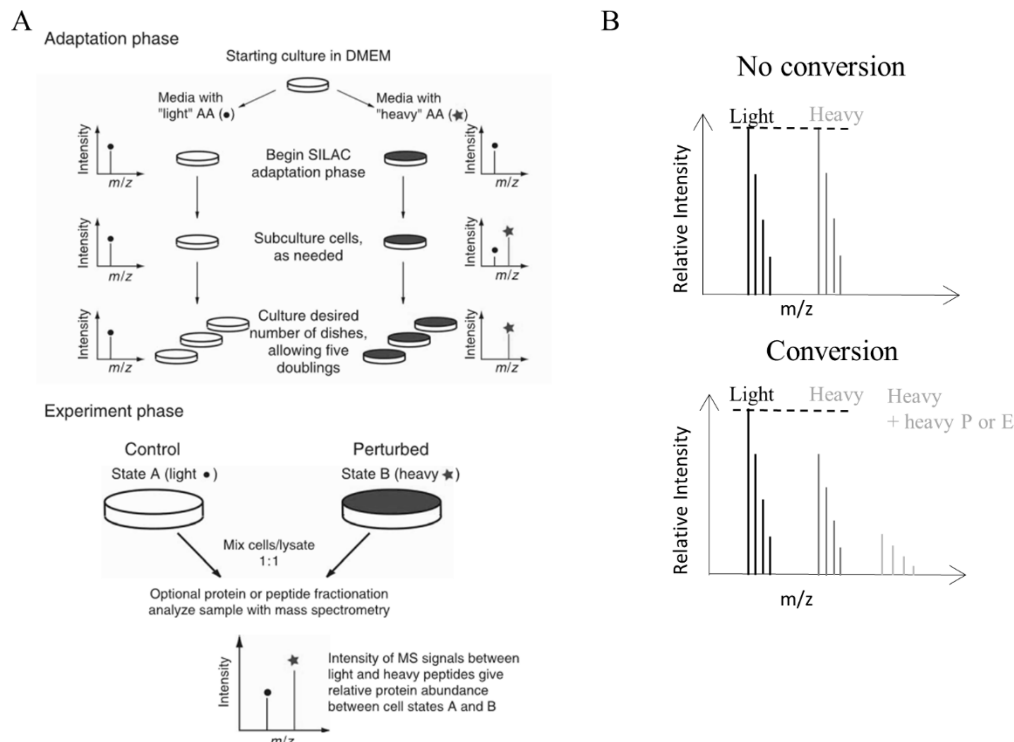


Figure 10. Stable Isotopic Labeling by Amino Acids in Cell culture (SILAC). (A) Cells are grown in light and heavy medium until complete incorporation is obtained (mostly after 5 doublings). After this adaptation phase, the cell populations are treated differently, combined, digested and analyzed with tandem mass spectrometry. In the MS spectrum: the precursor ion of the light and heavy sample are separated and relative quantification can be performed. (B) Metabolic conversion of arginine to proline (P) or glutamate (E) generates an extra peak (heavy + heavy P or E) resulting in incorrect quantitative results. (Image adapted from (115, 116))

SILAC was further optimized for several applications. Dynamic SILAC is used to determine protein turnover (117), while pulsed SILAC can be used to quantify protein translation events (118). Super SILAC can be used to quantify tissues by using a mix of SILAC-labeled cell lines as internal standard (119).

Absolute SILAC allows absolute quantification of selected proteins in a complex mixture. SILAC labeled proteins are spiked into a cell lysate after purification and concentration determination (by means of amino acid analysis or UV absorption at 280 nm) and work as an internal standard (120).

2.2.1.2 Chemical labeling

A lot of reagents are available for chemical labeling of peptides. In this introduction, only a limited number of these reagents are described: Isotope Coded Affinity Tag (ICAT) (121), isobaric Tag for Absolute and Relative Quantification (iTRAQ) (122) and finally ^{18}O labeling (123).

ICAT is a technique to identify and quantify cysteine containing proteins in a sample. In short, the sample is treated with an ICAT tag, consisting of a reactive group which reacts with the sulfhydryl group of cysteine, an isotopically coded linker (^{12}C or ^{13}C) and an acid-labile biotin affinity tag. One sample is treated with ^{12}C ICAT, another sample with ^{13}C ICAT. Samples are mixed, digested and cysteine containing proteins are purified with affinity chromatography (binding with avidin). After LC-MS/MS analysis, peptides are quantified in the MS mode. A disadvantage of ICAT is that this technique can only be used for cysteine containing proteins (121, 124).

Other labeling strategies were developed such as iTRAQ and tandem mass tags (TMT) (125). Both techniques quantify at the MS/MS level and the labeling strategy is similar in both technologies. For this reason, only iTRAQ is discussed into detail (**Figure 11**). The iTRAQ reagent consists out of an isobaric tag (reporter group (mass 114-117 in the 4-plex) and balance group) and an amine specific peptide reactive group. Samples are first digested into peptides with trypsin/LysC in an amine-free buffer (for example triethylammonium bicarbonate) and subsequently the samples are treated with the iTRAQ reagent (114, 115, 116 or 117) which reacts with the free amines in the sample (N-termini and lysine on the peptide). After LC-tandem mass spectrometry, relative quantification occurs at the MS/MS level (122, 124).

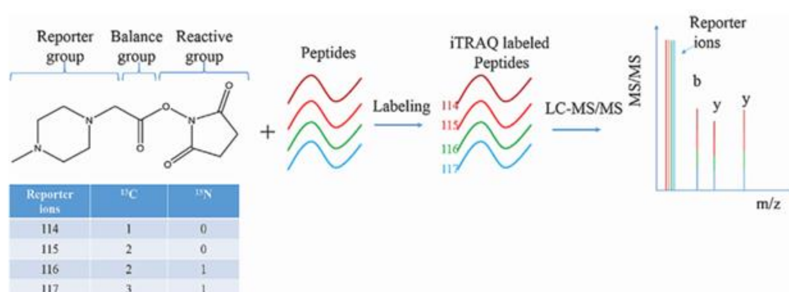


Figure 11. Isobaric Tag for Relative and Absolute Quantification (iTRAQ). Presentation of the iTRAQ protocol. The iTRAQ tag (consisting out of a reporter group, a balance group and a reactive group) reacts with the free amino groups on the peptides. After labeling, the reporter ions in the MS2 can be used for relative quantification. (Image taken from (126, 127))

^{18}O labeling is an example of an enzyme-catalyzed labeling technique (126). In here, two ^{18}O atoms are incorporated in the C-terminal carboxyl group of a peptide in the presence of a serine protease (for example trypsin) and $H_2{}^{18}O$. By treating one of the samples with this method, relative quantification is possible at the MS level (123, 126).

2.2.1.3 Label-free quantification:

Disadvantages of methods that are based on labeling are amongst others (1) the increased sample preparation time, (2) an increased sample complexity, (3) limited sample number that can be analyzed at once (SILAC= 5 samples at once), and (4) the high reagents cost (91, 126). Label-free quantitation was developed as a solution to provide faster, cleaner and simpler quantification results. Two quantitation methods can be distinguished in label-free quantitation: Spectral Counting or Area Under the Curve (AUC) measurement (91, 128).

Spectral counting is based on the fact that abundant peptides are more selected in DDA for fragmentation, resulting in more MS/MS spectra. Several approaches are developed: emPAI, absolute protein expression (APEX) and normalized spectral abundance factor (NSAF) are some examples. The protein abundance index (PAI) quantifies the proteins in the sample by calculating the number of observed peptides divided by the number of observable tryptic peptides for each protein within a given mass range of the mass spectrometer. emPAI (10^{PAI} -1) is the exponentially modified form of PAI and

is directly related to the protein content in the sample. emPAI is considered to give only an indication of the protein amount. It does not take into account (1) the length of the protein (a longer protein will generate more peptides resulting in more MS/MS fragments) or (2) the probability that the peptide can be detected by the mass spectrometer as is done by respectively APEX and NSAF (72, 91). A great disadvantage of spectral counting is that saturation effects (i.e. no additional peptides are detected even when increasing the protein concentration) are common, resulting in incorrect quantification of proteins in a complex sample mixture. In addition, peptides need to be identified before they can be used for quantitation and this is dependent of the used database (129).

AUC measurement on the other hand relatively quantifies the sample by defining the area under the curve or signal intensity measurement of the precursor (91). In this thesis, AUC measurement in UDMS^E mode at the MS1 level was obtained with the Synapt G2Si and analyzed with the commercially available Progenesis Q1 Software for Proteomics (Nonlinear Dynamics, Waters). The survey scan (MS1 data) is herein used for ion abundance quantitation. The peptide abundance is calculated by summing the intensities of all isotopes of the peptide (130) .

2.2.2 *Absolute quantification*

Targeted mass spectrometry under the form of selected reaction monitoring (SRM) and multiple reaction monitoring (MRM) is the gold standard for accurate absolute quantification. SRM is performed with a triple quadrupole instrument in which the first and the third quadrupole are used as mass filters and the second quadrupole as a collision cell. SRM is limited to a hypothesis-driven method. This means that *a priori* knowledge is necessary to perform SRM, so it is mostly used as a validation tool (107).

Absolute peptide quantification can be performed using peptide standards (e.g. synthetic peptides) or protein standards (e.g. absolute SILAC) as internal standard. These standards are spiked in the sample before or after the digestion step (107).

CHAPTER 2: AIMS & OVERVIEW

Studying the molecular biology of stem cells is crucial for a plethora of reasons. Their application in toxicology, regenerative medicine, etc. is entirely dependent on the knowledge that is generated in fundamental studies. Over the course of three different chapters, the research described in this thesis aims at making a valuable contribution to this field by providing a detailed description of several experimental considerations in both cell biology and proteomics approaches.

First, the characterization and culture of hESC is addressed. When hESC culture was introduced at our lab, this was a feeder cell culture system (on MEF). Throughout this doctorate, a feeder-free culture was optimized because of several advantages described in detail in **Chapter 1-Introduction**. The main focus during the culture optimization is monitoring the pluripotency state of these cells. This technique had to meet several requirements. First, it should allow a check-up of the pluripotency of hESC without substantial cell-loss. Secondly, it had to be user friendly, permitting everyone to use the technique without extensive training. Finally, the technique had to be compatible with the imaging instrument present at the lab. The development of this non-invasive technique and its validation by means of flow cytometry is described in **Chapter 3**.

In the second part of this dissertation, the sample preparation for a bottom-up proteomics experiment in human cells was optimized in terms of protein identification and repeatability for label-free quantitation. Herein, proteins were analyzed with tandem mass spectrometry (DIA mode) after being prepared through (1) cell lysis, (2) protein digestion into peptides by means of trypsin/LysC and (3) peptide separation and analysis by LC-MS/MS. This sample preparation protocol was optimized in three cell lines and is described in **Chapter 4**.

In the last part of the dissertation, SILAC was optimized for hESC analysis. SILAC is a quantitative method in which the samples are metabolically labeled during cell culture. The samples can be mixed even before cell lysis resulting in minimal technical variability, but the technique can suffer from metabolic conversion of the heavy amino acids. More specifically, conversion of arginine was shown to be a considerable problem in hESC and had to be reduced in a fully defined and xenofree feeder-free

culture, E8 medium in combination with vitronectin, as routinely used in our lab. Several solutions were proposed to reduce this arginine conversion in hESC as discussed in **Chapter 5**.

Together, the results presented in this dissertation should allow the reader to adapt his hESC cell culture system and to be able to monitor the changes thereof, both directly, using the reporter cell line and more generally, by quantifying changes in the proteome by either label-free or SILAC-based proteomics approaches.

CHAPTER 3: MONITORING OF PLURIPOTENCY IN HESC

Detailed method description for non-invasive monitoring of differentiation status of human embryonic stem cells.

Based on "*Detailed method description for non-invasive monitoring of differentiation status of human embryonic stem cells.*" Ellen Scheerlinck, Katleen Van Steendam, Mado Vandewoestyne , Trees Lepez, Veerle Gobin, Paulien Meert, Liesbeth Vossaert, Filip Van Nieuwerburgh, Ann Van Soom, Luc Peelman, Björn Heindryckx, Petra De Sutter, Maarten Dhaenens*, Dieter Deforce*

*Authors contributed equally

Published in *Anal Biochem* **2014,** 461, 60-66

Abstract

The (non-)differentiation status of hESC is usually analyzed by determination of key pluripotency defining markers (OCT4, NANOG, SOX2, …) by means of RT-qPCR, flow cytometry (FC) and immunostaining. Despite proven usefulness of these techniques, their destructive nature makes it impossible to follow-up on the same hESC colonies during several days, leading to a loss of information. In 2003, an OCT4-eGFP Knock-In hESC line to monitor OCT4-expression was developed and commercialized. However, to the best of our knowledge, the use of fluorescence microscopy (FM) for monitoring the OCT4-eGFP expression of these cells without sacrificing them has not been described in detail to date. Here, we describe such a method in detail, emphasizing both its resolving power and complementary nature to FC, as well as the potential pitfalls in standardizing the output of the FM measurements. The potential of the method is demonstrated by comparison of hESC cultured in several conditions, both feeder-free (vitronectin) and grown on feeder cells (MEF).

Introduction

Human embryonic stem cells are characterized by their unlimited proliferation potential (self-renewal) and their ability to differentiate into all cell types of the mesoderm, ectoderm and endoderm germ layers (pluripotency). These hESC could therefore potentially be useful in the field of regenerative medicine (1, 4). The process of differentiation has been extensively explored, but finding ways of keeping hESC undifferentiated is equally essential for fundamental clinical research and toxicological screenings (131). Commonly used markers for identifying this undifferentiated status include the two key pluripotent transcription factors OCT4 & NANOG (14, 132).

Currently, the differentiation status is routinely measured using immunostaining, real-time PCR and/or flow cytometry. Although these techniques have proven their usefulness in analyzing hESC, it is necessary to sacrifice cells, which makes it impossible to monitor the same cells during the experiment. Here, we evaluate the applicability of a commercially available OCT4-eGFP Knock-In hESC line (WiCell Research Institute, Madison, WI) in combination with fluorescence microscopy for non-invasive examination of (non-)differentiation of hESC. This OCT4-eGFP Knock-In hESC line (= OCT4 reporter hESC line) was developed in 2003 by means of homologous recombination whereby the transcription of enhanced green fluorescent protein (eGFP) is regulated by the promoter region of OCT4 (61). The pluripotent status of hESC can be verified in this hESC line by means of eGFP detection: a decrease in eGFP represents a decrease in OCT4 expression and thus a decrease in pluripotency, indicating that the hESC line is differentiating. This hESC line has been used for different purposes, for example to analyze cell division and to create induced pluripotent stem cells (133, 134). Non-invasive monitoring of OCT4 can be advantageous to e.g. investigate the effect of different culture conditions on hESC pluripotency. To our knowledge however, no detailed description of a methodology in which this hESC line is used in a non-destructive time-lapse experiment has been published to date.

A non-destructive and fast way to define the differentiation status of the cells of this hESC line is to measure the fluorescence of the hESC colonies by means of fluorescence microscopy (FM). By

determination of the densitometric means of a specific colony and the background, the signal to noise ratio (s/n ratio) can be compared between different conditions on a daily basis without any loss of cells. Flow cytometry (FC) on the other hand allows determining fluorescence at the single-cell level and is arguably the gold standard, despite its destructive nature. Our goal was to correlate the measurements of both techniques.

To validate this non-destructive method, hESC were differentiated using 2 µM retinoic acid (RA). Their s/n ratio was compared with hESC cultured in medium containing basic fibroblast growth factor (bFGF), a well-known growth factor to maintain self-renewal and pluripotency. Subsequently, feeder-free culture and feeder cell culture of hESC were analyzed in parallel in order to determine the effect of mouse embryonic fibroblasts on our hESC colonies. Fluorescence microscopy measurements added valuable information in interpreting FC experiments. More specifically, the use of FM has the additional advantage that it allows monitoring of hESC colony morphology and colony homogeneity, which we demonstrate to be a considerable source of variance undetected at the single cell level.

Finally, one application of this method is given where mouse embryonic fibroblast conditioned medium (CM) is used to analyze its beneficial impact on hESC growth.

Material & Methods

1.1 Materials

All products were purchased from Life Technologies, Carlsbad, CA, USA unless stated otherwise.

1.2 hESC culture on feeder cells

Human embryonic stem cells were cultured on feeder layers of inactivated MEF. Mouse embryonic fibroblasts were grown to confluence in a T75 culture flask (37 °C, 5 % CO_2, 5 % O_2) using medium composed of DMEM, 10 % FBS, 100 u/ml penicillin, 100 µg/ml streptomycin and 2 mM L-glutamine.

Inactivation of MEF occurred by incubation with 10 µg/ml mitomycin C (Sigma-Aldrich, St. Louis, MO, USA) during 2.5 hours at 37 °C**.** The cells were detached from a T75 flask with 0.25 % trypsin-EDTA and plated on a pre-coated 0.1 % gelatin 6-well plate at a density of 20 000 cells/cm^2, and cultured as described above. The next day, the WA01 OCT4-eGFP Knock-In hESC (WiCell, Madison, WI, USA) were plated on the MEF, and cultured in hESC medium consisting of DMEM/F12 with 20 % KO-SR, 100 u/ml penicillin, 100 µg/ml streptomycin, 1 % Non-Essential Amino Acids, 2 mM L-glutamine and 4 ng/ml basic fibroblast growth factor (bFGF). Splitting of the cells was performed every 4-5 days with 0.5 mM EDTA in dPBS, based on the manufacturer's protocol for feeder-free splitting.
Differentiation of hESC was induced by adding 2 µM retinoic acid (RA) and by removing bFGF from the hESC medium.

1.3 Feeder-free culture of hESC

In addition to feeder cell culture, feeder-free conditions were also used to validate the method. For feeder-free culture, hESC were plated on a pre-coated vitronectin (VN) plate (coating concentration = 0.5 µg/cm^2) and cultured in Essential 8™ (E8) medium. Splitting was performed every 4-5 days with 0.5 mM EDTA in dPBS according to the manufacturer's protocol of culturing hESC in Essential 8™ medium. Differentiation of hESC was induced by adding 2 µM retinoic acid (RA) and by removing bFGF from the hESC medium used in feeder cell culture.

For the application of the screening methodology, different media were tested. The composition of these media is summarized in **Table 1**.

Table 1 Composition of the different media used for feeder-free culture of hESC

Components	UCM -	UCM +	CM -	CM +
DMEM/F12	+	+	+	+
1.25 % ITS-A	+	+	+	+
2.5 mM L-glutamine	+	+	+	+
1.25 % NEAA	+	+	+	+
MEF secretome	-	-	+	+
4 ng/ml bFGF	-	+	-	+

Insulin-Transferrin-Sodium selenite-Sodium pyruvate (ITS-A) Conditioned Medium (CM) was made by adding 15 ml ITS-A Unconditioned Medium (UCM) to an inactivated MEF T75 flask (20 000 cells/cm^2). After culturing for 24 hours, CM was collected and filtered through a 0.22 µm Sterivex GP Filter Unit (Millipore, Bedford, MA, USA) for removal of cells and cell debris.

1.4 Fluorescence microscopy & data analysis

After removal of medium, eGFP expression (ex. 485 nm, em. 515 nm, exposure time 5000 ms) of 6 colonies/condition was measured daily on an Axiovert 200M inverted fluorescence microscope equipped with the Axiovision multichannel fluorescence module and an AxioCam MRM camera (Carl Zeiss, München, Germany). Colonies were screened at 10x magnification using a Carl Zeiss Fluar® objective (Carl Zeiss) and visualized using Zeiss filter set no. 38 (BP 470/40, FT 495, BP 525/50). For larger colonies, different TIFF-images were stitched using Photoshop CS4 (Adobe, San Jose, CA, USA). The s/n ratio was determined by dividing the densitometric mean of the colony by the densitometric mean of the background.

1.5 Flow cytometry

In general, FC was carried out at the end of each experiment (day 5). In order to obtain single cells, all cell cultures were incubated with 0.25 % trypsin-EDTA for 6 minutes. Prior to FC analysis, cells were resuspended in PBS + 1 % BSA solution. Flow count beads (Analis SA/NV, Suarlée, Belgium) were added to acquire absolute cell counts. The different conditions were analyzed using Beckman Coulter Cytomics FC500 and CXP analysis software. A minimum of 10000 events was acquired for each condition.

1.6 Celigo S®

Celigo S® Imaging Cell Cytometer (Brooks, Poway, CA, USA) was used to evaluate the confluency (%) during the experiment (from day 0 until day 5). Confluency can be defined as the total coverage of the plate. Medium was removed because of auto-fluorescence and PBS was added to the culture to prevent dehydration during measurement due to the warmth generated by the system.

1.7 Statistical analysis

Statistical analysis was performed using Graph Pad Prism 5 (San Diego, CA, USA). Differences were evaluated by a Student's t-test. A p-value < 0.05 was considered statistically significant.

Results

1 Method optimization of non-invasive monitoring of differentiation status of hESC

During this method optimization a detailed comparison was made between monitoring OCT4 expression by means of non-invasive fluorescence microscopy (FM) and by "destructive" flow cytometry (FC) as gold standard.

1.1 Assessing auto-fluorescence

Auto-fluorescence of cells due to the presence of cellular metabolites such as NADH is a well-known issue and must be investigated to avoid misinterpretation of fluorescence data (135). The auto-fluorescence of MEF was determined by FM and FC, and a non-reporter hESC line (UGENT2, (136)) was included as a negative control.

1.1.1 Fluorescence microscopy

For FM, signal to noise ratio (s/n ratio) is measured by dividing the densitometric mean of the colony by the densitometric mean of its background (Figure 1A). For feeder cultures, it is therefore essential to determine the background noise that might derive from the auto-fluorescence of the MEF. The MEF signal is not detectable near hESC colonies and as such has no impact on the s/n ratio when compared to hESC growing in feeder-free culture (on VN) (**Figure 1A**).

1.1.2 Flow cytometry

Next, these FM measurements were compared to FC. The fluorescence histogram of the OCT4 reporter hESC line (cells detached from a VN plate), from MEF, and of a non-reporter hESC line (UGENT2, cells detached from a VN plate), used as a baseline control, is presented in Figure 1B. No auto-fluorescence was observed in the UGENT2 cell line (fluorescent signal $< 10^0$). MEF auto-fluorescence however, is clearly present but is 10-fold less compared to the true signal of the OCT4 reporter hESC line.

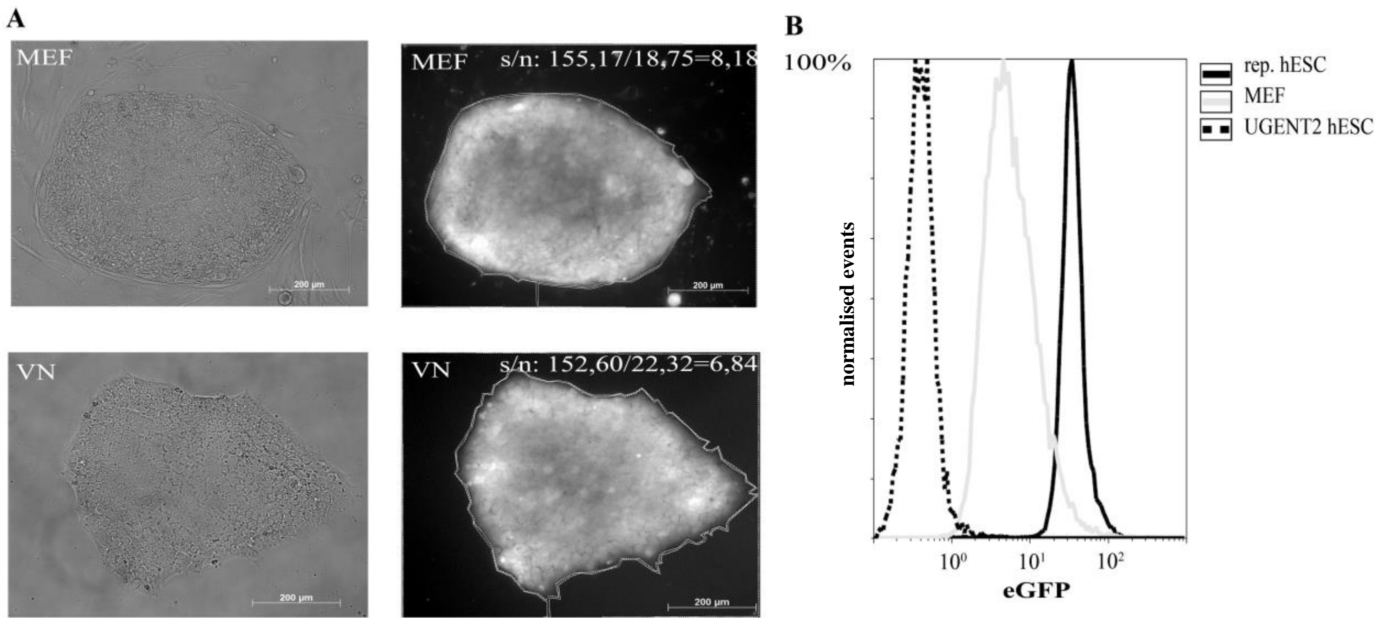


Figure 1. Auto-fluorescence of MEF and hESC. (A) Brightfield image (left) and fluorescence image (right) of a OCT4 reporter hESC colony cultured on MEF (top) and on VN (bottom) obtained by fluorescence microscopy (FM). No auto-fluorescence of the MEF can be observed. The dotted white lines in the right panels illustrate how colonies are manually delineated in all experiments to assess the s/n ratio. (B) Flow cytometry results of MEF, OCT4 reporter hESC line (rep. hESC) and a non-reporter UGENT2 stem cell line (UGENT2 hESC). Each histogram was scaled to 100% of the peak value. No auto-fluorescence of the UGENT2 cell line is observed. MEF show a weak auto-fluorescent signal, but 10 times lower compared to the true signal of the undifferentiated OCT4 reporter hESC line.

Note that the auto-fluorescence from MEF will have an influence on the fluorescence histogram of detached hESC from a plate cultured on MEF. Gating on the FS/SS plot to exclude the MEF from the histogram is not possible because no distinction could be made between MEF and hESC in terms of FS/SS. However, the relative portion of the inactivated MEF compared to the growing colonies reduces over time. As FC is only used at the end of the subsequent experiments, only a small contribution of the MEF (<10 %) to the fluorescence histogram is expected for feeder cell experiments.

Yet, even with decreasing signal during differentiation caution needs to be taken when directly comparing the values of FC measurements of hESC grown on MEF and on VN.

To assess the resolving power of the FC, complete differentiation of the OCT4 reporter hESC line (by means of 2 µM RA, culture on VN) was performed. The fluorescent signal rapidly decreased during the first 6 days and completely disappeared after 15 days in culture, falling back to the same level as the UGENT2 line, with no detectable auto-fluorescence (*data not shown*). Further experiments were done on 6 day-cultures as these showed adequate reduction in fluorescent signal.

1.2 Fluorescent signal in OCT4-eGFP Knock-In hESC line on MEF and VN during differentiation

1.2.1 Fluorescence microscopy

The resolving power of the FM to determine differences in differentiation status was verified as described below. The OCT4 reporter hESC line was cultured on feeder cells (MEF) in three different media to investigate respectively non-differentiation (hESC medium with 4 ng/ml bFGF), spontaneous differentiation (hESC medium without bFGF) and directed differentiation (hESC medium with 2 µM RA) and was also compared with feeder-free conditions (human recombinant VN in two different media (E8^{tm} medium and hESC medium with 2 µM RA)). Only the conditions with media that contain bFGF are assumed to keep the hESC undifferentiated (137). Colony fluorescence (5/6 colonies per condition) was assessed daily during 6 days and results were expressed as signal/noise ratios (Figure 2A). The experiment was carried out in triplicate.

As expected, addition of 2 µM RA caused a significant decrease in s/n ratio during the time of the experiment both on hESC cultured on MEF and VN (Figure 2A). This decrease was linked to a lower expression of OCT4, leading to differentiation. Of note, hESC cultured on VN tend to detach during forced differentiation by RA, something which can be easily detected when colonies are monitored through time by FM based on their coordinates.

Since 4 ng/ml bFGF is thought to be necessary to keep hESC on MEF undifferentiated, culture of hESC in the absence of bFGF would lead to differentiation, but to a slower rate than when differentiation is

artificially induced as for example by adding RA. In Figure 2A, the condition without bFGF has a lower s/n ratio on day 5 than the condition with bFGF on MEF (s/n ratio of 8.34 ± 2.16 in comparison with s/n ratio of 14.58 ± 2.98). This s/n ratio is however higher than when RA was added (s/n ratio of 8.34 ± 2.16 in comparison with s/n ratio of 2.49).

Interestingly, an increase in fluorescence in the "undifferentiated" conditions with bFGF both on MEF and VN was observed on day 5. This might be explained by (i) an increase in eGFP/cell or by (ii) the formation of multilayers (3D growth) resulting in an accumulation of fluorescent signal.

No significant differences were found in s/n ratio from hESC cultures on MEF and those on VN for both the non-differentiating (with bFGF) and the differentiating (with RA) conditions, confirming the low impact of auto-fluorescence of the MEF on FM measurements.

1.2.2 Flow cytometry

The abovementioned data of FM were compared to FC measurements on the hESC population. Because of the destructive nature of this technique, analysis is only performed at the end of the experiment (day 5) (**Figure 2B**). Both feeder-free and MEF grown hESC in the presence of bFGF retained the undifferentiated status (fluorescent signal > 10^1). Of note, a small portion of cells in the latter population had an eGFP expression between 10^0 and 10^1 (Figure 2B asterisk). These cells were most probably MEF, as mentioned earlier (auto-fluorescence between 10^0 and 10^1 (1.1.2)). The finding that the eGFP/cell remained constant in the undifferentiated conditions strongly suggests that the daily increase in fluorescence of the undifferentiated conditions as observed by FM is not due to the increase of eGFP/cell but rather to a multilayer effect (3D growth) resulting in accumulated fluorescent signal (1.2.1).

In the MEF condition without bFGF, most of the cells were still undifferentiated after a 6-day culture (fluorescent signal > 10^1) but in comparison with the MEF condition with bFGF, a significantly higher number of cells with an eGFP expression between 10^0 and 10^1 were observed. These results are in line with the FM measurements, in which it was shown morphologically that there was a mix of

differentiated (low fluorescence; s/n ratio = 2.26) and undifferentiated hESC (high fluorescence, s/n ratio = 13.75) on day 5 when bFGF was absent (**Figure 2C left**).

The conditions with RA showed a clear drop in fluorescence on day 5 which is due to differentiation (**Figure 2B**). This finding is in line with the results obtained by FM (**Figure 2A**). Remarkably, FM images of hESC colonies differentiated with RA revealed the existence of zones ("islands") with high accumulated fluorescence (s/n ratio = ca. 24) (**Figure 2C right**). This small population of high fluorescent "islands" could not be discriminated using flow cytometry as these individual highly fluorescent cells were somewhat hidden in the tail of the fluorescence histogram obtained with FC (**Figure 2B**).

In conclusion, when looking only at the s/n ratios measured by FM for pluripotency assessment, it is important to take into account that an increase of signal of a whole colony does not correlate with an increased eGFP signal per cell and that only a decrease in FM signal can be directly interpreted as an ongoing differentiation. A flat signal in FM can be interpreted as a hESC culture with a population of differentiating and non-differentiating cells.

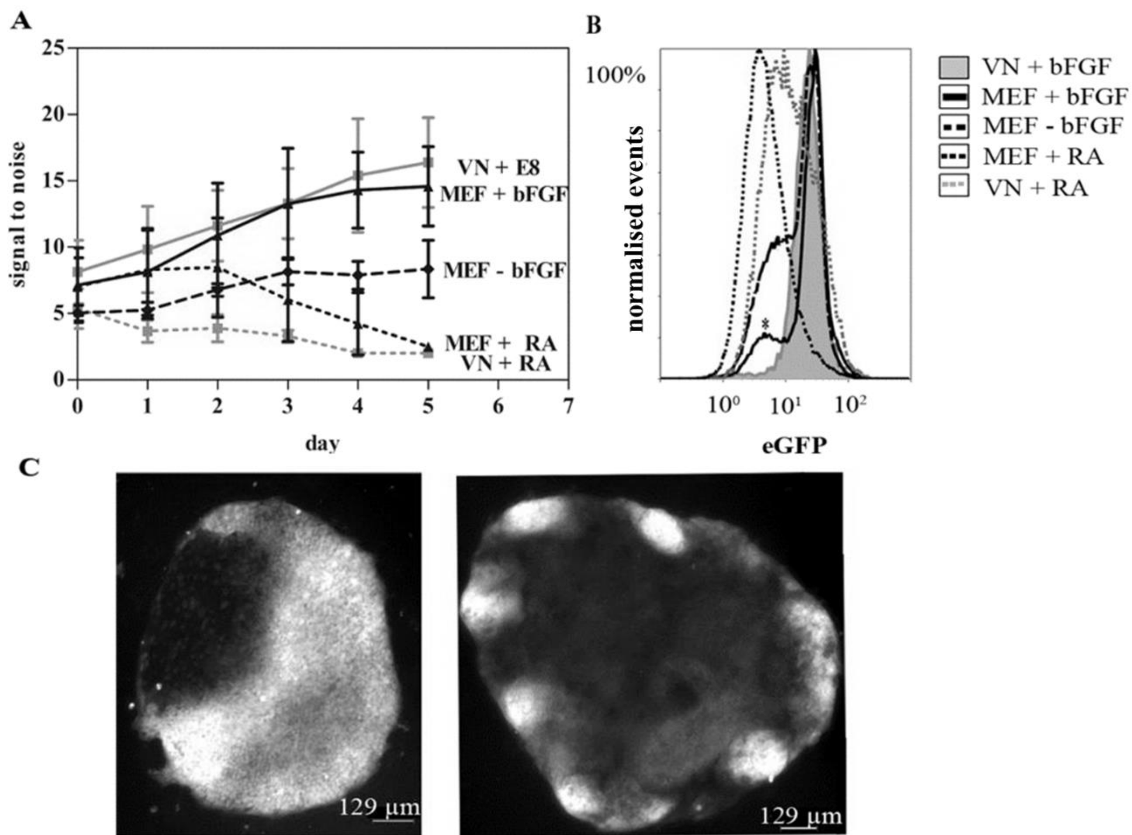


Figure 2. Fluorescence microscopy (FM) and flow cytometry (FC) results of hESC cultured on MEF and on VN. n= total amount of replicates (A) Signal to noise ratio (s/n ratio) measured by fluorescence microscopy of OCT4 reporter hESC cultured on MEF in different conditions: hESC medium with 4 ng/ml bFGF (MEF + bFGF, n=3), without bFGF (MEF – bFGF, n=3) and with 2 µM RA (MEF + RA, n=3), and on VN in different conditions: E8^{TM} medium (VN + bFGF, n=3) and with 2 µM RA (VN + RA, n=5). In the conditions with RA, no error bars are displayed on day 5 because of both colony detachment and because low s/n ratio makes image stitching of the colony infeasible (signal colony ≈ signal background). The experiment was carried out in triplicate (n=3) and 6 colonies were monitored in each experiment (B) FC results at day 5 of the same conditions as described in A. Each histogram was scaled to 100 % of the peak value. Asteriks indicates MEF contamination in the plot of the OCT4 reporter hESC line on MEF with hESC medium + 4 ng/ml bFGF. (C) Fluorescent images of an OCT4 reporter hESC colony on MEF in hESC medium without bFGF (left image) and with RA (right image). A mix of high fluorescent and low fluorescent cells can be observed in the left image. In the right image, "islands" of high OCT4 expression can be observed.

2. Application of non-invasive monitoring of the differentiation status: MEF conditioned medium

A possible application of the above-mentioned non-invasive method is a comparison of different media to test their ability to maintain hESC in an undifferentiated state: E8^{tm} medium on VN (positive control) versus insulin-transferrin-sodium selenite-sodium pyruvate conditioned medium (CM) versus ITS-A

unconditioned medium (UCM). The difference between CM and UCM is that CM contains MEF secreted proteins (Method & Materials). ITS-A was used instead of KnockOut Serum Replacement (KO-SR), since ITS-A contains no albumin in comparison with KO-SR, a great advantage when subsequent mass-spectrometry analysis of the media is envisioned. To our knowledge, this is the first time that ITS-A-containing MEF CM without KO-SR was evaluated for its ability to keep hESC pluripotent.

On top of validating the impact of MEF secreted proteins, the influence of the addition of bFGF (CM+ and UCM+ contain bFGF while absent in CM- and UCM-) was also verified. Besides OCT4 analyses, cell number could be determined as well with FC.

2.1 eGFP expression assessed by fluorescence microscopy

After a 6-day culture, no significant difference in eGFP expression could surprisingly be observed between the positive control (E8^{TM}) and the other conditions (CM+, CM-, UCM+, UCM-) (**Figure 3A**), suggesting that none of the tested supplements significantly downregulated OCT4 expression during short term culture. Morphological assessment of the colonies showed a more or less uniform distribution of the fluorescence in the colony in all conditions, indicating a comparable multilayer formation if all cells have the same OCT4 expression (confirmed with FC).

2.2 eGFP expression assessed by flow cytometry

Flow cytometry was used to validate the results of FM at the end of the experiment (**Figure 3B & C**). In addition to OCT4 expression at day 5 (by means of x-mean of the green channel, **Figure 3B**), the amount of cells was also counted by using flow count beads as a reference (**Figure 3C**). Both values are expressed as the logarithmically normalized ratio with respect to E8^{TM} (positive control). As seen in **Figure 3B**, a slight decreasing trend (not significant) in OCT4 expression at the single cell level can be observed over the different conditions at day 5, confirming FM results. In contrast, the total cell number

at the end of the experiments was on average higher in CM compare to UCM with statistical significance only found between CM- and UCM- (p-value < 0.05).

2.3 Confluence

This difference in cell number between CM and UCM was verified by assessing the confluence of colonies by use of a Celigo imaging cell cytometer (**Figure 3D**). Less confluence could be observed in the conditions with UCM, confirming the results obtained with FC.

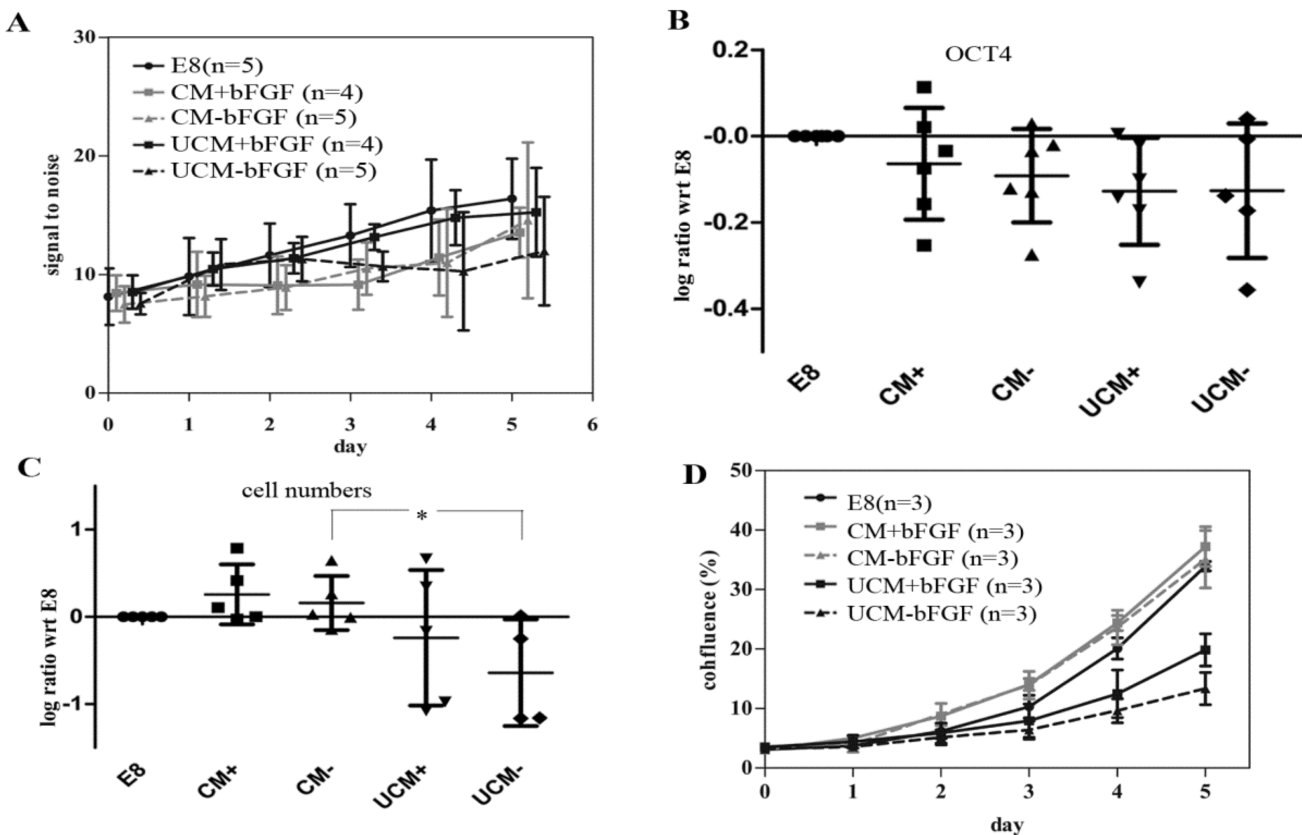


Figure 3 Influence of different media on hESC pluripotency and cell growth measured by means of FM and FC. The experiment was carried out in triplicate or more (n = 3, 4 or 5) and 6 colonies were monitored in each experiment (A) Signal to noise ratio (s/n ratio) after FM of OCT4-eGFP Knock-In hESC on VN in combination with different media: E8tm, CM (with and without bFGF) and, UCM (with and without bFGF) during 6 days. (B) OCT4 expression by means of x-mean of the abovementioned conditions determined with FC on day 6. Results are expressed as the log ratio with respect to E8TM (positive control) for normalized representation. (C) Cell number of the abovementioned conditions determined with FC after 6 days. Flow count beads were used to assess absolute count concentration. Results are expressed as the log ratio with respect to E8tm (positive control) for normalized representation. *p-value < 0.05 (D) Confluence (%) determined by means of the Celigo® cytometer of the same abovementioned conditions during 6 days.

2.4 Stress

We noticed that especially in UCM media repeatability of the experiments was very low. In both FM and Celigo however, some stress is induced by removal of the medium before measurement and by the lack of a controlled environment in the apparatus used during these experiments (temperature, CO_2, O_2). We hypothesize that this stress caused the lack in repeatability in UCM cultures. No stress was induced by analyzing hESC only at the end of the experiment (on day 5) with FC without intermediate FM or Celigo measurements (no stress during time of the experiment) and these results were displayed in **Figure 4A** (OCT4 expression) & **4B** (cell numbers). Note that media were replaced throughout the time course of the experiment and that detached cells were removed each day, as was the case in the earlier described experiments.

Importantly, no significant influence of stress on OCT4 expression could be observed. When comparing the cell number however, it is clear that the difference in cell number in the different conditions (**Figure 3C**) is strongly reduced when hESC are grown without stress. This stress-caused effect can probably be avoided by using auto-fluorescence free medium in combination with a controlled environment.

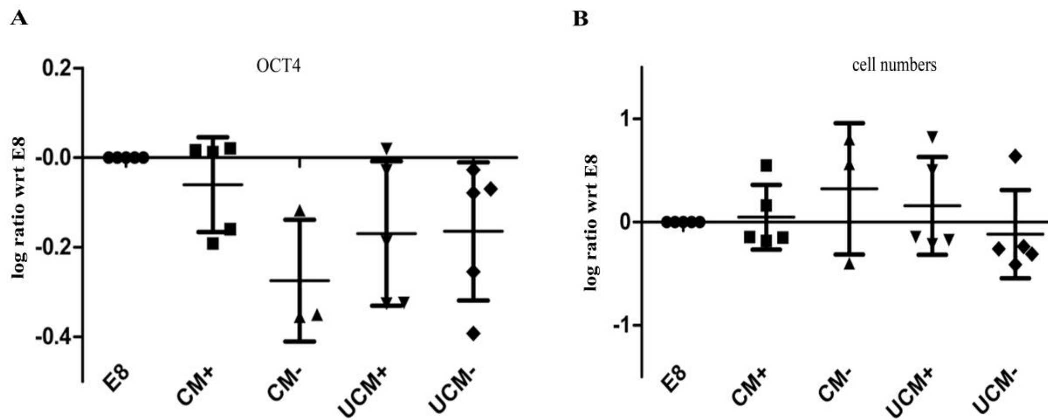


Figure 4 Influence of stress induced by measurement on the FM. Flow cytometry results on day 6 of the different media as mentioned in Figure 3. No FM measurement was carried out. Results are expressed as the log ratio with respect to E8[tm] (positive control) for normalized representation (A) OCT4 expression by means of x-mean. (B) Cell numbers.

Discussion

The goal of this study was to validate an easy and non-destructive method to follow up pluripotency (by expression of OCT4) and morphology of hESC. To our knowledge, this is the first detailed description of such a method for evaluating the (non-)differentiation status of hESC. It is based on the use of a commercially available OCT4-eGFP Knock-In hESC line (WiCell) in combination with a fluorescence microscope (FM). Method validation and cell number assessment were accomplished by means of flow cytometry (FC) as the gold standard. Therefore, no special live-cell imaging instruments are required when implementing this method.

As a measure of OCT4-expression, colony signal to noise ratio (s/n ratio=densitometric $mean_{colony}$/densitometric $mean_{background}$) and single-cell x-mean of the fluorescence were measured by FM and FC, respectively. Using FM, we found a surprising daily increase in s/n ratio of hESC colonies in the undifferentiated conditions (+bFGF). When measuring OCT4-expression at single cell level with FC, this increase in fluorescence was not observed (same eGFP/cell during the experiment). Therefore, increased s/n ratio of whole colonies seen by FM is likely due to 3 dimensional growth (= multilayer effect). Although FM can be used for following up the OCT4 expression of hESC colonies, one should

keep in mind that converting these images into a single number (s/n ratio) will, by definition, result in a loss of information content. Before implementing FM as a non-destructive screening tool, s/n ratio values first need to be compared to additional FC measurements to examine the OCT4-expression at single cell level.

FM however, and not FC, can assess changes in colony homogeneity and morphology. This was clearly illustrated by “islands” of high fluorescence in RA-differentiated hESC colonies. These “islands” expressed high amounts of eGFP and thus OCT4, suggesting the existence of small remaining populations of non-differentiated cells after 6 days of RA-induced differentiation. These cells were somewhat “buried” in the tail of the FC histogram and are difficulty to be detected by this technique. Co-staining with other germ line markers is needed to elucidate the origin of these islands, but this lies outside the scope of this study.

Of note, we tried to define colony homogeneity by following up the standard deviation (SD) of the colony during the time of the experiment with FM (colony homogeneity cannot be assessed by FC (only single cell analysis)). In theory, a polymorph colony (regions of high and low fluorescent areas) will have a higher SD than a uniform colony and the formation of a polymorph colony will lead to an increase in SD during the time of the experiment. Different SD calculations were tested, but to our surprise, none showed in full the polymorphism of the colony in comparison with the morphological images taken by FM (*data not shown*). However, other analyzing software packages such as ImageJ, can possible be used to determine this colony homogeneity.

Although FC will still be needed to assess information at single-cell level, FM is unique in providing daily information about the distribution of OCT4-expression in different colonies in a non-destructive way that no other technique can accomplish up to now thereby giving new insights in how cells will respond to different stimuli in terms of hESC differentiation. For defining lineage commitment however, one must still stain with other markers such as SSEA-1 & GATA4.

Once FM measurements have been essayed against FC, our method can be used to e.g. analyze the effect of different compounds (small molecules, proteins) on hESC pluripotency, morphology and cell growth. This application was tested by observing differences in colony growth in media that were conditioned by MEF prior to culture of the hESC. Here, FM was used to monitor the OCT4-eGFP Knock-In hESC line in both conditioned and unconditioned medium (UCM) with and without the addition of bFGF. The commercially available Essential 8™ medium was used as a positive control (42). Our results indicated that all media can retain the hESC in their undifferentiated status during at least 6 days. These findings were corroborated by FC measurements at day 6. FC (after FM analysis) and Celigo® however, did show a higher cell number at day 6 with hESC grown in CM compared with UCM. This observation was not seen by analyzing hESC with FC (no FM analysis) alone. FM analysis and Celigo will induce some stress that can probably be avoided by using an incubator system with auto-fluorescent free medium (riboflavin is for example a well-known auto-fluorescent substance) (138). More colonies can then also be monitored. Indeed, software for automatic measurement of colony fluorescence is already under construction by companies like Brooks (Celigo™ system). However, because of the difference in half-life between eGFP (half-life ≥ 24 hours) and OCT4 (half-life in mice = 6-8 hours), the fluorescent signal follows the OCT4 expression and only the absence of a signal can directly be interpreted as a lack of OCT4 in the cell (139-141).

Of note, a disadvantage of the FM method described here is that for larger colonies, several images need to be taken of each colony, which subsequently need to be stitched to visualize the whole colony in a single image. In our hands, a lower magnification objective with a lower numerical aperture (NA) (2.5x, NA = 0.15 instead of 10x, NA= 0.50) could not be used as this led to lower fluorescent signals of the colonies, resulting in unusable s/n ratios. Stitching results in an increased handling time, therefore increasing the stress that these cells need to undergo. This is because media needed to be removed to avoid auto-fluorescence and the microscope used during these experiments was not equipped with an incubation chamber.

Conclusion

We describe in detail a method for the use of fluorescence microscopy (FM) to monitor the (non-) differentiation status of human embryonic stem cell (hESC) colonies by using a commercially available OCT4-eGFP Knock-In hESC line. We focus on the pitfalls and the benefits of this non-invasive screening method by testing it against flow cytometry (FC) as gold standard. FM is capable of following the (non-)differentiation status of different colonies during several days, but has the added value of observing morphological changes indiscernible by FC. Together with complementary FC data, such as cell number and eGFP/cell, this provides an additional dimension in defining the (non-)differentiation status of a culture.

This optimized FM setup can be used to analyze the impact of different media on the (non-) differentiation status of the hESC line growing on vitronectin.

CHAPTER 4: SAMPLE PREPARATION PROTOCOL FOR PROTEOMICS

Minimizing technical variation during sample preparation prior to label-free quantitative mass spectrometry

E. Scheerlinck; M. Dhaenens; A. Van Soom; L. Peelman; P. De Sutter; K. Van Steendam*; D. Deforce[a]*

* Authors contributed equally

Published in *Anal Biochem* **2015**, 490, 14-19

Abstract

Sample preparation is the crucial starting point to obtain high quality mass spectrometry data and can be divided into two main steps in a bottom up proteomics approach: cell/tissue lysis with or without detergents and a(n) (in-solution) digest comprising denaturation, reduction, alkylation/blocking and digesting of the proteins. Some important considerations herein are, amongst others, that the reagents used for sample preparation can inhibit the digestion enzyme (e.g. 0.1 % (w/v) SDS & 0.5 M guanidine HCl), give rise to ion suppression (e.g. PEG), be incompatible with LC-MS/MS (e.g. SDS) or can induce additional modifications (e.g. urea). Taken together, all these irreproducible effects are gradually becoming a problem when label-free quantitation of the samples is envisioned, such as during the increasingly popular UDMS^{E} and SWATH DIA strategies. Here, we describe the detailed validation of a reproducible method with sufficient protein yield for sample preparation without any known LC-MS/MS interfering substances by using 1 % SDC both during cell lysis and in-solution digest.

Introduction

Sample preparation is a critical step towards high quality LC-MS/MS data in proteomics. In addition, not only protein identification, but also repeatability between samples becomes very important when using label-free strategies such as UDMS^E and SWATH DIA (142). When starting from intact cells for a proteome analysis, sample preparation most often involves the use of a surfactant to increase the protein recovery during cell lysis. Commonly used surfactants for cell lysis, prior to mass spectrometry, are TritonX, NP40, SDS, SDC and 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS). NP40 and TritonX are non-ionic, non-denaturant surfactants which are chosen for mild cell lysis when conservation of native biological structure is required. Sodium dodecyl sulphate and sodium deoxycholate, on the other hand, are ionic and denaturant surfactants which can disrupt cell membranes and can cause protein denaturation by breaking protein:protein interactions. The zwitterionic and non-denaturating surfactant CHAPS disrupts protein aggregates and is most often used for 2D gel electrophoresis instead of ionic surfactants (74). An important disadvantage of the use of TritonX/NP40 is that these surfactants are composed of PEG structures (143). Polyethylene glycol, a hydrophobic agent, can give rise to ion suppression at the ion source of a mass spectrometer and can be observed in the mass spectrum as repeating elements of 44Da (144, 145). For this reason, removal of PEG is required resulting in sample loss and (possible) loss of repeatability in the case of label-free quantitative analysis. Sodium dodecyl sulphate also has disadvantages: (1) it denatures enzymes such as trypsin leading to impaired digestion, (2) it is liquid chromatography incompatible and (3) it causes ion suppression. Despite these unwanted side effects, SDS is still used on a regular basis for cell lysis and digestion prior to mass spectrometry. Although SDS can be removed after digestion by filter aided sample preparation (FASP) as described by Wisniewski et al. and Shevchenko et al. (146, 147), it has been reported that this time-consuming method was not able to deplete all SDS, still causing LC-MS problems (147, 148). Additionally, reproducible results, which are crucial for label-free quantitation during e.g. HDMS^E or SWATH are difficult to obtain with the FASP protocol (149). CHAPS equally is MS incompatible (ion suppression) and sample clean-up must be performed by for example C_{18} Zip Tips (150). Finally, SDC

also needs to be removed prior to MS, but this can be done by either acid precipitation or two-phase solvent extraction after digesting. After these removal steps no LC-MS/MS interference is detected (76, 151), but at least for two-phase solvent extraction higher variability in peptide and protein identification rate has been described in comparison with acid precipitation (152). For this study, a detergent with denaturant characteristics (SDC and SDS) was chosen for addition to the cell lysis buffer because one can expect that a loss in the native conformation of proteins will lead to more protein identifications in a bottom-up proteomics approach.

After cell lysis, extracted proteins are cleaved into peptides by means of a digesting enzyme, mostly trypsin. The addition of a denaturant in this step will keep hydrophobic proteins in solution and denature proteins making the cleavage by a digesting enzyme more efficient. The effect of different denaturants on protein denaturation and solubilization during digestion has been extensively studied and denaturants can be grouped as surfactants (SDS/SDC), chaotropic agents ((thio)urea) and solvents (methanol/acetonitrile). Because of the above mentioned problems with SDS, different companies have developed MS-compatible surfactants by (1) making them easily removable after digestion by acid precipitation before MS analysis (Rapigest^{TM}, PPS silent surfactant^{TM}, Protease Max^{TM}) or (2) by assuring that the surfactant did not co-elute with the peptides on a C^{18} reverse column liquid chromatography system (Invitrosol^{TM}) (81, 153-155). Although these surfactants are able to improve the digest efficiency of different proteins in comparison to no addition of any denaturant, they are expensive compared to SDS or SDC. In 2007, Masuda et al. (156) compared 27 additives, analyzing the effect on the solubilization of a membrane fraction derived from both *E. coli* and HeLa cells prior to digestion. SDS gave the best result on protein yield determined with bicinchoninic acid assay, followed by Rapigest^{TM} and SDC (76, 156). Proc et al. (2010) compared 14 different digesting protocols on their efficiency to digest soluble human plasma proteins. In particular, proteins resistant to digestion (myoglobin for example) showed a better digestion efficiency with SDS and SDC both in 4, 9 or 16 hours digesting time compared to urea or combinations of methanol with SDC or trifluoroethanol. They were the first to also consider repeatability, which scored best for SDC while the lowest repeatability

was observed for urea, which indicates once more the advantage of the use of SDC in $UDMS^E$ and SWATH above others (79). Finally, Leon et al. (2013) analyzed the digestion efficiency of an in-solution digest of 1 % (w/v) Rapigest, 8 M urea or 5 % (w/v) SDC for denaturation and solubilization of proteins (denaturants were diluted when trypsin was added to the sample). Sodium deoxycholate in general scored best for peptide/protein identifications and protein sequence coverage (152).

Taken together, these reports indicate that the low-cost denaturants, SDS and SDC, seem to be the best additives to be used for cell lysate in combination with subsequent digestion. In this report, we focus for the first time on the effect of using these reagents throughout the whole protocol, starting from cell lysis all the way to the final peptide samples. Hereby, we specifically focus on the repeatability of these approaches without substantial loss in protein identification, to assure their compatibility with emerging label-free quantitation strategies such as $UDMS^E$ and SWATH. A reproducible protocol was hereby accomplished by means of using one type of buffer and detergent throughout the whole sample preparation protocol (from cell lysis to mass spectrometry analysis).

Material & Methods

1.1 Materials

All products were purchased from Life Technologies (Carlsbad, CA, USA) unless stated otherwise.

1.2 Cell culture of cell lines (RAJI/HeLa/THP1)

Two suspension cell lines (RAJI/THP1) and one adherent cell line (HeLa) were cultured to confluence in a T175 flask (37 °C, 5 % CO_2) using medium composed of basal medium supplemented with 10 % fetal bovine serum, 100 u/ml penicillin, 100 µg/ml streptomycin and 2 mM L-glutamine. Dulbecco's modified Eagle's medium (DMEM), DMEM/F12 and Roswell Park Memorial Institute Medium 1640 were used as basal medium for respectively RAJI, HeLa and THP1 culture.

1.3 Cell lysis

After cell counting with a hemocytometer, four million cells were washed two times with 1x phosphate buffered saline and were subsequently lysed in a protein-low bind Eppendorf with 50 mM triethylammonium bicarbonate (TeABC, Sigma-Aldrich, St. Louis, MO, USA), supplemented with 100 units Benzonase Nuclease (Sigma-Aldrich, St. Louis, MO, USA) and 1x Halt Protease and Phosphatase Inhibitor Cocktail (PerBio, Erembodegem, Belgium) whether or not in combination with a denaturant (4 % (w/v) SDS (MP, Illkirch, France), 1 % (w/v) SDC, 4 % (w/v) SDC or 10 % (w/v) SDC (Sigma-Aldrich, St. Louis, MO, USA)).

Cells were vortexed and subsequently sonicated (Transsonic 460, Elma) for 10 minutes on ice.

After centrifugation (10 minutes at 17968xg), the supernatant was used for further analysis.

1.4 Analysis of the protein concentration

Protein concentration was determined by means of absorbance at 280 nm with a Nanodrop ND-1000 Spectrophotometer (Thermo Scientific, Rockford, USA). UV measurement was performed since no Coomassie results could be obtained because of incompatibility with SDC and SDS.

Statistical analysis was performed using Graph Pad Prism 5 (San Diego, CA, USA). Differences were evaluated by a Student's t-test. A p-value < 0.05 was considered statistically significant.

1.5 Trypsin digest

First, compatibility of 1 % (w/v) SDC with other digesting reagents was analyzed. Compatibility of 1 % (w/v) SDC with a digestion reagent was defined when no white precipitation in a blank sample (50 or 500 mM TeABC, 1 mM $CaCl_2$ (Sigma-Aldrich), 5 % acetonitrile (Biosolve)) was observed. Different reducing (10 mM DTT vs 5 mM TCEP (Sigma-Aldrich) and blocking/alkylating agents (10 mM MMTS vs 20 mM IAM (both from Sigma-Aldrich)) were added to a blank sample and visually checked for precipitation.

THP1 cell lysate was digested overnight at 37 °C in 500 mM TEABC, 1 % SDC (w/v), 1 mM $CaCl_2$, 5 % acetonitrile and trypsin/lysC (25:1 protein-enzyme ratio w/w; Promega, Madison, WI, USA) after reduction with 10 mM DTT for 60 minutes at 60 °C and blocking with 10 mM MMTS for 10 minutes at room temperature.

1.6 SDC removal by means of acid precipitation or two-phase solvent extraction

Removal of SDC after a trypsin digest from THP1 cell lysates with 1 % SDC was obtained by acid precipitation (pH=2) with 2 % (v/v) trifluoroacetic acid (TFA, Sigma-Aldrich) or by two-phase solvent extraction with ethyl acetate (1:1) followed by addition of 2 % (v/v) TFA. After centrifugation, the supernatant of the acid precipitation and the aqueous phase (lower phase) in the two-phase system contained the peptides and were transferred to another Eppendorf. The precipitates of acid precipitation and organic phase (after vacuum evaporation) in the two-phase system were washed with 3x 0.5 % (v/v) TFA. All samples were dried afterwards. Each removal protocol was performed on 5 replicas.

1.7 LC-UDMS^E

After digestion, dried peptides were dissolved in H_2O with 0.1 % (v/v) formic acid. Peptides were separated on a NanoACQUITY system (Waters Corp., Manchester, UK) with direct injection on a NanoACQUITY UPLC® column (1.7 µm BEH130 100 µm x 100 mm C18) at a flow rate of 300 nL/min. The column temperature was maintained at 35 °C. The LC-gradient (1 %- 40 % B in 60 min followed by 7 min on 85 % B) was obtained by a combination of mobile phase A (H_2O + 0.1 % (v/v) formic acid + 3 % (v/v) dimethyl sulfoxide) and mobile phase B (acetonitrile + 0.1 % (v/v) formic acid). All samples were analyzed by HDMS^E with an in-house optimized collision energy look up table (ultradefinition mass spectrometry), (97)) on a Synapt G2Si instrument (Waters Corporation). Therefore, ion-mobility dependent collision energy profiles (look up table) in the transfer region are assigned to each individual ion mobility separation cycle across the full ion mobility separation range. The ion mobility separation wave height was set to 40 V. All analyses were performed in resolution mode with a scan time of 0.8sec. Mass accuracy was maintained using a lock spray with glufib (*m/z* 785.8426, 100 fmol) and leu-

enkephalin (*m/z* 556.2771, 200 pg) with a flow rate of 0.5 µl/min. Traveling wave velocity was ramped from 1200 m/s to 400 m/s over the full ion mobility separation cycle. Wave heights in the trap and transfer were both set to 4 V and wave velocities to 311 m/s and 190 m/s respectively. In low and high energy mass spectrometry mode, the collision energy was set to 4 eV in the trap region.

1.8 UDMS^E data analysis

Uniform optimal processing parameters (low energy, high energy, intensity) for UDMS^E analysis were first determined with Protein Lynx Global Server Treshold Inspector and the data was subsequently analyzed with Progenesis 2.0 software (Waters). First, retention time correction between samples had to be performed. This was accomplished by the alignment of each sample run to a home-made quality control sample run, created by generating an equal mixture of all samples. Subsequently, peak picking was performed and data were filtered by charge state (only 2-4+ features were held for analysis). Next, normalization was performed to all proteins. After processing, the data were searched against a human databank with methylthio (on cysteine) as fixed modification and deamidation (on asparagine and/or glutamine) and oxidation (on methionine) as variable modifications. The enzyme specificity was set to trypsin, with maximum 1 missed cleavage. False discovery rate was set to 4 %, corresponding to a peptide score threshold in our search environment of ± 5.4. Two peptides were required to identify a protein.

1.9 Cell lysate analysis: addition of no detergent in comparison with addition of 1 % SDC

Possible protein/peptide differences between a cell lysate with or without 1 % (w/v) SDC were analyzed with LC-UDMS^E. In short, the same amount (µL) of both types of THP1 cell lysates (n = 3 per condition (1 % (w/v) SDC/no detergent)) was digested as described in 1.5. Each sample was digested in duplo (6 samples/condition). SDC was removed by acid precipitation as described in 1.6. No pellet wash was performed. Peptides were analyzed with LC-UDMS^E (1.7). Normalization was performed against all proteins. Data analysis was performed with Protein Lynx Global Server. Only common proteins (in 2 or

more replicas per condition) were retained to define differences in protein/peptide identifications in both conditions.

Possible differences in abundances between the same proteins and peptides were identified with Progenesis 2.0 software. An independent T-test with false discovery rate correction was performed with Excel. A Q-value ≤ 0.001 was considered statistically significant. All significant peptides/proteins were further analyzed on possible differences in their hydrophobicity (GRAVY) by means of Prot Param software (157).

Results

1. Cell lysis: the need for a detergent

Sodium dodecyl sulphate and sodium deoxycholate are two of the most used detergents for cell lysis. To validate their added value in terms of yield and repeatability, protein concentration was determined in triplicate by a UV measurement assay (280 nm) because colorimetric assays are incompatible with detergents. Cell lysis without any detergent, with different concentrations of SDC (1 % (w/v), 4 % (w/v) or 10 % (w/v)) and with 4 % (w/v) SDS (based on FASP protocol [6]) were compared after cell lysis in 2 suspension cell lines (RAJI and THP1) and 1 adherent cell line (HeLa) to coordinately exclude any cell line specific effects. Indeed, a significant increase in protein yield in both suspension cell lines (THP1, RAJI) for even the lowest concentration of SDC (1 % (w/v)) could be found (Figure 1). Compared to 1 % (w/v), addition of 4 % (w/v) or 10 % (w/v) SDC had no added value in increasing the protein content. When comparing 1 % SDC with 4 % (w/v) SDS (FASP protocol), one could observe a higher protein yield with the use of 1 % (w/v) SDC in comparison with 4 % (w/v) SDS (p-value < 0.05 in RAJI, p-value = 0.06 in THP-1 cell lysate) (**Figure 1**). We concluded that 1 % (w/v) SDC could be used as detergent during cell lysis and can be used as a substitute for 4 % (w/v) SDS.

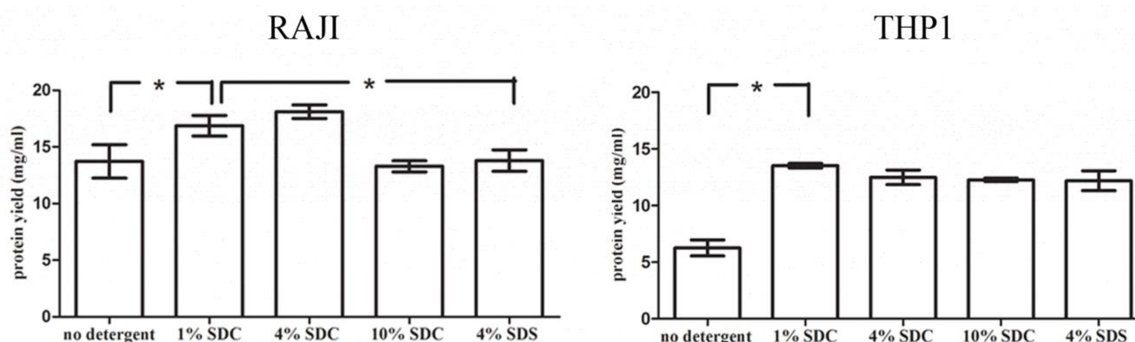


Figure 1. Overview optimization cell lysis protocol for RAJI and THP1 cells. One representative of each experiment is shown. UV measurement results concerning the effect of SDC and SDS on cell lysis in 2 cell lines. *= significant differences (p-value < 0.05) between no detergent and 1 % (w/v) SDC or 1 % (w/v) SDC and 4 % (w/v) SDS.

For the adherent HeLa cell line, 0.25 % trypsin-EDTA was used to detach cells prior to cell lysis. However, the addition of 1 % (w/v) SDC gave no significant higher protein yield compared to no detergent in this cell line: in multiple experiments the addition of 1 % (w/v) SDC gave rise to equal or higher protein yield compared to no detergent. This larger variability in protein yield between experiments might be due to differential clustering of cells after detachment from the culture plate which can interfere with subsequent protein extraction efficiency. Therefore, direct cell lysis of adherent cells without the detachment with 0.25 % trypsin-EDTA was analyzed as well. Interestingly, when cells were directly lysed from the washed plate using the different amounts of detergent more reproducible results were obtained compared to cell lysis after cell detachment with trypsin. Importantly, protein yield after direct cell lysis in 1 % (w/v) SDC was comparable to the 'standard' procedure with the 0.25 % trypsin-EDTA step.

2. <u>Trypsin digest optimization of a cell extract with SDC</u>

2.1 Compatibility of SDC with chemicals needed for trypsin digest

During the subsequent steps of the digest, we noticed that precipitation occurs by the addition of some reagents. Since SDC precipitates in acid environment it is important to avoid fluctuations in the buffer pH. We thus tested the use of TCEP or DTT as reducing agent as well as IAM or MMTS as alkylating

reagents in a 50 mM or 500 mM TeABC buffer environment. Sodium deoxycholate precipitation occurred when either TCEP or MMTS were added to the blank sample (in 50 mM TeABC, (1 mM $CaCl_2$, 5 % (v/v) ACN)). Indeed, these solutions (made in respectively 50 or 500 mM TeABC) have a pH of 1 and 4 respectively. However, precipitation could be avoided for all tested reducing and alkylating agents when using 500 mM TeABC. Increasing the buffer capacity is thus strongly recommended when using SDC for sample preparation.

Of note, using 500 mM TeABC instead of 50 mM TeABC also during cell lysis would greatly increase the simplicity of the protocol, resulting in better repeatability. While it is known that osmolarity can theoretically have an influence on cell lysis (74), in our hands no significant difference in protein yield could be observed in any of the cell lines tested above when using 50 or 500 mM TeABC. In conclusion, we recommend the use of 500 mM TeABC in both cell lysis and in-solution digest.

2.2 SDC removal by means of acid precipitation or two-phase solvent extraction

After the tryptic digest, the SDC needs to be removed. Using the THP-1 and RAJI cell line, we compared the use of acid precipitation (AP) and two-phase solvent extraction (PT) in terms of protein and peptide identification efficiency. As can be seen in **Figure 2** for the THP-1 cell line, no significant differences were detected in the number of proteins (white bars) and peptides (grey bars) between PT and AP. Yet, the use of AP produced a more reproducible list of identifications (**Supplementary Table 1**). Of note, when washing the pellet, as suggested by Lin et al. (2010), a small number of proteins/peptides could be identified in the pellet wash of both PT and AP. However, no new peptide identifications could be detected in these pellet washes and we thus discarded this additional step.

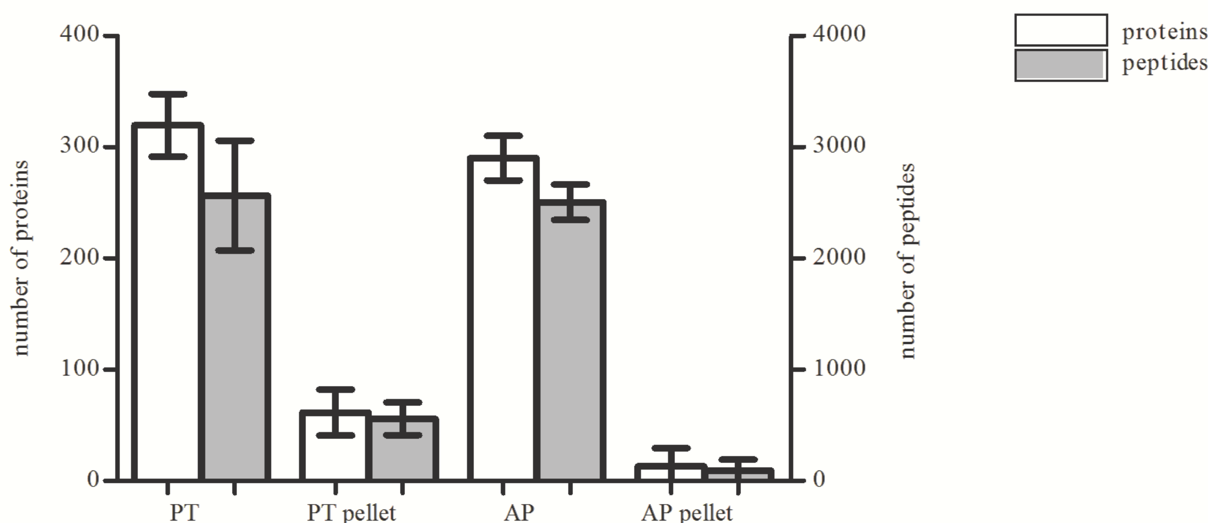


Figure 2. SDC removal by means of two-phase solvent extraction (PT) or acid precipitation (AP) for THP1 cell lysates. The need for a pellet wash was also taken under consideration (PT pellet and AP pellet). The number of proteins (left axis, white bars) and peptides (right axis, grey bars) of the different conditions are presented. Each removal protocol was performed on 5 samples.

3. Proteomic analysis of cell lysates obtained with or without 1% (w/v) SDC

In a final analysis, identical amounts of THP1 cell lysates with and without 1 % (w/v) SDC were analyzed with HDMS^E. Indeed, only small normalization factors were calculated when normalization was done against all proteins. Surprisingly, the same number of protein/peptide identifications in both conditions was observed (with around 83 % of all identified proteins common between both conditions). The identification efficiency (% annotation) was also the same in both conditions. However, in the SDC samples, additional unidentified precursor masses (10 %, charge 2-4+) were found at the peptide level. By defining the repeatability at the level of feature intensity, we can directly define the technical variability that would actually interfere with each of the features present in a sample. Within the different replicas (n= 6) in THP1 cell lysates with and without 1 % (w/v) SDC, we therefore calculated the relative standard deviation (RSD) of all features (**Figure 3**). RSD was determined by dividing the standard deviation by the mean of the normalized abundances of all replicas/condition for all features separately. Out of this data, a frequency plot was made. As can be seen in Figure 3, over 60 % of all features had a RSD lower than 20 % in a cell lysate with 1% SDC (grey bars). In contrast, the same RSD was achieved for ± 38 % of all features for a cell lysate with no detergent (white bars), indicating a higher repeatability

when using 1% (w/v) SDC. While protein identification is a metric that is used in most protocol optimizations, the above results illustrate that this actually only shows a limited picture.

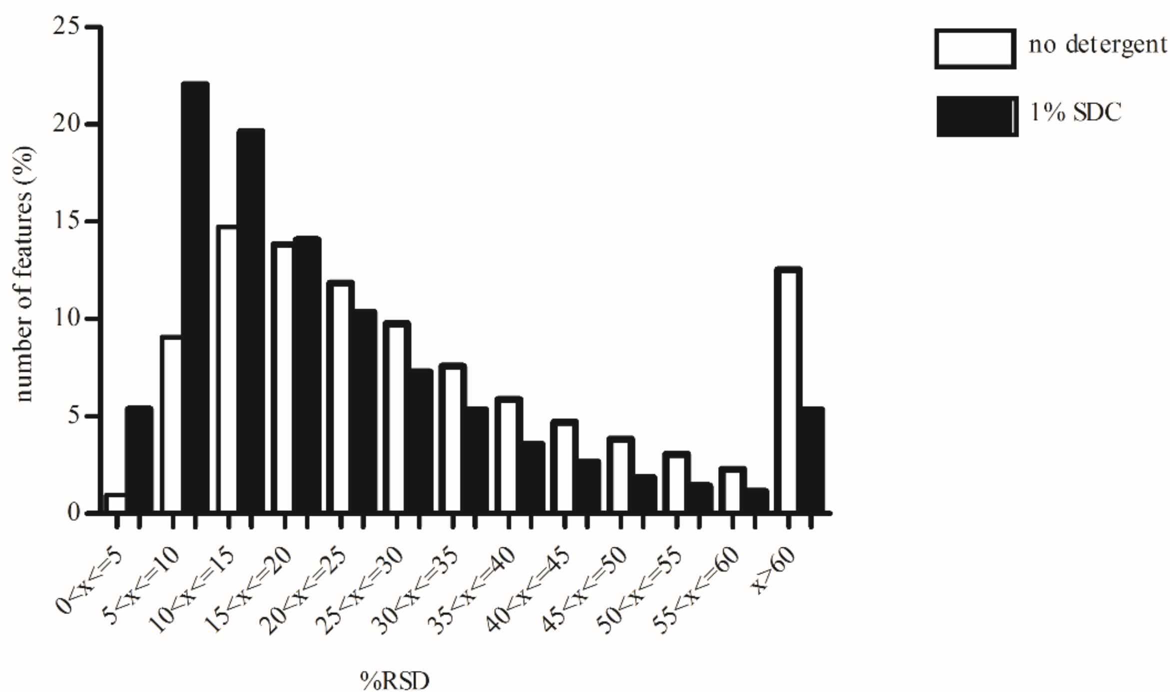


Figure 3. Frequency plot considering repeatability. The x-axis represents the relative standard deviation (stdev normalized abundances/mean normalized abundances of all replicas/condition x 100) (RSD), the y-axis represents the number of features with this RSD (%). Cell lysates with 1 % (w/v) SDC (grey bars) have a higher repeatability in comparison with cell lysates without 1 % (w/v) SDC (no detergent, white bars).

Discussion

The majority of studies focusing on optimizing sample preparation use the number of peptide or protein identifications as the metric of validation. However, repeatability in sample preparation is of main importance when using label-free quantification approaches like $HDMS^E$ and SWATH (142). This repeatability during sample preparation can only be obtained by using a protocol with a minimum of steps. Here, we present a reproducible protocol by using a single buffer and a single detergent throughout the entire protocol.

Tris(hydroxymethyl) aminomethane(TRIS)-HCl, a commonly used buffer, was not the first choice because of its reported ion suppression effect, the formation of TRIS-protein adduct ions and its

incompatibility with iTRAQ (contains primary amines) (158). TeABC on the other hand, can be evaporated by means of vacuum drying and is compatible with iTRAQ analysis. Therefore, TeABC is widely used as digesting buffer (pH 8.0). Because of its suitability during in-solution digestion, this buffer is also the first choice for cell lysis. In literature, the concentration of TeABC in digest protocols is not uniform (concentration between 50 mM and 500 mM). However, we observed precipitation of SDC when MMTS and TCEP were added to an in-solution digest containing 50 mM TeABC. Because the change in osmolarity does not change cell lysis efficiency, we here argue for the use of 500 mM TeABC where no precipitation was observed. Of note, because of the known side reactions with IAM (N alkylation and O alkylation), MMTS was used as blocking agent in an in-solution digest (80).

Next, the use of detergents for cell lysis was validated. A detergent is added to the cell lysis buffer for several reasons: (1) solubilization of hydrophobic proteins or membranes or (2) denaturation of proteins (breaking protein:protein interactions) (74). Several detergents are available; each having its own advantages and disadvantages. In our experiment, SDC and SDS were chosen because of their denaturant characteristics, low cost, and promising results based on digest optimizations (76, 146) . Different concentrations (1-4-10 % (w/v)) of SDC were compared with 4 % (w/v) SDS (amount used during FASP (146)) and no detergent addition. In general, one could observe a reproducible and significant higher protein yield with 1 % (w/v) SDC in comparison with no detergent or 4 % (w/v) SDS addition. Therefore, SDC can be considered as an alternative for SDS for cell lysis. Next, we have shown that higher repeatability between experiments is obtained in adherent cell lines when using direct cell lysis instead of first using 0.25 % trypsin-EDTA for detaching the cells. Direct cell lysis is therefore recommended for the use in HDMS^{E} and SWATH data-independent acquisition strategies.

After cell lysis, a trypsin digest is carried out to obtain peptides which can be analyzed in LC-MS/MS. As for cell lysis itself, different denaturants can be added to promote the unfolding of proteins. We reasoned however, that using the same buffer as for the cell lysis would benefit repeatability. Indeed, one must keep in mind that addition of denaturants needs to be done with care: (1) inhibition of trypsin activity: 0.1 % (w/v) SDS, 4 M urea and 50 % methanol will lead to a trypsin activity of respectively

20, 71 and 31 % (153); (2) incompatibility with LC-MS/MS: SDS gives rise to ion suppression (76) and (3) introduction of modifications: high temperature will convert (thio)urea to cyanate resulting in carbamylation of lysine, arginine and N termini (146). One percent SDC was chosen as denaturant during digestion because of the already reported promising results and its LC-MS/MS compatibility (76, 151, 152, 156).

Removal of SDC can be achieved by AP with TFA or by PT by means of 1:1 ethyl acetate with TFA. In short, addition of TFA (=AP) leads to SDC precipitation since SDC is insoluble in an acid and aqueous environment. During PT, SDC solubilizes in the organic solvent (ethyl acetate) while the peptides remain in the aqueous part. Masuda et al. (2007) found that more peptides (±32 %)/proteins (±37 %), in particular hydrophobic peptides/proteins, could be identified in an *E. coli* membrane fraction by means of PT compared to AP. They hypothesized that hydrophobic proteins will precipitate with SDC when using AP (156). In contrast, Lin et al. (2010) found that more hydrophilic and hydrophobic peptides and proteins (±11-12 %) could be identified in rat liver membrane with AP in comparison with PT (151). In both articles, no information concerning repeatability is available of the comparison between AP and PT (151, 156). Leon et al. (2013) compared the repeatability of PT and AP in a rat liver mitochondrial sample: no difference was found between AP and PT at the protein level. In contrast, more peptides (11.51 %) were found in PT in comparison with AP resulting in slightly higher protein coverage for PT. Higher repeatability was observed in AP (152). Our comparison led to the conclusion that no difference between protein or peptide numbers could be observed between AP and PT. A slightly higher repeatability appeared to be present with AP. These results are in general a confirmation of the results as described in Leon et al. (2013) (152).

In a final experiment, the impact of the addition of 1 % (w/v) SDC during cell lysis of THP1 cells on protein identification and quantification was examined with $HDMS^E$. The same proteins and peptides were identified with and without SDC. An increased coverage will undoubtedly be obtained by using a longer LC gradient. Higher repeatability was observed in a cell lysate with 1 % (w/v) SDC (lower %

RSD) than without SDC, which is a major advantage when using $HDMS^E$ and SWATH data-independent acquisition strategies.

In conclusion, a sample preparation protocol is presented here with good repeatability and protein yield by using the same buffer (500 mM TeABC) and same detergent (1 % (w/v) SDC) starting from cell lysis to $HDMS^E$ analysis.

Conclusion

We have demonstrated that the addition of 1 % (w/v) SDC to a cell lysis buffer resulted in a higher and more reproducible protein yield in comparison with no detergent addition in 3 different cell lines, making it the most recommend method for $HDMS^E$ and SWATH data-independent acquisition strategies.

Acknowledgements

This research was (partially) financed by GOA- project no. 01G01112 – Pathways to pluripotency and differentiation in embryos and embryonic stem cells.

Supplementary data

Supplementary data can be found online on:

http://www.sciencedirect.com/science/article/pii/S0003269715003942

Of note, supplementary data was not mentioned in the hard-copy of the thesis. The data, consisting out of a list of all identified proteins in the different samples amongst others, was too large to display in the thesis and was not necessary to understand this chapter.

CHAPTER 5: SILAC OPTIMIZATION IN HESC

The development of a fully defined SILAC culture medium with minimal arginine conversion in human embryonic stem cells

Based on *"The development of a fully defined SILAC culture medium with minimal arginine conversion in human embryonic stem cells".* Ellen Scheerlinck; Katleen Van Steendam; Simon Daled; Elisabeth Govaert; Liesbeth Vossaert; Paulien Meert; Filip Van Nieuwerburgh; Ann Van Soom; Luc Peelman; Petra De Sutter; Björn Heindryckx; Maarten Dhaenens*; Dieter Deforce*.

Submitted to *J. of Prot. Res*

* Authors contributed equally to this study

Abstract

We present a fully defined culture system (adapted Essential8^{TM} (E8^{TM}) medium in combination with vitronectin) for hESC that can be used for Stable Isotopic Labeling of Amino acids for Cell culture (SILAC) purposes. Although a complete incorporation of the labels was observed after 4 days in culture, only 7.39 % of all mass spectrometry (MS) precursors displayed a conversion of L-arginine (R) to L-proline (P) or L-glutamate (E) where the converted peak was less than 10 % of total peak intensities of that peptide. To reduce this arginine conversion, E8^{TM} medium was modified by adding (1) L-proline, (2) L-ornithine, (3) N^{ω}-hydroxy-nor-L-arginine (Nor-NOHA) acetate or by (4) lowering the arginine concentration. Reduction of arginine conversion was best obtained by adding 5 mM L-ornithine, followed by 3.5 mM L-proline and by lowering the arginine concentration in the medium to 99.5 µM. No major changes in the proteome, pluripotency and cell amount could be observed for these adapted Essential8^{TM} media with ornithine and proline. A sudden cell death however, was observed with the use of 99.5 µM L-arginine. In conclusion, we suggest using 5 mM L-ornithine to reduce arginine conversion.

Introduction

Stable Isotopic Labeling of Amino acids in Cell culture (SILAC), developed in 2002 by the lab of M. Mann, is used to quantify differences in protein abundance between two cell culture conditions by means of incorporating stable isotopically labeled or "heavy" amino acid(s) (AA) in a culture (e.g. $^{13}C_6$ lysine in one culture and $^{12}C_6$ lysine (K) in the other) (111). $^{13}C_6$ lysine and $^{13}C_6$ arginine are commonly used as heavy AA, in order to quantify every peptide during tandem mass spectrometry since most digests are performed with trypsin (115, 159).

Importantly, incomplete incorporation of heavy AA can lead to skewed light-over-heavy ratios. For most cell types, five population doublings are therefore performed for near-complete incorporation (~97 %) of the heavy AA into the cell proteome (115).

A second problem is the metabolic conversion of AA during SILAC experiments resulting in incorrect light-over-heavy ratios. Metabolic conversion of L-arginine (R) to L-proline (P) and, to lesser extent, L-glutamate (E) generates (an) extra peak(s) of + 5 Dalton per heavy P or E in the peptide in the MS spectrum (159). Especially fast metabolizing cell types like human embryonic stem cells (hESC) suffer greatly from this arginine conversion problem (116). Arginine titration as well as a proline titration can be used to reduce this arginine conversion (115, 116). In hESC cultures however, lowering the arginine concentration can induce differentiation and cell death (116). For this reason, P titration was used as a solution to overcome this problem, since the addition of P can theoretically reduce its formation out of R (116). However, these reports used undefined culture conditions such as conditioned medium from mouse embryonic fibroblasts or an undefined coating like Matrigel™, making it impossible to determine the actual amount of AA present in these media (116, 160, 161). Additionally, the impact of these approaches on pluripotency or the rest of the proteome was not assessed. To date, no fully defined culture medium for SILAC application is available for hESC, which makes it difficult for laboratories to use SILAC directly out of literature for their own applications. Recently, Essential8™ (E8™) medium in combination with vitronectin was developed for culturing hESC in a full xenofree and defined way (42).

Here, we present the comparison between different arginine conversion reducing strategies on hESC in a fully controlled experimental pipeline, coordinately monitoring pluripotency and general proteome changes. First, we used a completely defined culture for SILAC by respectively adding different concentrations of L-proline, L-ornithine or Nor-NOHA acetate to the E8 culture medium or by lowering the L-arginine concentration in this medium. Second, by using an OCT4–eGFP Knock-In hESC line (WiCell Research Institute, Madison, WI, USA), cell number and pluripotency could be monitored based on OCT4 expression during the time of the experiment by means of flow cytometry (FC) and fluorescence microscopy (FM). FM allows non-invasive monitoring of pluripotency and simultaneous verification of colony morphology to assess general cell culture health (162). Within the time span of the experiment (4 days, complete heavy label incorporation) no differences in OCT4 expression and cell number were seen when different concentrations of L-ornithine and L-proline was added to the media. However, when lowering the arginine concentration to 99.5 µM, a complete cell loss during the experiment was observed in 2 out of 4 experiments. Addition of Nor-NOHA gave no effect on OCT4 expression, but a significant increase in cell number was observed in comparison with the control (no adaptation of the E8 medium). Subsequently, the effect of the different adapted media on the reduction of the arginine conversion were analyzed. Herein, 5 mM ornithine, 3.5 mM proline and 99.5 µM arginine were the best in reducing the arginine conversion. Finally, these three adapted culture conditions were further analyzed concerning their possible effect onto the proteome and were compared to our control (no adaptation of the E8 medium). The analysis of the possible effects onto the proteome was obtained by means of a label-free data independent acquisition approach named UDMS^{E}. This technique is similar to the MS^{E} parallel fragmentation approach but uses ion mobility as an extra dimension of separation, greatly increasing peak capacity and specificity for linking fragments and precursor masses during data analysis. This analysis revealed that none of the culture conditions had a major effect within this time span of 4 days, but the minor differences found suggest that these effects might well exacerbate over longer time periods. In depth analysis of these differences showed that histones and tubulins were upregulated in all conditions (in comparison with control), pointing towards a potential impact on

mitosis/protein folding and chromatin organization. The addition of L-proline to the cell culture resulted in downregulation of pathways related to RNA degradation, heat stress and TGFβ-signaling.

Material & Methods

1.1 Materials

All products were purchased from Life Technologies (Carlsbad, CA, USA) unless stated otherwise.

1.2 Feeder-free culture of hESC

WA01 OCT4-eGFP Knock-In hESC (Wicell Research Institute, Madison, WI, USA) were plated on a pre-coated xenofree vitronectin (VN XF, Primorigen Biosciences, Madison, WI, USA) 6-well plate (coating concentration = 0.5 µg/cm^2) and cultured in E8^{TM} medium (37 °C, 5 % CO_2 and 5 % O_2). E8^{TM} medium was made by diluting Essential 8^{TM} 50x supplement 1:50 with “arginine and lysine free” DMEM/F12 (Thermo Scientific, Rockford, IL, USA) supplemented with 398 µM L-arginine HCl and 499 µM L-lysine HCl (both from Sigma-Aldrich, St. Louis, MO, USA). Splitting was performed every 4-5 days with 0.5 mM ETDA in Dulbecco’s Phosphate Buffered Saline according to the manufacturer’s protocol of culturing hESC in E8^{TM} medium.

1.3 SILAC labeling of hESC and culture conditions

1.3.1 Incorporation of heavy labels into the proteome

For labeling, E8^{TM} medium was supplemented with $^{13}C_6$ lysine and $^{13}C_6$ arginine (both from Thermo Scientific) in the same concentrations as the light variant. hESC were harvested after 4, 5 or 6 days in culture (3, 4 or 5 population doublings in which a population doubling is defined as a doubling of the amount of cells between 2 consecutive days) to examine the time needed for a full incorporation of heavy labels. No splitting was performed during the time of the experiment. Media were changed daily. The experiment was performed in triplicate.

1.3.2 Inhibition of the arginine conversion to P and E

The composition of the heavy labeled E8™ medium was changed to examine the effect on inhibition of arginine conversion. Four different culture conditions were examined: (1) addition of L-proline (3.5-6.9-10.4-13.9 mM) (Sigma-Aldrich), (2) L-ornithine HCl (0.05-0.5-1-2-5 mM) (Sigma-Aldrich), (3) Nor-NOHA acetate (50-100 µM) (Enzo Life Sciences, Farmingdale, NY, USA) and (4) L-arginine concentration was decreased from 398 µM (concentration in no adapted E8 medium) to respectively 199 and 99.5 µM. hESC grown in heavy isotopically labeled E8™ medium were used as a control for all SILAC related experiments.

The experiment was done in triplicate for flow cytometry analysis (FC) and a fourth well was used for fluorescence microscopy (FM) analysis to assess colony morphology (M&M 1.4).

1.4 OCT4-expression & cell count

The influence of the different conditions on hESC differentiation was examined by analyzing OCT4-expression by means of FM and FC.

Daily monitoring of OCT4-expression was assessed by non-invasive monitoring by means of FM. To assess influence of stress, a plate without daily monitoring (FM) but with FC at the end of the experiment was investigated to obtain information about OCT4 expression/cell and cell number as described earlier (162).

1.5 Cell lysis and digest

After culture medium removal, cells were detached with 0.25 % trypsin-EDTA (4 min., 37 °C) and trypsin was subsequently inactivated with trypsin inhibitor (1:1 w/w) (Sigma-Aldrich). 10 % of the cells were transferred to another Eppendorf for flow cytometry analysis. After centrifugation (200xg, 5 min.), cells were resuspended in Phosphate Buffered Saline with 1 % (w/v) bovine serum albumin and kept on ice until flow cytometry analysis. After centrifugation (200xg, 5 min.) and a wash step with 1x

Phosphate Buffered Saline, remaining cells (90 %) were lysed in a protein LoBind Eppendorf containing 250 µL of 500 mM TeABC (Sigma Aldrich) supplemented with 1x Halt protease & phosphatase inhibitors (Thermo Scientific), 125 units benzonase nuclease (Sigma-Aldrich) and 1 % (w/v) SDC (Sigma-Aldrich) as a detergent. The Eppendorf was vortexed (30 sec) and subsequently sonicated (10 min., on ice, Transsonic 460, Elma).

Cell lysates were centrifuged for 10 min at 17968xg and supernatant was further analyzed. Protein content of the supernatant was determined by means of a Coomassie Bradford Assay (standard curve obtained using bovine serum albumin (0-2000 µg/ml in 10 times diluted cell lysate buffer), Thermo Scientific). The cell lysate was digested overnight at 37 °C in 500 mM TeABC, 1 % SDC (w/v), 1 mM $CaCl_2$, 5 % acetonitrile (v/v) and trypsin/lysC (25:1 protein:enzyme ratio w/w; Promega, Madison, WI, USA), after reduction with 10mM DTT for 60 min at 60 °C and blocking with 10mM MMTS for 10 min at room temperature. Sodium deoxycholate was subsequently removed by means of acid precipitation. Detailed information about this method is described in (163).

1.6 LC-MS/MS

After vacuum drying in a Centrivap®, peptides were dissolved in H_2O with 0.1 % (v/v) formic acid. A trapped HPLC system, Dionex Ultimate 3000 (Thermo Scientific), was used to separate the peptides (1 µg loaded) on an Acclaim PepMap 100® C18 column (75 µm x 25 cm) (Thermo Scientific) at a flow rate of 0.3 µL/min. The LC-gradient used for elution was obtained by a combination of mobile phase A (H_2O + 0.1 % (v/v) formic acid) and mobile phase B (80 % (v/v) acetonitrile + 0.1 % (v/v) formic acid): 4 % B- 100 % B in 66 min. Data directed acquisition (DDA) on a Triple TOF^{TM} 5600 mass spectrometer (Sciex) with a NanoSpray source operating in positive ESI mode was used to assess MS and MS/MS data in dynamic accumulation mode. In short, the scan range for MS ranged from *m/z* 400 to *m/z* 1250 with a 250 ms accumulation time. In MS/MS, a scan range from *m/z* 65 to *m/z* 2000 with a minimum of 25 msec accumulation time was used. Rolling collision energy was used in MS/MS. DDA was triggered

for m/z with a charge state from 2+ to 4+ which exceeds 50 cps. Former target ions were excluded for 30 sec.

1.7 Data analysis of DDA data

1.7.1 Incorporation

RAW DDA data (wiff files) were loaded into Mascot Distiller (Matrix Science) and processed. A Mascot search was subsequently performed with the following parameters: enzyme specificity was set to trypsin with maximum 2 missed cleavages. Methylthio (on cysteine) was used as fixed modification and deamidation (on asparagine and/or glutamine) and oxidation (on methionine) as variable modifications. The precursor tolerance was set to 20 ppm and the MS/MS tolerance to 0.1 Da. Identification was considered positive with a p-value < 0.05. After identification, incorporation rate was determined by analysis of the L/H ratio. This ratio was determined by defining the light component as a peptide having a $^{12}C_6$ arginine or/and $^{12}C_6$ lysine and the heavy component as a peptide having a $^{13}C_6$ arginine or/and $^{13}C_6$ lysine. In addition, the conversion of heavy arginine to heavy proline is taken under consideration by defining the heavy proline as a satellite modification group. The ratio was accepted by Distiller by applying thresholds to 2 measurements: correlation (threshold = 0.9) and fraction (threshold = 0.5).

For incorporation and conversion calculation, identification is of minor importance and analysis was thus extended by using an in-house developed Python script Conversion Finder (CoFi) running at the MS precursor mass level. Following the “Quantify then Identify” principle (quantification-driven analysis), all MS precursors and not only the identified ones can be analyzed in this way to obtain a more complete picture of all precursors which are present in the sample. First, MS precursors exported from the Progenesis software appointed the same retention time (tolerance window: ± 1 min) and same *m/z* (tolerance window: ± 0.05 Da) were excluded from the analysis because these features could belong to different peptides. Secondly, all m/z were sorted from low to high and MS precursors with no full incorporation were isolated by using the formula: *m/z* + 6.02013*x/charge (x = represents the number

of R and K that can be present in the sequence, ranging from 1 to 5) (same retention time (tolerance window = ± 0.1 min).

1.7.2 Quantitative analysis of arginine conversion

1.7.2.1 AUC measurement

RAW DDA data were loaded into Progenesis 2.0 Software (Nonlinear Dynamics, Waters). Data alignment for retention time correction, peak picking and subsequent normalization against all proteins were performed. MS precursors with a charge from 2+ to 4+ were exported to Excel 2010. Because we want to analyze arginine conversion on all MS precursors (Quantify then Identify principle), the CoFi script was used. The formula in the CoFi script was adapted to m/z + 5.0168*x/charge (x= represents the number of P and E that can be present in the sequence, ranging from 1 to 5) because arginine conversion is observed in a mass spectrum as a mass shift of 5.0168 Da between two peaks. Arginine conversion (%) was calculated from the reported normalized abundances by dividing the sum of the normalized abundances of the heavy labeled MS precursors by the normalized abundances of all MS precursors (light + heavy) * 100.

1.7.2.2 *Quantitative analysis by means of Spectral Counting to confirm AUC measurement*

In a second analysis, RAW data were converted to mgf format by Peak View (Sciex) and searched against a human databank using Mascot with the following parameters: enzyme specificity was set to trypsin with maximum 2 missed cleavages. Methylthio (on cysteine) was used as fixed modification and deamidation (on asparagine and/or glutamine), oxidation (on methionine), Label: 13C(5) P, Label : 13C(5) E, Label : 13C(6) K, Label : 13C(6) R as variable modifications. The precursor tolerance was set to 20 ppm and the MS/MS tolerance to 0.1 Da. Only peptides with a p-value ≤ 0.05 were kept for analysis. Arginine conversion (%) of identified peptides was determined by dividing the number of MS/MS spectra that were identified with a P or E heavy label by the total number of MS/MS spectra identified with a sequence containing P and/or E.

1.8 LC-UDMS^E as label-free method to analyze the effect of the conditions on the proteome

After digestion, dried peptides were dissolved in H_2O with 0.1 % (v/v) formic acid. Peptides (100 ng loaded) were separated on a NanoACQUITY system (Waters Corp., Manchester, UK) with direct injection on a NanoACQUITY column (UPLC® 1.7 µm BEH130 100 µm x 100 mm C18) at a flow rate of 300 nL/min. The column temperature was maintained at 35 °C. The LC-gradient (1 % - 40 % B in 60 min followed by 7 min on 85 % B) was obtained by a combination of mobile phase A (H_2O + 0.1 % (v/v) formic acid + 3 % (v/v) dimethyl sulfoxide) and mobile phase B (acetonitrile + 0.1 % (v/v) formic acid). All samples were analyzed by UDMS^E with an in-house optimized collision energy look up table on a Synapt G2Si instrument (Waters Corporation) (97). This method is described in detail in (163).

1.9 UDMS^E data analysis

Uniform optimal processing parameters (low energy, high energy, intensity) for UDMS^E analysis were first determined with Protein Lynx Global Server Treshold Inspector and the data was subsequently analyzed with Progenesis 2.0 software (Nonlinear Dynamics, Waters). Briefly, an intelligent peak-modelling algorithm using a wavelet based approach identifies individual precursor peaks retaining all relevant quantification and positional information. To combine and compare results from different runs, Progenesis QI aligns them to compensate for between-run variation. This results in increased reliability and reproducibility of the peptide abundance measurements.

Subsequently, peak picking was performed and data were filtered by charge state (only 2-4+ MS precursors were used for analysis). Next, normalization was performed to all proteins. After processing, the data were searched against a human databank (SwissProt) with methylthio (on cysteine) as fixed modification and deamidation (on asparagine and/or glutamine) and oxidation (on methionine) as variable modifications. The enzyme specificity was set to trypsin, with maximum 1 miss cleavage. False discovery rate was set to 4 % at the protein level, corresponding to a UDMS^E score threshold in our search environment of ± 5.4.

The measurements are then combined into protein measurements. After removal of any outliers in the replicas by means of Multivariate Statistics (Principal Component Analysis, Progenesis software), proteins with a minimum of 2 peptides from which at least 1 was unique, were kept for analysis. A p-value ≤ 0.05 was used to identify proteins which were significantly different between control and condition (L-proline/L-arginine or L-ornithine). These proteins were exported to Excel 2010 and were further analyzed with Reactome (164, 165).

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (166) via the PRIDE partner repository with the dataset identifier PXD002859.

Results & Discussion

1. <u>Incorporation of heavy labels into the proteome</u>

Before any SILAC analysis can be performed, it is not only recommended to check the incorporation of the heavy labels into the proteome, but equally to verify the occurrence of any conversion of the heavy label to another amino acid (e.g. L-arginine conversion to L-proline/L-glutamate) (115). While some corrective data analysis tools are available, full incorporation without conversion remains the preferred point of departure to obtain correct quantitative information. WA01 OCT4-eGFP Knock-In hESC grown in E8™ medium were therefore analyzed after 4, 5 or 6 days in culture (3, 4 or 5 population doublings). Importantly, no measurable effect on OCT4 expression was observed when using heavy arginine in comparison to light arginine. Cell number analysis showed a population doubling rate of 24 hours between two consecutive days (*data not shown*).

Full incorporation was defined when no light counterpart of a peptide could be identified or for which the light peptide was only ≤ 5 % of the total peak area (heavy + light) and was first confirmed with Mascot Distiller by focusing on the roughly 3000 identified peptides. Of note, some of the completely light peptides (2 % in total) were identified as keratins, which are most probably contaminants from sample preparation (167). Importantly, identification is not essential for the calculation of incorporation and conversion. On balance, only an estimated 16 % of peptide-like precursors gets selected for

fragmentation during a conventional DDA run and no more than 58 % of all MSMS spectra gets annotated with a false discovery rate of 1 % (96). Additionally any bias that is introduced by sequence-dependent fragmentation and/or prior knowledge in e.g. selecting search parameters such as variable modifications can be avoided. Indeed, using an in-house developed Conversion Finder (CoFi) script, all of the 22.000 peptide-like features in the samples could be interrogated for incorporation by looking for any mass pair differing 6.02013 Da. Here too, over 97 % of features had no mass pair and was thus completely heavy (or entirely light). The same incorporation rate was found at day 4, 5 and 6. Of note, for the annotated features ratios from the CoFi script and from Distiller were very comparable (*data not shown*). In conclusion, 3 population doublings (4 days during experiment) is sufficient to obtain a full heavy label incorporation into the WA01 OCT4-eGFP Knock-In hESC proteome and this is used in the following experiments.

2. Arginine conversion

With complete incorporation being achieved, arginine conversion was investigated as second possible bottleneck when using SILAC. Arginine conversion can now be quantified on all MS precursors without prior bias introduced by MS/MS acquisition or identification, as it can be described as an intensity ratio of co-eluting pairs of precursors differing m/z + 5.0168*x/charge (x=1-5), wherein x represents the number of P and E that can be present in the sequence. Remarkably, only 7.39 % of all MS precursors suffered a R to P or E conversion of less than 10 % of the total MS intensity of that peptide, showing that arginine conversion in the human embryonic stem cell line is a major issue (Figure 4). In addition, around 50 % of all tryptic peptides (between 700-6000 Da) contain at least one proline according to the human database of the international protein index (116). For this reason, it is important to inhibit this arginine conversion. Analysis of the pathway mediating the conversion from R to P and E (**Figure 1**) indicates that arginine conversion could be inhibited by (1) the addition of L-proline or (2) the addition of L-ornithine to the media or (3) by decreasing the arginine concentration (115, 116, 168). Another solution is to inhibit arginase, an enzyme that is responsible for the conversion of arginine to ornithine.

Arginase can be inhibited by Nor-NOHA (169). Nor-NOHA works as a competitive inhibitor of arginase at physiological pH and is around 40 times more potent in inhibiting arginase than the endogenous formed N^{ω}-hydroxy-L-arginine in murine macrophages (169, 170). This inhibition of arginase theoretically blocks the conversion of L-arginine to L-ornithine. L-ornithine is amongst others a precursor of polyamines, needed for cell proliferation in a growing fetus (170).

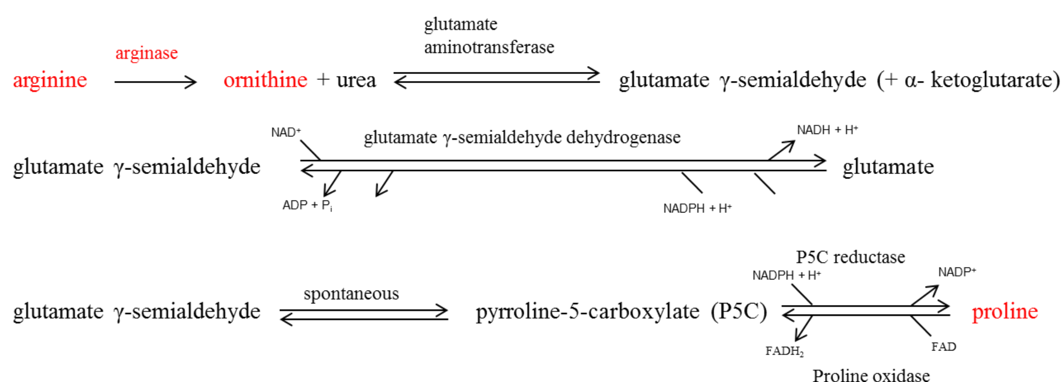


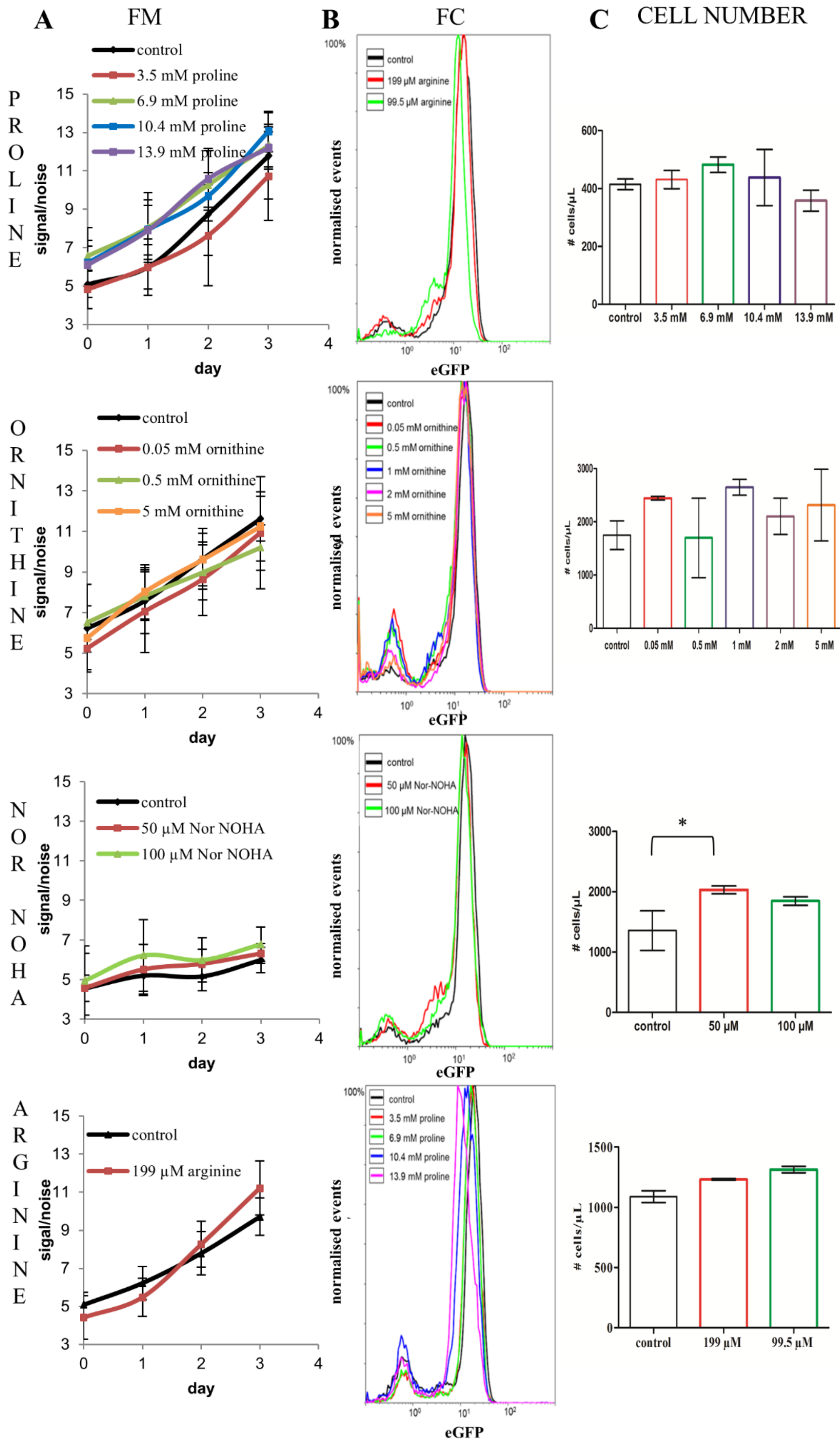
Figure 1. Arginine conversion. L-arginine is converted to L-proline and L-glutamate by several intermediate steps. The steps in which we intervene for inhibiting the arginine conversion are indicated in the figure as bold red.

2.1 Effect of the different conditions on OCT4 expression and cell number

Because hESC are known to start differentiating spontaneously upon changes to their culture environment, the different conditions (L-proline/L-ornithine/Nor-NOHA acetate/L-arginine) were first analyzed for their effect on the differentiation status of the hESC by means of an optimized screening method using FM and FC (**Figure 2**) (162). A low eGFP expression (fluorescence signal $<10^1$), correlating to a low OCT4 expression, indicates differentiation. When using FM, both colony shape and fluorescence signal/surface unit can be measured in a non-invasive manner. This can give important additional information on the impact of an experiment on hESC. In such experiments the overall fluorescence of the colonies increases with time when no differentiation is induced, as we described earlier (**Figure 2A**) (162). Normal E8™ medium was used as a control. No decrease in OCT4 expression or change in colony shape was observed with FM in all tested conditions within the analyzed time period of 4 days, needed for complete AA incorporation. For FC, only a small but insignificant loss in eGFP signal could be seen (OCT4 expression) at L-proline concentrations above 10.4 mM.

When assessing the effect of different treatments on cell number (**Figure 3C**), addition of 50/100 µM Nor-NOHA acetate displayed a positive effect on cell growth (p-value ≤ 0.05). The inhibition of arginase would theoretically lead to an increase in arginine. Deamidation of arginine by nitric oxide synthase will form nitric oxide (NO). Dependent on the available concentration of NO, NO promotes cell proliferation (picomolar and nanomolar) or induces cell arrest (micomolar) (171). The cell proliferation as observed in our experiment can therefore be explained if NO is indeed present in low concentrations. Of note, a higher cell amount with Nor-NOHA acetate was also reported in neural stem cells (172). When decreasing the arginine concentration, complete loss of cells in 1 or more wells was observed in 2 out of 4 experiments. No significant effect on cell growth could be observed for the addition of proline and ornithine to the media.

In conclusion, a decrease in arginine concentration can lead to complete cell loss and the addition of Nor-NOHA acetate can have an influence on cell growth, making these two conditions the least favorable options to reduce arginine conversion during a SILAC experiment.



A
FM
B
FC
C
CELL NUMBER
PROLINE
control
3.5 mM proline
6.9 mM proline
10.4 mM proline
13.9 mM proline
signal/noise
day
control
199 µM arginine
99.5 µM arginine
normalised events
eGFP
cells/µL
control
3.5 mM
6.9 mM
10.4 mM
13.9 mM
ORNITHINE
control
0.05 mM ornithine
0.5 mM ornithine
5 mM ornithine
1 mM ornithine
2 mM ornithine
0.05 mM
0.5 mM
1 mM
2 mM
5 mM
NOR NOHA
control
50 µM Nor NOHA
100 µM Nor NOHA
50 µM Nor-NOHA
100 µM Nor-NOHA
*
50 µM
100 µM
ARGININE
control
199 µM arginine
sigal/noise
199 µM
99.5 µM

Figure 2. Analysis of the different conditions (L-arginine/L-proline/L-ornithine/Nor-NOHA) on OCT4 expression (FM (A) and FC (B)) and cell number (C). (A) FM results of the different conditions. The y-axis represents the mean of the fluorescence (eGFP signal) of 5 hESC colonies/surface unit (= background) (signal/noise ratios); the x-axis represents the day of the experiment. A loss in eGFP expression and thus of signal/noise indicates loss of pluripotency. Because of the 3D growth of hESC colonies, an increase in signal/noise is observed during the time of the experiment in undifferentiated colonies as described in Scheerlinck et al., 2014 (162). In none of the conditions, significant loss in eGFP was observed. (B) FC results. At the last day of the experiment (day 4), a part of the cells were analyzed with flow cytometry in addition to fluorescence microscopy analysis to confirm FM results and to obtain information concerning cell amount. The expression of eGFP signal in all tested conditions was measured (a minimum of 10 000 events was analyzed). No significant loss of eGFP in any of the conditions was observed in comparison with the control, confirming FM results. (C) cell number results (FC, day 4). The amount of cells/µL (y-axis) was determined by adding a known number of fluorescent beads as spike-in to the FC samples. Only the addition of Nor-NOHA induced a significant increase in cell number. Inconsistent results were observed when arginine concentration was decreased. The asterisk depicts a p-value ≤ 0.05 (by unpaired T-test).

2.2 Inhibition of the arginine conversion

Next, the CoFi script was used to analyze all MS precursors for possible arginine conversion which is not possible with Mascot to obtain a more complete picture. All conditions were assessed with DDA for their ability to inhibit the arginine conversion. A frequency plot (**Figure 3**) of all MS precursors (identified and non-identified) was made. Arginine conversion was most reduced by 5 mM L-ornithine (56.85 % of all MS precursors with ≤10 % arginine conversion), followed by L-proline (40-45 % of all MS precursors) and 99.5 µM L-arginine (33.30 % of all MS precursors). Surprisingly, Nor-NOHA acetate has no effect on inhibition of the arginine conversion. Although spectral counting of identified spectra could be considered less accurate, the reduction in arginine conversion is so intense that the trend is reflected even at the level of the identified spectra (**Supplementary Data Figure 1**). In addition, the data presented in Figure 4 were examined in more detail. Herein, the conversion rate of peptides containing multiple converted prolines was compared to all identified peptides (**Supplementary Data Figure 2***)*. In the control, a higher conversion rate for peptides containing multiple P was observed in comparison to all peptides. The reduction of the arginine conversion by the adapted culture conditions showed the same results: the main reduction was observed with 5 mM ornithine followed by proline and

99.5 µM arginine. Of note, 99.5 µM arginine was less able to reduce the arginine conversion rate in peptides containing multiple P in their sequence.

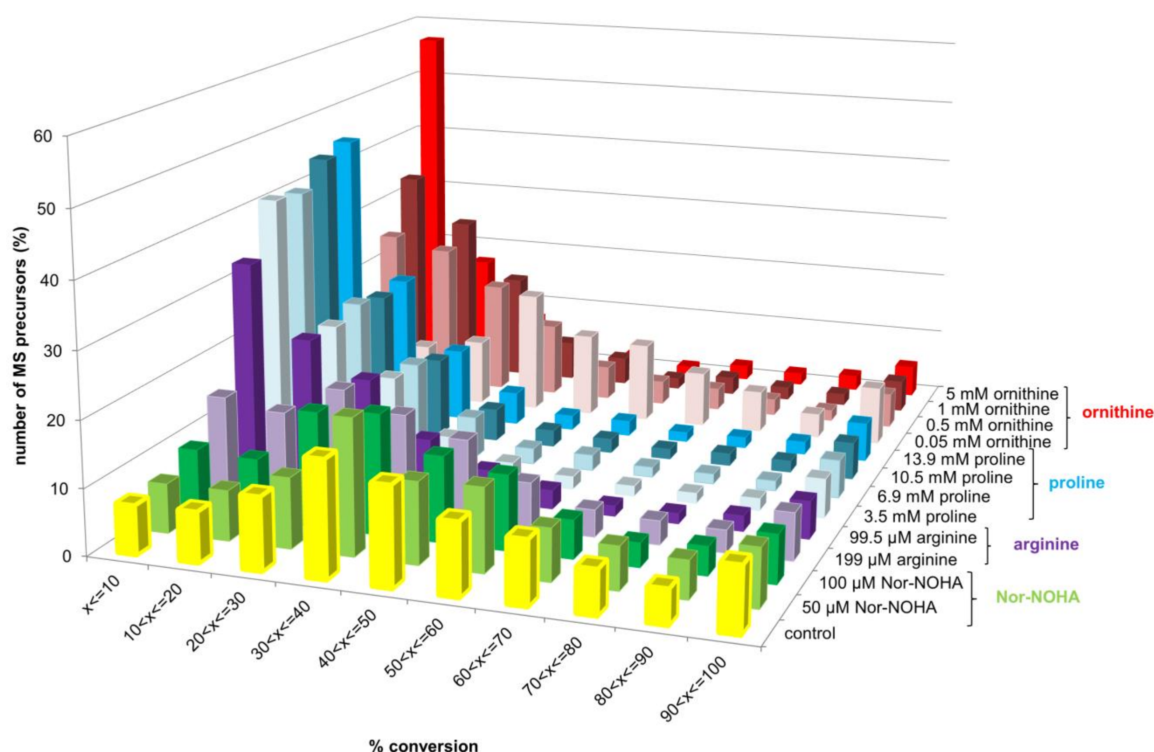


Figure 3. **Overview of the impact of the different conditions on arginine conversion**. The graph represents the relative number of MS precursors (y-axis) with x % conversion. The percentage conversion is defined as the intensity ratio of co-eluting pairs of precursors (= identified and non-identified peptides) differing *m/z* + 5*x/charge (x=1-5) wherein x represents the number of P and E that can be present in the sequence and was automatically analyzed with the CoFi script. Conversion was most inhibited by means of 5 mM ornithine (pink-red bars, 56.85 % of all MS precursors with ≤10% arginine conversion), followed by addition of proline (blue bars, 40-45 % of all MS precursors) and 99.5 µM L-arginine (purple bars, 33.30 % of all MS precursors). 50 -100 µM Nor-NOHA acetate (green bars) has no effect on inhibition of the arginine conversion as compared to the control (yellow). (control= E8 medium in combination with vitronectin)

As a validation of the CoFi script, 5 peptides in all conditions were manually analyzed in Peak View to investigate the arginine conversion by means of summed intensities in which one representative peptide is shown in **Supplementary Data Figure 3**. Similar results are obtained with Peak View in comparison

with the CoFi script. The same trend was observed in the other 4 peptides (*data not shown*) and confirmed the results as described above.

Of note, L-arginine can theoretically be formed out of L-proline and L-ornithine. Addition of these two (light) amino acids to the culture medium can in this way lead to a decrease in incorporation efficiency) (116). However, in our experiment, the incorporation efficiency was the same for every condition (~97 %).

3. After 4 days - only minor changes in the proteome are induced by different treatments

Based on the findings as described above, 5 mM L-ornithine/3.5 mM L-proline and 99.5 µM L-arginine were repeated to analyze the effect of the different treatments on the hESC proteome. HDMS^E was used as a label-free quantification technique to obtain quantitative data that were further analyzed with Reactome to reveal affected pathways.

First, the total set of all identified proteins (2008 proteins, analyzed with Progenesis 2.0) was examined with Reactome to define the enrichment of certain pathways simply due to sample preparation. DNA replication (FDR: 9.4E-5) was most enriched during extraction and is thus intrinsically enriched in the dataset.

Next, significantly different proteins (p-value ≤ 0.05) between control (no change to the medium) and condition (addition of either ornithine and proline concentration or decrease in arginine concentration) were isolated from Progenesis QI 2.0 as follows. First, outliers in the technical replicate were excluded from the analysis by means of PCA (173). Secondly, only proteins identified with a minimum of 2 peptides of which minimum 1 peptide was unique for quantitation were retained (1096 proteins (54.58 % of all identified proteins)). Finally, significant different proteins (p-value ≤ 0.05) between control and condition were analyzed. Only small differences were observed: 69/82/188 proteins (6.30/7.48/17.15%

of all quantifiable proteins (min 2 peptides of which 1 unique)) were significantly different between L-arginine/L-ornithine/L-proline and control, respectively. In depth analysis of these potentially different proteins and pathway(s) was performed with Reactome wherein only those pathways were considered in which all proteins (with a minimum of 3) were consistently up- or downregulated. Histones (H2B and H4) and tubulins were upregulated in all conditions in comparison with the control. These proteins are involved in different pathways such as mitosis, protein folding and chromatin organization. Other pathways that were potentially upregulated in comparison with the control were pathways involved with mRNA splicing (up in L-proline and L-ornithine) and membrane trafficking (up in L-arginine and L-ornithine). In addition, the enzyme pyrroline 5-carboxylate reductase 1, responsible for the conversion of pyrroline 5-carboxylate to L-proline, was upregulated when L-arginine was reduced in the medium (compared to control). Downregulated pathways were only found in the condition with L-proline: pathways involved in RNA degradation, cellular response to heat stress and TGFβ receptor complex signaling were possibly affected. All proteins (with their respectively normalized abundance) involved in all above mentioned pathways can be found in **Table 1**.

Longer incubation periods are needed to analyze more profound effects on the proteome, but our data suggest that only minor changes in the proteome are found at the onset of a SILAC experiment on hESC if incorporation took 3 doublings.

Table 1. Significantly different proteins (p-value ≤ 0.05) in one condition relative to control. Data were obtained as follows: first, $HDMS^E$ data were analyzed with Progenesis 2.0 software (Waters). Only the proteins, identified with a minimum of two peptides of which one peptide was unique are retained for analysis. Subsequently, only the significantly different proteins (p-value ≤ 0.05) in the conditions (proline, arginine or ornithine) relative to control (no change to the medium) were further analyzed with Reactome to identify potentially up- or downregulated pathways. In the table, the different proteins (with their normalized abundance value) belonging to the upregulated or downregulated pathways are displayed (in relative to control). Normalized abundances of which the p-value between the condition and control (t-test) was greater than 0.05 (insignificant) are displayed in bold.

mitose/protein folding/chromatine organisatie * relative to control

Accession number	Name	Normalized abundance (mean ± stdev)			
		proline	arginine	ornithine	control
P07437	Tubulin beta chain	13.15 ± 0.78	**13.80 ± 1.63**	14.08 ± 1.04	**10.50 ± 1.73**
Q13885	Tubulin beta-2A chain	7.53 ± 0.52	8.65 ± 0.54	7.75 ± 0.35	6.27 ± 0.77
Q9BVA1	Tubulin beta-2B chain	9.36 ± 0.74	9.43 ± 0.67	9.04 ± 0.79	7.25 ± 0.68
P04350	Tubulin beta-4A chain	0.40 ± 0.16	**0.32 ± 0.25**	**0.47 ± 0.28**	**0.24 ± 0.05**
Q9BUF5	Tubulin beta-6 chain	**0.44 ± 0.05**	**0.43 ± 0.06**	0.47 ± 0.07	**0.38 ± 0.04**
P62805	Histone H4	37.31 ± 2.45	34.33 ± 1.21	**33.22 ± 3.44**	**25.56 ± 4.85**
P33778	Histone H2B type 1-B	7.26 ± 0.82	7.16 ± 0.63	7.79 ± 0.68	5.70 ± 0.74
P42677	40S ribosomal protein S27	3.61 ± 0.23	**2.94 ± 0.36**	**3.33 ± 0.37**	**3.06 ± 0.27**
P06493	Cyclin-dependent kinase 1	3.39 ± 0.28	**3.16 ± 0.13**	**3.19 ± 0.31**	**2.89 ± 0.20**
P30153	Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform	1.72 ± 0.16	**1.61 ± 0.07**	1.70 ± 0.16	**1.51 ± 0.09**
P20618	Proteasome subunit beta type-1	**1.94 ± 0.44**	2.06 ± 0.41	**1.93 ± 0.37**	**1.50 ± 0.26**
Q5BJF6	Outer dense fiber protein 2	0.80 ± 0.19	**1.11 ± 0.27**	**0.94 ± 0.10**	**1.04 ± 0.12**
P61981	14-3-3 protein gamma	4.05 ± 0.47	**5.16 ± 0.39**	**4.17 ± 1.45**	**4.91 ± 0.45**
Q9UNM6	26S proteasome non-ATPase regulatory subunit 13	4.22 ± 0.34	**4.98 ± 0.41**	**4.74 ± 0.93**	**5.19 ± 0.52**
P25786	Proteasome subunit alpha type-1	1.74 ± 0.15	**2.02 ± 0.18**	**1.97 ± 0.43**	**2.03 ± 0.14**

membrane trafficking

Accession number	Name	Normalized abundance (mean ± stdev)			
		proline	arginine	ornithine	control
Q13885	Tubulin beta-2A chain	7.53 ± 0.52	8.65 ± 0.54	7.75 ± 0.35	6.27 ± 0.77
Q9BVA1	Tubulin beta-2B chain	9.36 ± 0.74	9.43 ± 0.67	9.04 ± 0.79	7.25 ± 0.68
P62158	Calmodulin	**0.30 ± 0.07**	0.40 ± 0.12	**0.27 ± 0.17**	**0.19 ± 0.12**
Q9NP79	Vacuolar protein sorting-associated protein VTA1 homolog	**0.70 ± 0.15**	1.08 ± 0.11	**0.95 ± 0.09**	**0.79 ± 0.14**
Q8NBS9	Thioredoxin domain-containing protein 5	**0.95 ± 0.25**	**0.95 ± 0.30**	1.24 ± 0.09	**0.98 ± 0.11**
Q9BUF5	Tubulin beta-6 chain	**0.44 ± 0.05**	**0.43 ± 0.06**	0.47 ± 0.07	**0.38 ± 0.04**

pyrroline-5-carboxylate reductase 1

Accession number	Name	Normalized abundance (mean ± stdev)			
		proline	arginine	ornithine	control
P32322	Pyrroline-5-carboxylate reductase 1. mitochondrial	**0.24 ± 0.08**	0.39 ± 0.12	**0.21 ± 0.05**	**0.16 ± 0.10**

RNA splicing

Accession number	Name	Normalized abundance (mean ± stdev)			
		proline	arginine	ornithine	control
P22626	Heterogeneous nuclear ribonucleoprotein A2/B1	29.74 ± 5.82	27.23 ± 1.36	28.67 ± 4.31	22.09 ± 2.73
P51991	Heterogeneous nuclear ribonucleoprotein A3	6.17 ± 0.35	4.96 ± 0.48	5.38 ± 1.04	3.81 ± 0.36
Q13242	Serine/arginine-rich splicing factor 9	0.92 ± 0.07	**0.72 ± 0.06**	0.93 ± 0.04	**0.71 ± 0.07**

mRNA stability

Accession number	Name	Normalized abundance (mean ± stdev)			
		proline	arginine	ornithine	control
Q15717	ELAV-like protein 1	2.95 ± 0.21	**3.67 ± 0.15**	**3.23 ± 0.23**	**3.49 ± 0.45**
Q9UNM6	26S proteasome non-ATPase regulatory subunit 13	4.22 ± 0.34	**4.98 ± 0.41**	**4.74 ± 0.93**	**5.19 ± 0.52**
P11940	Polyadenylate-binding protein 1	1.00 ± 0.09	**1.12 ± 0.12**	**1.05 ± 0.05**	**1.14 ± 0.09**
Q92945	Far upstream element-binding protein 2	4.70 ± 0.21	5.45 ± 0.40	4.90 ± 0.64	7.50 ± 1.25
P11142	Heat shock cognate 71 kDa protein	37.83 ± 2.98	**43.59 ± 2.29**	**42.69 ± 3.37**	**43.63 ± 1.24**
P25786	Proteasome subunit alpha type-1	1.74 ± 0.15	**2.02 ± 0.18**	**1.97 ± 0.43**	**2.03 ± 0.14**

cellular response to heat stress

Accession number	Name	Normalized abundance (mean ± stdev)			
		proline	arginine	ornithine	control
P55072	Transitional endoplasmic reticulum ATPase	5.77 ± 0.49	**6.14 ± 0.20**	5.92 ± 0.07	**6.46 ± 0.35**
P11142	Heat shock cognate 71 kDa protein	37.83 ± 2.98	**43.59 ± 2.29**	**42.69 ± 3.37**	**43.63 ± 1.24**
Q02790	Peptidyl-prolyl cis-trans isomerase FKBP4	2.98 ± 0.10	**3.40 ± 0.24**	3.10 ± 0.21	**3.65 ± 0.38**

TGF beta receptor complex signaling

Accession number	Name	Normalized abundance (mean ± stdev)			
		proline	arginine	ornithine	control
P62942	Peptidyl-prolyl cis-trans isomerase FKBP1A	6.02 ± 0.74	**6.56 ± 0.57**	**7.03 ± 1.69**	**7.73 ± 0.91**
Q16254	Transcription factor E2F4	10.25 ± 0.99	**12.27 ± 0.93**	**10.89 ± 2.35**	**13.84 ± 1.95**
O95405	Zinc finger FYVE domain-containing protein 9	0.68 ± 0.04	**0.94 ± 0.17**	**0.97 ± 0.17**	**1.15 ± 0.26**

Conclusion

In conclusion, a fully defined medium is presented to perform SILAC experiments on hESC. With the smallest direct effects on the proteome, OCT4 expression and cell number, we suggest to use 5 mM L-ornithine to reduce arginine conversion in hESC. Thus, this is the first completely defined media for SILAC purposes that can directly be implemented.

Acknowledgements

This research was financed by GOA-project no. 01G01112 – Pathways to pluripotency and differentiation in embryos and embryonic stem cells. MD, EG and PM were respectively supported by a grant from the Fund of Scientific Research Flanders (FWO: 12E9716N and FWO: 3G073112), BOF mandate and a PhD grant from the Institute for the Promotion of Innovation through Science and

Technology in Flanders (IWT-Vlaanderen). P.D.S. is holder of a Fundamental Clinical Research Mandate from the Flemish Foundation of Scientific Research (FWO-Vlaanderen).

We thank Dr. Ir. Koen Decock for writing the Python script, named Conversion Finder that made it possible to automatize data analysis.

Supplementary data

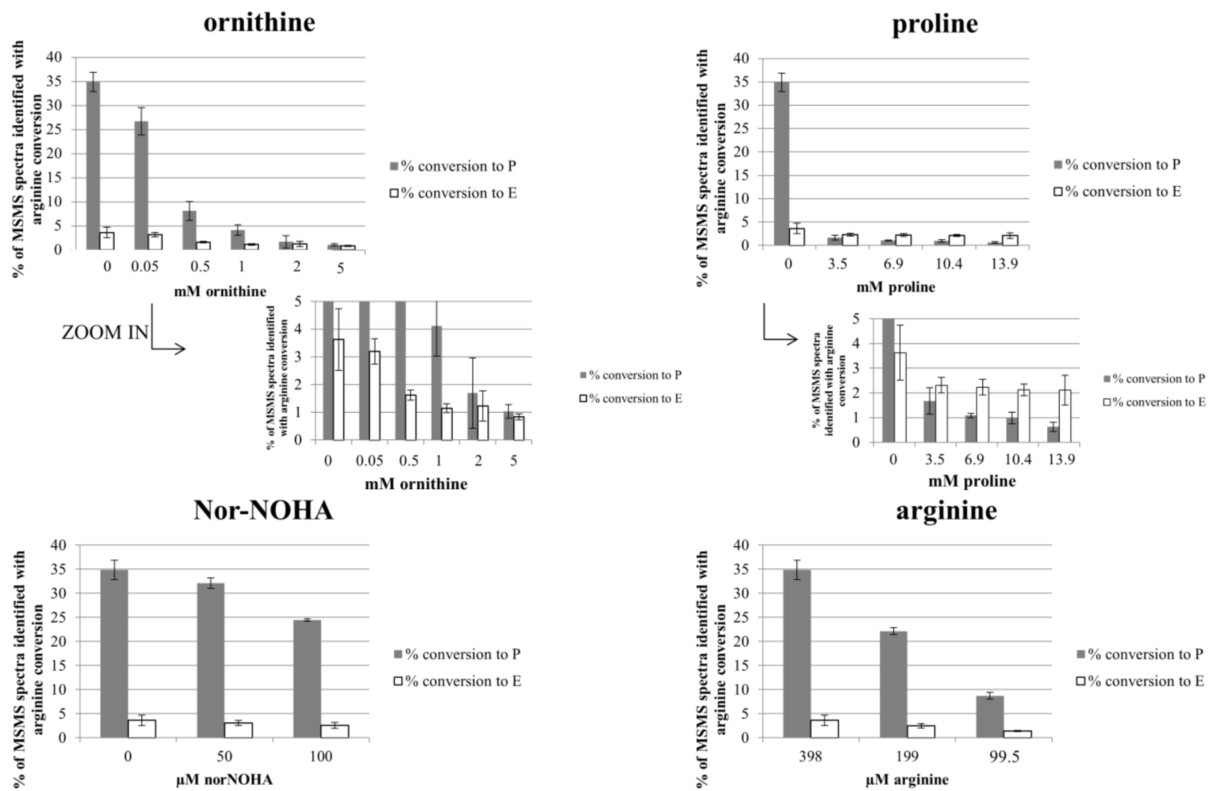


Figure 1. Spectral counting data of all conditions. These data were calculated as follows: first, spectra were annotated using the Mascot search engine (Matrix Science) as described in Experimental Section 1.7. Secondly, the peptides containing a P or/and E in their sequence were isolated for further analysis. Finally, the % of MSMS spectra identified with arginine conversion was calculated by dividing the number of peptides with an identified heavy P or E label to all peptides containing a P or E in their sequence*100. As can be seen in the Figure, the highest reduction in arginine conversion can be obtained by adding 2 mM ornithine or higher or by adding 3.5 mM proline or higher to E8 medium. The lowest reduction in arginine conversion can be observed with the addition of Nor-NOHA. These results confirmed the AUC results.

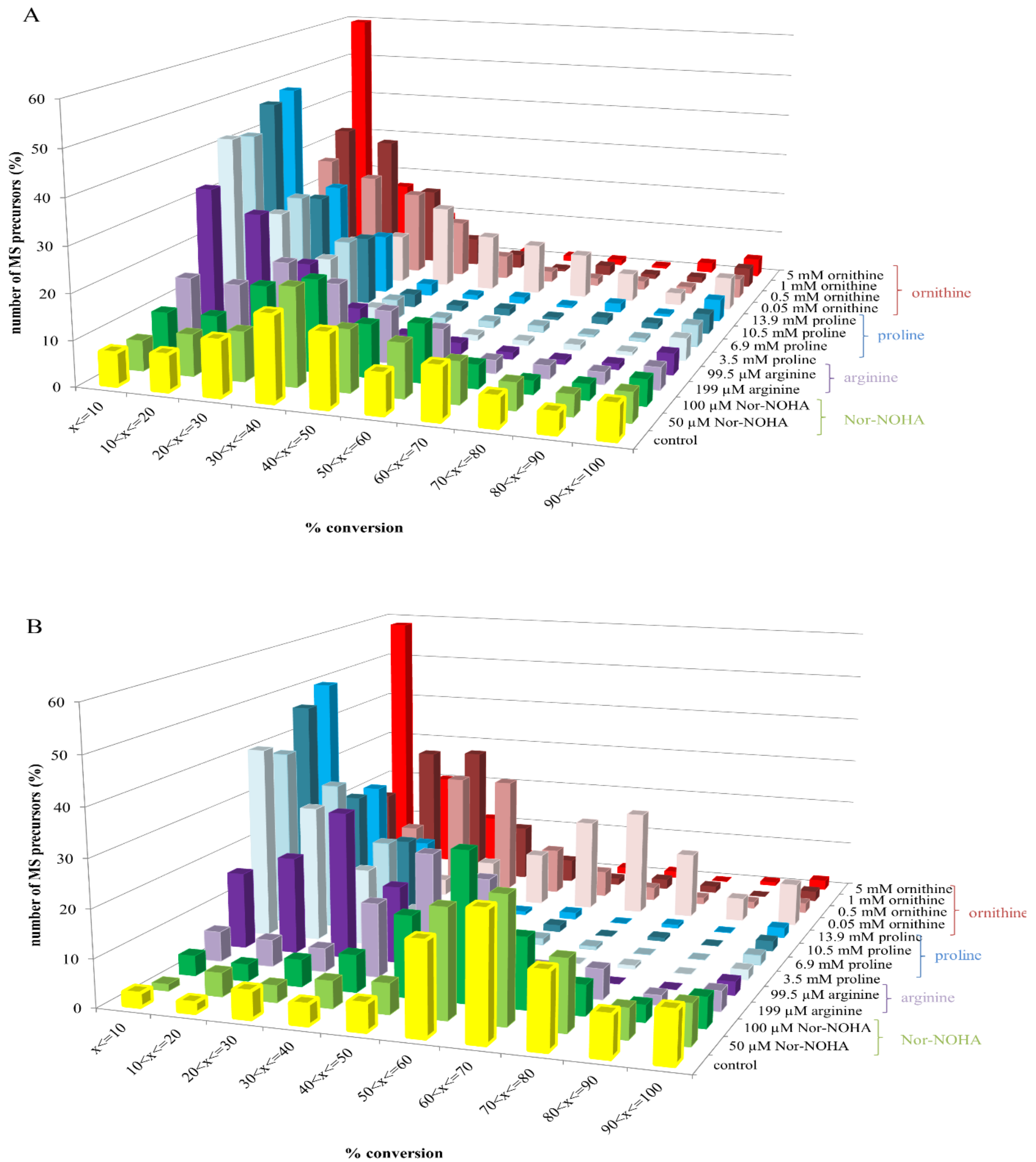


Figure 2. Analysis of the arginine conversion on multiple P in the sequence. A. Representation of the arginine conversion of the identified peptides. B. Representation of the arginine conversion of peptides containing multiple P in their sequence. As can be observed, more conversion of the control (no adaptation of E8 medium) took place in peptides containing more than one P in their sequence in comparison with all peptides. The same trend is observed for the reduction of the arginine conversion: 5 mM ornithine is the best in reduction followed by proline and 99.5 µM arginine. Of note, 99.5 µM arginine was less able to reduce the conversion rate in peptides with multiple P in comparison with all peptides.

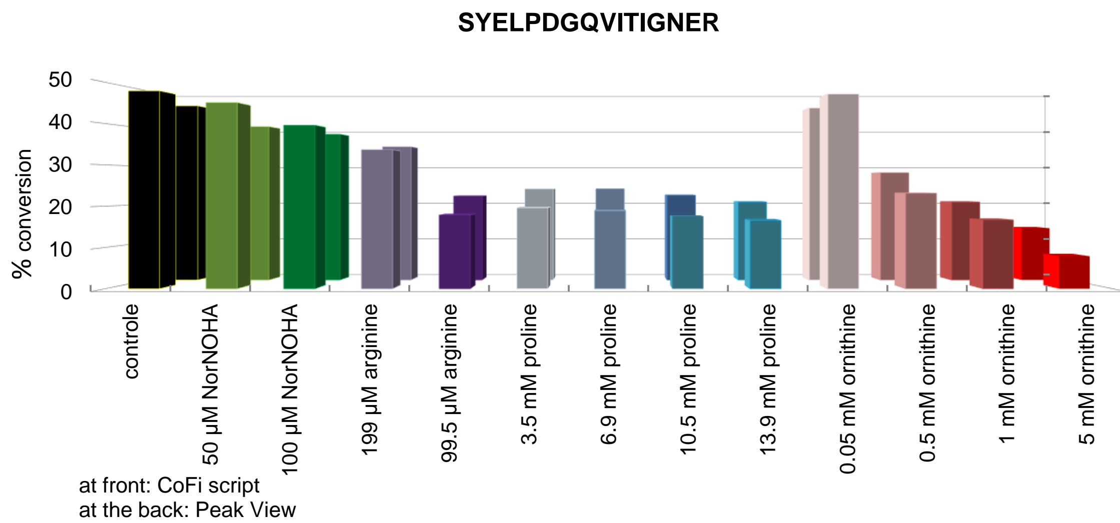


Figure 3. Comparison of the results obtained by means of Peak View (columns at the back) and CoFi script (columns at the front) on the effect of the different conditions for the peptide "SYELPDGQVITIGNER". The same trend can be observed in both data analysis methods: the addition of 5 mM ornithine can reduce the arginine conversion the most, followed by the addition of 3.5 mM proline and the decrease in arginine concentration to 99.5 µM arginine to the E8 medium.

CHAPTER 6

BROADER INTERNATIONAL CONTEXT, RELEVANCE, AND FUTURE PERSPECTIVES:

GOOD PRACTICE IN HESC STUDY

Because of their unique cell characteristics, hESC are considered to be useful for a plethora of applications: (1) to broaden our knowledge considering human development and disease, (2) in cell replacement technologies and (3) in toxicity studies, hereby reducing animal based experiments (174). When studying hESC, it is important to realize that these cells are exceptionally prone to cellular changes following small adaptations to their surroundings. In this dissertation, several tools are provided to both monitor hESC cultures in general and to quantify more specific molecular changes at the proteome level. Together, these results contribute to the necessary experimental knowledge and "good practice" when using hESC in the lab.

The main concern in hESC culture is the maintenance of pluripotency, self-renewal and karyotype. Good cell culture practice requests that these characteristics are routinely examined. Characterization of pluripotency is usually done by investigating the expression of transcription factors and/or surface markers by means of staining (fluorescence measurement by flow cytometry or fluorescence microscopy) or real time PCR (**Chapter 1**). These screening techniques are labor-intensive and expensive. Alternatively, a quick and easy way to check differentiation is by means of studying the colony morphology. hESC form round and flat colonies with defined borders of which the cells are homogenous in shape as shown in Figure 1A. During differentiation, hESC colonies become patchy (**Figure 1B**) or the individual cells start changing shape (**Figure 1C**).

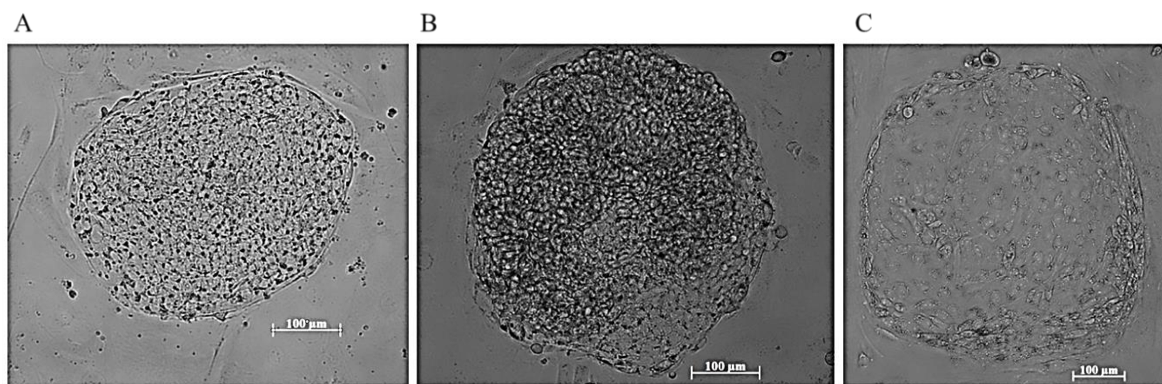


Figure 1 hESC morphology. hESC are kept in culture on inactivated MEF with hESC medium. (A) undifferentiated colony. (B,C) differentiated colonies.

Studying these morphological changes is however subjective and demands a trained eye. In **Chapter 3**, we describe a more straightforward and objective screening methodology for assessing pluripotency by using a commercial OCT4-eGFP Knock-In hESC line (WiCell). Herein, both daily colony morphology and expression of OCT4 (by means of eGFP) can be simultaneously monitored for both feeder (on MEF) and feeder-free (on vitronectin) culture. This method made it for others possible to follow-up the hESC colony not only morphologically, but also the OCT4 expression could be monitored by means of eGFP expression. A loss of eGFP expression means a loss in OCT4 expression and thus differentiation. A limitation of our method is that only OCT4 expression and no other stem cell-specific markers can be followed-up. To our knowledge, no reporter cell line exists in which two stem cell-specific markers can be followed-up in hESC. A solution to this problem can be the addition of the analysis of alkaline phosphatase. Alkaline phosphatase is an enzyme which is highly expressed in undifferentiated hESC (4). Non-invasive detection of alkaline phosphatase is possible using a cell-permeable, non-toxic small molecule substrate which becomes fluorescent after dephosphorylation by alkaline phosphatase. This substrate diffuses out of the cells during time (23). Several companies (ThermoFisher and Abcam) provide an alkaline phosphatase determination kit with a different fluorescent signal (ThermoFisher: ex/em 495 nm/519 nm, Abcam: ex/em 360nm/440nm) (175, 176). Therefore, only the Abcam kit can be used in our case when combining this kit with the analysis of the expression of OCT 4 in the OCT4-eGFP Knock-in hESC line (eGFP: ex/em 489 nm/509 nm). Of note, when using this combination for the first time, one should first confirm that no spectral overlap took place or if so, compensation was performed.

In addition, other remaining hurdles (in particular the need for an incubation chamber and the auto-fluorescence of the medium for FM analysis) have to be overcome. At this moment, imaging systems with an incubation chamber regulating heat, carbon dioxide or oxygen can be purchased, for example the imaging system “cell observer” of Zeiss. In addition, this instrument makes it possible to follow-up the cells in a time-lapse experiment. For the latter hurdle, we have analyzed the cells without the use of any medium or with the use of PBS. To minimize the effect on the culture however, analysis had to be

performed in a minimal amount of time. A solution would be the use of a medium without any auto-fluorescent compound. For this purpose, ThermoFisher has recently developed “FluoroBrite DMEM” which has a 90 % lower auto-fluorescence compared to phenol red-free medium. Importantly, it remains to be determined whether it can be used for hESC culture, especially in light of the abovementioned “metastable” cellular state of hESC. An integrative pipeline to validate this adapted culture medium is presented in this dissertation: (i) The impact of “FluoroBrite DMEM” on pluripotency can be assessed by either FC or FM (**Chapter 3).** (ii) If indeed pluripotency, colony morphology or cell amount is not compromised in this medium, a proteome analysis should be performed to define other potential changes in the cell. The latter can be attained by either label-free DDA or DIA analysis with minimized technical variation (**Chapter 4**) or by using SILAC (**Chapter 5).** Of note, in addition to proteomic changes, also metabolomic changes can occur when changing culture conditions. For example, naïve stem cells catabolize glutamine to maintain a high α-ketoglutarate/succinate ratio. This high ratio promotes histone/DNA methylation and maintains pluripotency (177). This analysis show that metabolomics is certainly valuable and essential in the analysis of hESC and can be implemented in the future to characterize hESC.

Next to the monitoring of the differentiation status of stem cells, we thus focused in a second part of this dissertation on the proteomic analysis of these cells. Under circumstances where a change of hESC culture had to be examined, one can analyze the effect on pluripotency and proteome of these cells. The field of proteomics has emerged during the last years: mass spectrometry instruments became more evolved, obtaining higher speed, sensitivity and resolution resulting in increased peptide and protein identifications. In our lab, the ESI-Q-TOF Premier which was used at the beginning of my work was replaced by a TripleTOF 5600 (Sciex) and Synapt G2Si (Waters). The growing technical capabilities of LC-MS instrumentation shown in **Table 1** illustrates the increasing pressure that is put on sample preparation (178).

Table 1. Overview of mass spectrometers used during this dissertation. Technical information considering data generation during data dependent acquisition is displayed. The last row shows data generated with a HeLa cell lysate as described in (178). Improved data quality is obtained with Synapt G2Si and TripleTOF 5600 in comparison with the Premier.

DDA mode	Premier (Waters)	Synapt G2Si (Waters)	TripleTOF 5600 (Sciex)
Mass accuracy	100-250 ppm	1-15 ppm	1-5 ppm
Resolution in MS	10k FWHM	10-50k FWHM	10-45k FWHM
Max MSMS/sec	1	30	50
Average MSMS/min	10-20	100-200	200-300
MSMS/µg sample in x time (178)	*25k/500 µg in 1 month (2-3D LC)*	*20k/0.5 µg in 2 hours*	*30k/1µg in 2 hours*

In addition, data-independent acquisition (DIA) is developed, in which theoretically every peptide can now be analyzed. Quantitative differences between two or more experiments can be addressed by several label-based and label-free approaches. At this moment, the field of proteomics has focused on label-free analysis in combination with DIA or DDA. In label-free analysis, samples are mixed at the step of data-analysis, which makes sample preparation a step of great importance. Regarding DIA or DDA, quantitation results based on AUC are improved using DIA in comparison with DDA: (1) every precursor peptide ion is observed with DIA and (2) the duty cycle is almost 100 % resulting in more accurate and reproducible data (99). Applying these label-free approaches thus increasingly requires a thoroughly validated extraction and digestion protocol in which not only identification, but also technical repeatability has to be taken into consideration.

Cell lysis is the first step in the sample preparation protocol. In **Chapter 4**, we compare different protein extraction protocols on different cell lines (two suspension cell lines and one adherent cell line). A greater and reproducible protein yield was observed by addition of 1 % SDC in all of these three cell cultures in comparison with no detergent. Because its removal by acid-precipitation does not require more than one additional step, technical repeatability is not greatly compromised. One very suitable application of this detergent is thus in histone extraction by means of acid precipitation on isolated

nuclei, as routinely done for epigenetic studies (179). An acid-insoluble detergent like SDC, improves protein recovery while being removed at the mandatory acid precipitation step. Of note, not every cell has the same structural characteristics and cell lysis protocols should be adapted for the cell type under investigation.

As for protein extraction, protein digestion should be as reproducible as possible. Herein, the enzymes used during in-solution digest need to work in an optimal way. This implies that the protein sequence needs to be available for cleavage. This is obtained (1) by unfolding the protein to its primary structure (addition of a detergent), (2) by breaking disulphide bridges (addition of reductans and alkylans) and (3) by using 5 % acetonitrile and 1 mM $CaCl_2$. The addition of acetonitrile allows for higher protein accessibility, while Ca^{2+}-ions improves the working of trypsin (180). Again, SDC is the detergent of choice because peptides remain soluble in 2 % TFA, while SDC precipitates.

Similarly, increased enzyme specificity can be considered to increase technical repeatability as reported in literature. The combination of trypsin with lysC makes it possible to cleave peptides on the C-terminal side of lysine even when lysine is followed by proline. Of note, when a quantitative workflow using chemical labeling is preferable, such as in iTRAQ, it is important to realize that using e.g. TeABC as buffer avoids that primary amines in the buffer (as for example with TRIS) would interfere with label efficiency and that TeABC easily is removed by vacuum drying (180). We thus suggest here to use TeABC as a standard buffer system compatible with all conventional workflows in quantitative proteomics.

In summary, we here described a detailed protocol that can be used from cell lysis to peptide analysis wherein we thoroughly validate (1) repeatability, (2) peptide identification and (3) MS compatibility. In addition, the possibility of peptide loss during SDC removal is analyzed. To our knowledge, this is the first sample preparation protocol from cell lysis to LC-MS/MS which uses a minimal number of steps and in which repeatability and identification is addressed. This protocol can easily be implemented in

other laboratories and can be used for label and label-based purposes. In addition, it makes no use of expensive detergents such as Rapigest or ProteaseMax.

Despite the optimized extraction and digestion protocol presented in this dissertation, the intrinsically accumulating technical variability in the label-free approaches will always leave room for the parallel existence of metabolic labeling strategies. Most accurate information can be provided by using stable-isotopic labeling by amino acids (SILAC), in which samples are already mixed before cell lysis. Before using SILAC, one should investigate if full incorporation of the heavy label took place and if no *in vitro* conversion of the heavy amino acid is visible. Although the technique was developed using deuterated leucine, it was quickly evident that this heavy label was not the ideal choice because (1) not every peptide contains leucine and (2) the use of a deuterated label evokes a shift in retention time making data analysis afterwards more difficult (115). The use of $^{13}C_6$ arginine and $^{3}C_6$ lysine in combination with trypsin as digestive enzyme made it possible to theoretically quantify each peptide. However, the use of arginine as heavy label has shown an *in vitro* conversion to proline, resulting in quantification errors. This was also the case in our hESC culture system. Of note, the culture system used at our lab is E8 medium in combination with vitronectin as coating, which is ideal for proteomic approaches: no co-culture of other cells and a fully defined medium which made it possible to discriminate between proteins derived from hESC and from medium. The arginine conversion in our culture system could be inhibited by several solutions in which the addition of 5 mM ornithine to the medium resulted in the largest reduced conversion without major effects onto the proteome (**Chapter 5**). Of note, hESC in their adapted media were first monitored for their effect on pluripotency as described in **Chapter 3** and the sample preparation for proteomics was performed using the protocol as described in **Chapter 4**, showing the usefulness of both methods. This adapted culture system can now be used for SILAC experiments with hESC. The advantage of our method over other described methods is that our method uses a fully defined culture system avoiding lot variability, as would be the case when using for example MEF CM as medium or Matrigel as coating. Undefined culture systems can in particular lead to inconsistent results between different laboratories. In addition, the use of ornithine is a new way to reduce conversion in

cell lines and can consequently be tested in other cell lines as well which have also a problem in arginine conversion such as HeLa or HEK cells.

In conclusion, in this dissertation we have described a workflow for respectively the cellular monitoring of human embryonic stem cells and for the quantitative and qualitative analysis of the proteome either by means of a label-based (SILAC) or label-free approach. Besides these issues in cell culture and sample preparation, future efforts will need to address (1) LC-MS data acquisition and (2) data analysis software tools, especially when considering the relatively new DIA approach. With the same amount of wet lab work, now come considerably larger datasets. Increasingly, this is skewing the field of proteomics to the data analysis part of the workflow, where bioinformaticians have by now become an essential link in the chain. Yet, the main pitfall of this evolution is that it might lead to negligence during the sample preparation. This dissertation emphasizes the importance of continuously keeping a careful watch on the whole experimental pipeline, from biological sample to data analysis.

CHAPTER 7: CONCLUSION

hESC are an important model system and display great promise for applications in toxicology and medicine. However, a hallmark of these cells is their “meta-stable state” in culture, meaning that even small disruptions to the culture system can easily result in the loss of pluripotency and “spontaneous” differentiation. In this work, a non-invasive method for monitoring hESC pluripotency is presented which is easy to use. While proven to be essential, OCT4 monitoring alone can never match a complete proteomics assessment of the culture system. Thus, either as an additional screening or to more specifically define changes following e.g. toxicological treatments, a proteomic analysis can be used to address potential proteomic alterations.

Quantitative information can be obtained either with label-free quantitation or with label-based approaches (SILAC, iTRAQ). Label-based quantitation is prone to less technical variation, because samples are mixed in an earlier stage of the experimental workflow. In hESC however, this technique can not directly be applied to the culture used in our lab, because of the observed metabolic conversion of the heavy labeled arginine to proline. Indeed, this conversion could be reduced, hereby making SILAC applicable to hESC. Because of the use of a fully defined and xenofree hESC culture, others can use this protocol directly for their own purposes. When SILAC cannot be used, label-free quantitation is increasingly becoming the method of choice, because of its straightforward workflow. However, to minimize technical variation, a sample preparation protocol was developed which is reproducible and provides sufficient peptide identification.

In the future, label-free quantitation with DIA will increasingly become the method of choice because of its cheapness and the possibility to analyze a lot of samples at once. With a well-monitored cell culture and a reproducible extraction method, it might even compete with SILAC in terms of accurate quantitation when defining biological changes. In conclusion, however, thoroughly optimizing and validating an experimental protocol remains the central pillar of good practice.

CHAPTER 8: SUMMARY/SAMENVATTING

The cooperation with Ghent Fertility and Stem cell Team (G-Fast) (Ghent University Hospital) made it possible to set up a human Embryonic Stem Cell (hESC) culture at our lab. At the start of this dissertation, hESC were kept in culture on a feeder layer of mouse embryonal fibroblasts. However, this type of culture results in a more difficult analysis of hESC and influences the study of hESC considerably. For this reason, a feeder-free culture was used, more specifically the culture of hESC in Essential 8 medium in combination with vitronectin. This culture is completely defined and xenofree.

The hESC culture was at that moment only morphologically examined in order to detect the presence of differentiated cells. This method is however very subjective, so the use of a commercially available reporter hESC line was validated in which the expression of eGFP is under the control of the pluripotency gene OCT4 (**Chapter 3**). The use of a fluorescence microscope made it possible to examine hESC objectively at that moment. A lower expression of eGFP is related to a lower expression of OCT4. Additional advantages of our method in comparison with flow cytometry, the gold standard in most screenings, are that differences in OCT4 expression in the colony itself can be shown and that the cells can be kept in culture after screening. The main obstacle in this approach is the auto-fluorescence of the medium. For this reason, no medium was used during the analysis. This led however to cellular stress.

hESC are very sensitive to changes in their environment. Even if no effect on pluripotency can be observed by means of the abovementioned method, it remains important to screen other changes for example at the protein level. The sequel of this dissertation was therefore characterized by the optimization of a workflow for the proteomic analysis of hESC.

Metabolic labeling strategy or SILAC is one way to obtain both quantitative and qualitative information about the proteome. In this method, isotopically labeled amino acids (such as arginine and lysine) are added to the culture medium and are built into the proteome. Before this technique can be used for a comparative study in a certain cell culture, one should examine if (1) a complete incorporation of heavy labels took place and (2) no *in vitro* conversion was observed of arginine to proline. Several possibilities are described in literature to inhibit this conversion: addition of proline or lowering the amount of

arginine in the medium. In this dissertation, both solutions were analyzed, but also the addition of respectively ornithine or Nor-NOHA to the medium was examined as described in **Chapter 5**. By using the optimized screening methodology with the OCT4 reporter hESC, as described in **Chapter 3**, the different "adapted" media were examined in terms of pluripotency, colony morphology and cell amount. Only a positive effect on the cell growth by the addition of 50 µM Nor-NOHA and a suddenly cell death using low concentrations of arginine were observed. The other media were usable for hESC culture.

Further examination showed that addition of 5 mM ornithine and 3.5 mM proline as well decreasing the amount of arginine to 99.5 µM in the medium resulted in a significant decrease of the conversion of arginine to proline. Finally, these abovementioned culture conditions were further investigated with $UDMS^E$ to investigate the possible effect onto the proteome. This label-free method makes it possible to quantify peptides by means of "Area-Under-the-Curve" quantitation using Progenesis QI (Nonlinear Dynamics, Waters). Out of this analysis, one should conclude that no major effect on the proteome in comparison with the control ($E8^{TM}$ medium without any addition) could be observed in this short time frame.

To make this label-free analysis possible, the existing sample preparation protocol had to be optimized for protein extraction on human cells. Repeatability during sample preparation is in this label-free technique of great importance to obtain accurate quantitative information. In contrast to SILAC, samples are only analyzed together at the step of data-analysis.

This protocol, as described in **Chapter 4**, is optimized from cell lysis up to "in-solution" digest and makes use of one detergent, 1 % (w/v) SDC, and one buffer, 500 mM TeABC. 1 % (w/v) SDC was chosen because of (1) the reported efficiency to denature proteins (similar as SDS), (2) the cost (much cheaper in comparison with for example $Rapigest^{TM}$) and (3) the possibility to remove SDC by means of acid precipitation in which peptides remain in solution.

This dissertation describes and validates in detail the several technical considerations that have to be addressed for the study of the hESC proteome. Cells need to be monitored in detail in terms of

pluripotency and colony morphology, certainly if something needs to be changed to the culture conditions, as for example was the case in techniques using metabolic labeling like SILAC. However, because data-independent data acquisition will become more and more important, strong validation of the protein extraction and digestion in terms of repeatability is needed, certainly if these label-free approaches will compete with label-based approaches.

De samenwerking met het Ghent Fertility and Stem cell Team (G-FaST) (Universitair Ziekenhuis Gent) maakte het mogelijk om een humane Embryonale Stam Cel (hESC) cultuur op te zetten in ons labo. Bij aanvang van dit doctoraat werden deze cellen op een feeder layer van muis embryonale fibroblasten in cultuur gehouden, maar deze vorm van cultuur bemoeilijkt en beïnvloedt de studie van hESC aanzienlijk (**Hoofdstuk 1**). Om deze reden werd er gekozen voor een feeder-vrije cultuur genaamd Essential 8 medium in combinatie met vitronectine. Deze cultuur is volledig gedefinieerd en xenovrij.

De hESC cultuur werd op dat moment enkel morfologisch beoordeeld op de aanwezigheid van differentiërende cellen. Doordat deze methode echter te subjectief is, werd het gebruik van een commerciële reporter hESC lijn gevalideerd waarbij de expressie van eGFP onder controle staat van het pluripotentie kenmerkend gen OCT4 (**Hoofdstuk 3**). Door gebruik te maken van een fluorescentiemicroscoop is het nu mogelijk om hESC objectief te beoordelen. Een lagere eGFP expressie is immers gerelateerd aan een lagere expressie van OCT4. Bijkomende voordelen van onze methode ten opzichte van flow cytometrie, de gouden standaard in de meeste screenings, zijn dat verschillen in OCT4 expressie in de kolonie zelf aangetoond kunnen worden en dat de cellen na screening verder in cultuur kunnen worden gehouden. Het belangrijkste obstakel in de methode is de auto-fluorescentie van het medium. Hierdoor werd er geopteerd om geen medium te gebruiken tijdens de analyse zelf. Dit leidde echter tot cellulaire stress.

hESC zijn zeer gevoelig aan veranderingen in hun omgeving. Zelfs indien er geen effect op de pluripotentie zichtbaar is op basis van de bovenvermelde methode, blijft het belangrijk om ook andere veranderingen, bijvoorbeeld op gebied van proteïneniveau te screenen. Het vervolg van deze scriptie stond dan ook in het teken van de optimalisatie van een workflow voor de proteoomanalyze van hESC.

“Metabolic labeling” of SILAC is een manier om zowel kwantitatieve als kwalitatieve informatie over het proteoom te verkrijgen. In deze methode worden isotopisch zwaar gelabelde aminozuren (zoals arginine en lysine) aan het cultuurmedium toegevoegd en ingebouwd in het proteoom. Alvorens een dergelijke techniek toe te passen voor een vergelijkende studie in een bepaalde celcultuur, moet

nagegaan worden of (1) er volledige incorporatie is van de labels en (2) er een *in vitro* conversie plaatsvindt van arginine naar proline. Verschillende mogelijkheden zijn beschreven in de literatuur om deze conversie tegen te gaan: toevoegen van proline of verlagen van de hoeveelheid arginine in het medium. In deze thesis werden beide oplossingen uitgetest, alsook de toevoeging van respectievelijk ornithine en Nor-NOHA aan het medium zoals beschreven in **Hoofdstuk 5**. Door het gebruik van de in **Hoofdstuk 3** geoptimaliseerde screening methodologie met OCT4 reporter hESC, zijn de verschillende "aangepaste" media eerst onderzocht op hun effect op pluripotentie, kolonie morfologie en cel aantal. Naast een mogelijk positief effect op de celgroei bij toevoeging van 50 µM Nor-NOHA en een soms plotselinge celdood bij lage concentraties arginine, waren de meeste media bruikbaar voor hESC cultuur.

Uit verder onderzoek bleek dat de toevoeging van 5 mM ornithine of 3.5 mM proline, alsook de verlaging van de concentratie arginine naar 99.5 µM arginine resulteerde in een significante reductie van de conversie van arginine in proline. Deze cultuurcondities werden verder onderzocht met $UDMS^E$ om het potentiële effect op het proteoom te onderzoeken. Deze label-vrije methode laat peptide kwantificatie toe aan de hand van "oppervlakte-onder-de-curve" kwantificatie met behulp van Progenesis QI (Nonlinear Dynamics, Waters). Hieruit bleek dat er alvast op het onderzochte tijdsverloop geen groot effect op het proteoom kon worden gevonden ten opzichte van de controle.

Om deze label-vrije analyse mogelijk te maken werd het bestaande staalvoorbereidingsprotocol geoptimaliseerd voor proteïne extractie op humane cellen. Herhaalbaarheid gedurende staalvoorbereiding is in deze label-vrije techniek namelijk van groot belang om tot accurate kwantitatieve informatie te komen. Dit is omdat bij deze techniek, in tegenstelling tot SILAC, de stalen pas bij data-analyse samen worden geanalyseerd.

Dit protocol, beschreven in **Hoofdstuk 4**, is geoptimaliseerd vanaf cellyse tot en met het "in oplossing" digest en maakt gebruik van één detergent, met name 1 % (w/v) SDC en één buffer, met name 500 mM TeABC. Er werd voor 1 % (w/v) SDC gekozen omwille van (1) de gerapporteerde doeltreffendheid om eiwitten te denatureren (gelijkaardig aan SDS), (2) de kostprijs (veel goedkoper dan bijvoorbeeld

Rapigest^{TM}) en (3) de mogelijkheid om SDC te verwijderen met behulp van zuur precipitatie waarbij peptiden in oplossing blijven.

Deze scriptie beschrijft en valideert in detail de verschillende technische overwegingen die moeten worden genomen bij de studie van het proteoom van hESC. De cellen zelf dienen in detail te worden gemonitord op vlak van pluripotentie en kolonie-morfologie, zeker als er iets aan de cultuurcondities dient te veranderen, zoals het geval bij technieken die berusten op metabole labeling zoals SILAC. Nu data-afhankelijke data acquisitie steeds meer aan belang wint, zullen proteïne extractie en digestie zeer sterk gevalideerd moeten worden op vlak van herhaalbaarheid, willen deze label-vrije benaderingen de intrinsiek minder variabele labelingstechnieken ook naar de kroon stoten.

CHAPTER 9: REFERENCES

1. Hoffman, L. M.; Carpenter, M. K., Characterization and culture of human embryonic stem cells. *Nat Biotechnol* **2005,** 23, (6), 699-708.

2. O'Leary, T.; Heindryckx, B.; Lierman, S.; van Bruggen, D.; Goeman, J. J.; Vandewoestyne, M.; Deforce, D.; de Sousa Lopes, S. M.; De Sutter, P., Tracking the progression of the human inner cell mass during embryonic stem cell derivation. *Nat Biotechnol* **2012,** 30, (3), 278-82.

3. Wright, W. E.; Piatyszek, M. A.; Rainey, W. E.; Byrd, W.; Shay, J. W., Telomerase activity in human germline and embryonic tissues and cells. *Developmental Genetics* **1996,** 18, (2), 173-179.

4. Thomson, J. A.; Itskovitz-Eldor, J.; Shapiro, S. S.; Waknitz, M. A.; Swiergiel, J. J.; Marshall, V. S.; Jones, J. M., Embryonic stem cell lines derived from human blastocysts. *Science* **1998,** 282, (5391), 1145-7.

5. Van Hoof, D.; Heck, A. J. R.; Krijgsveld, J.; Mummery, C. L., Proteomics and human embryonic stem cells. *Stem Cell Research* **2008,** 1, (3), 169-182.

6. Hughes, C. S.; Nuhn, A. A.; Postovit, L. M.; Lajoie, G. A., Proteomics of human embryonic stem cells. *Proteomics* **2011,** 11, (4), 675-690.

7. Burridge, P. W.; Keller, G.; Gold, J. D.; Wu, J. C., Production of De Novo Cardiomyocytes: Human Pluripotent Stem Cell Differentiation and Direct Reprogramming. *Cell Stem Cell* **2012,** 10, (1), 16-28.

8. Xu, C. H.; Inokuma, M. S.; Denham, J.; Golds, K.; Kundu, P.; Gold, J. D.; Carpenter, M. K., Feeder-free growth of undifferentiated human embryonic stem cells. *Nature Biotechnology* **2001,** 19, (10), 971-974.

9. Takahashi, K.; Yamanaka, S., Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **2006,** 126, (4), 663-676.

10. Yamanaka, S., Induced Pluripotent Stem Cells: Past, Present, and Future. *Cell Stem Cell* **2012,** 10, (6), 678-684.

11. Drews, K.; Jozefczuk, J.; Prigione, A.; Adjaye, J., Human induced pluripotent stem cells-from mechanisms to clinical applications. *Journal of Molecular Medicine-Jmm* **2012,** 90, (7), 735-745.

12. Takahashi, K.; Tanabe, K.; Ohnuki, M.; Narita, M.; Ichisaka, T.; Tomoda, K.; Yamanaka, S., Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **2007,** 131, (5), 861-872.

13. Bhowmik, S.; Li, Y., Induced pluripotent stem cells. *Chinese Medical Journal* **2011,** 124, (12), 1897-1900.

14. Medvedev, S. P.; Shevchenko, A. I.; Mazurok, N. A.; Zakian, S. M., OCT4 and NANOG are the key genes in the system of pluripotency maintenance in mammalian cells. *Russian Journal of Genetics* **2008,** 44, (12), 1377-1393.

15. Nichols, J.; Zevnik, B.; Anastassiadis, K.; Niwa, H.; Klewe-Nebenius, D.; Chambers, I.; Scholer, H.; Smith, A., Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **1998,** 95, (3), 379-391.

16. Walker, E.; Ohishi, M.; Davey, R. E.; Zhang, W.; Cassar, P. A.; Tanaka, T. S.; Der, S. D.; Morris, Q.; Hughes, T. R.; Zandstra, P. W.; Stanford, W. L., Prediction and testing of novel transcriptional networks regulating embryonic stem cell self-renewal and commitment. *Cell Stem Cell* **2007,** 1, (1), 71-86.

17. Wang, Z.; Oron, E.; Nelson, B.; Razis, S.; Ivanova, N., Distinct Lineage Specification Roles for NANOG, OCT4, and SOX2 in Human Embryonic Stem Cells. *Cell Stem Cell* **2012,** 10, (4), 440-454.

18. Ng, H. H.; Surani, M. A., The transcriptional and signalling networks of pluripotency. *Nature Cell Biology* **2011,** 13, (5), 490-496.

19. Xu, R. H.; Sampsell-Barron, T. L.; Gu, F.; Root, S.; Peck, R. M.; Pan, G. J.; Yu, J. Y.; Antosiewicz-Bourget, J.; Tian, S. L.; Stewart, R.; Thomson, J. A., NANOG is a direct target of TGF beta/Activin-mediated SMAD signaling in human ESCs. *Cell Stem Cell* **2008,** 3, (2), 196-206.

20. Yang, A. X.; Mejido, J.; Luo, Y. Q.; Zeng, X. M.; Schwartz, C.; Wu, T. X.; Thies, R. S.; Bhattacharya, B.; Han, J.; Freed, B.; Rao, M.; Puri, R. K., Development of a focused microarray to assess human embryonic stem cell differentiation. *Stem Cells and Development* **2005,** 14, (3), 270-284.

21. Noaksson, K.; Zoric, N.; Zeng, X. M.; Rao, M. S.; Hyllner, J.; Semb, H.; Kubista, M.; Sartipy, P., Monitoring differentiation of human embryonic stem cells using real-time PCR. *Stem Cells* **2005,** 23, (10), 1460-1467.

22. Bongso, A.; Richards, M., History and perspective of stem cell research. *Best Practice & Research in Clinical Obstetrics & Gynaecology* **2004,** 18, (6), 827-842.

23. Singh, U.; Quintanilla, R. H.; Grecian, S.; Gee, K. R.; Rao, M. S.; Lakshmipathy, U., Novel Live Alkaline Phosphatase Substrate for Identification of Pluripotent Stem Cells. *Stem Cell Reviews and Reports* **2012,** 8, (3), 1021-1029.

24. James, D.; Levine, A. J.; Besser, D.; Hemmati-Brivanlou, A., TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development* **2005,** 132, (6), 1273-82.

25. Sato, N.; Meijer, L.; Skaltsounis, L.; Greengard, P.; Brivanlou, A. H., Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nature Medicine* **2004,** 10, (1), 55-63.

26. McLean, A. B.; D'Amour, K. A.; Jones, K. L.; Krishnamoorthy, M.; Kulik, M. J.; Reynolds, D. M.; Sheppard, A. M.; Liu, H. Q.; Xu, Y.; Baetge, E. E.; Dalton, S., Activin a efficiently specifies definitive endoderm from human embryonic stem cells only when phosphatidylinositol 3-kinase signaling is suppressed. *Stem Cells* **2007,** 25, (1), 29-38.

27. Armstrong, L.; Hughes, O.; Yung, S.; Hyslop, L.; Stewart, R.; Wappler, I.; Peters, H.; Walter, T.; Stojkovic, P.; Evans, J.; Stojkovic, M.; Lako, M., The role of PI3K/AKT, MAPK/ERK and NF kappa beta signalling in the maintenance of human embryonic stem cell pluripotency and viability highlighted by transcriptional profiling and functional analysis. *Human Molecular Genetics* **2006,** 15, (11), 1894-1913.

28. McCain, J., The MAPK (ERK) Pathway: Investigational Combinations for the Treatment Of BRAF-Mutated Metastatic Melanoma. *P T* **2013,** 38, (2), 96-108.

29. Ding, V. M. Y.; Ling, L.; Natarajan, S.; Yap, M. G. S.; Cool, S. M.; Choo, A. B. H., FGF-2 Modulates Wnt Signaling in Undifferentiated hESC and iPS Cells Through Activated PI3-K/GSK3 beta Signaling. *Journal of Cellular Physiology* **2010,** 225, (2), 417-428.

30. Walsh, J.; Andrews, P. W., Expression of Wnt and Notch pathway genes in a pluripotent human embryonal carcinoma cell line and embryonic stem cells. *Apmis* **2003,** 111, (1), 197-211.

31. Bieberich, E.; Wang, G., Molecular Mechanisms Underlying Pluripotency In *Pluripotent Stem Cells*, Bhartiya, D., Ed. 2013.

32. Xu, R. H.; Peck, R. M.; Li, D. S.; Feng, X. Z.; Ludwig, T.; Thomson, J. A., Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. *Nature Methods* **2005,** 2, (3), 185-190.

33. Lund, R. J.; Narva, E.; Lahesmaa, R., Genetic and epigenetic stability of human pluripotent stem cells. *Nature Reviews Genetics* **2012,** 13, (10), 732-744.

34. Imataka, G.; Arisaka, O., Chromosome Analysis Using Spectral Karyotyping (SKY). *Cell Biochemistry and Biophysics* **2012,** 62, (1), 13-17.

35. Leary, R. J.; Cummins, J.; Wang, T. L.; Velculescu, V. E., Digital karyotyping. *Nature Protocols* **2007,** 2, (8), 1973-1986.

36. Oostlander, A. E.; Meijer, G. A.; Ylstra, B., Microarray-based comparative genomic hybridization and its applications in human genetics. *Clinical Genetics* **2004,** 66, (6), 488-495.

37. Lim, J. W.; Bodnar, A., Proteome analysis of conditioned medium from mouse embryonic fibroblast feeder layers which support the growth of human embryonic stem cells. *Proteomics* **2002,** 2, (9), 1187-203.

38. Miyazaki, T.; Futaki, S.; Hasegawa, K.; Kawasaki, M.; Sanzen, N.; Hayashi, M.; Kawase, E.; Sekiguchi, K.; Nakatsuji, N.; Suemori, H., Recombinant human laminin isoforms can support the undifferentiated growth of human embryonic stem cells. *Biochemical and biophysical research communications* **2008,** 375, (1), 27-32.

39. Lu, J.; Hou, R.; Booth, C. J.; Yang, S. H.; Snyder, M., Defined culture conditions of human embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **2006,** 103, (15), 5688-93.

40. Chin, A. C.; Fong, W. J.; Goh, L. T.; Philp, R.; Oh, S. K.; Choo, A. B., Identification of proteins from feeder conditioned medium that support human embryonic stem cells. *Journal of biotechnology* **2007,** 130, (3), 320-8.

41. Vukicevic, S.; Kleinman, H. K.; Luyten, F. P.; Roberts, A. B.; Roche, N. S.; Reddi, A. H., Identification of Multiple Active Growth-Factors in Basement-Membrane Matrigel Suggests Caution in Interpretation of Cellular-Activity Related to Extracellular-Matrix Components. *Experimental Cell Research* **1992,** 202, (1), 1-8.

42. Chen, G. K.; Gulbranson, D. R.; Hou, Z. G.; Bolin, J. M.; Ruotti, V.; Probasco, M. D.; Smuga-Otto, K.; Howden, S. E.; Diol, N. R.; Propson, N. E.; Wagner, R.; Lee, G. O.; Antosiewicz-Bourget, J.; Teng, J. M. C.; Thomson, J. A., Chemically defined conditions for human iPSC derivation and culture. *Nature Methods* **2011,** 8, (5), 424-429.

43. T'Joen, V.; Declercq, H.; Cornelissen, M., Expansion of human embryonic stem cells: a comparative study. *Cell Proliferation* **2011,** 44, (5), 462-476.

44. C5914,Sigma-Aldrich http://www.sigmaaldrich.com/catalog/product/sigma/c5914?lang=en®ion=BE.

45. Hannoun, Z.; Fletcher, J.; Greenhough, S.; Medine, C.; Samuel, K.; Sharma, R.; Pryde, A.; Black, J. R.; Ross, J. A.; Wilmut, I.; Iredale, J. P.; Hay, D. C., The Comparison between Conditioned Media and Serum-Free Media in Human Embryonic Stem Cell Culture and Differentiation. *Cellular Reprogramming* **2010,** 12, (2), 133-140.

46. Yoon, T. M.; Chang, B.; Kim, H. T.; Jee, J. H.; Kim, D. W.; Hwang, D. Y., Human Embryonic Stem Cells (hESCs) Cultured Under Distinctive Feeder-Free Culture Conditions Display Global Gene Expression Patterns Similar to hESCs from Feeder-Dependent Culture Conditions. *Stem Cell Reviews and Reports* **2010,** 6, (3), 425-437.

47. Higuchi, A.; Ling, Q. D.; Ko, Y. A.; Chang, Y.; Umezawa, A., Biomaterials for the Feeder-Free Culture of Human Embryonic Stem Cells and Induced Pluripotent Stem Cells. *Chemical Reviews* **2011,** 111, (5), 3021-3035.

48. Li, Y.; Powell, S.; Brunette, E.; Lebkowski, J.; Mandalam, R., Expansion of human embryonic stem cells in defined serum-free medium devoid of animal-derived products. *Biotechnology and bioengineering* **2005,** 91, (6), 688-98.

49. Li, Y.; Lin, C.; Wang, L.; Liu, Y.; Mu, X.; Ma, Y.; Li, L., Maintenance of human embryonic stem cells on gelatin. *Chinese Science Bulletin* **2009,** 54, (22), 4214-4220.

50. Blair, K.; Wray, J.; Smith, A., The Liberation of Embryonic Stem Cells. *Plos Genetics* **2011,** 7, (4), e1002019

51. Williams, R. L.; Hilton, D. J.; Pease, S.; Willson, T. A.; Stewart, C. L.; Gearing, D. P.; Wagner, E. F.; Metcalf, D.; Nicola, N. A.; Gough, N. M., Myeloid-Leukemia Inhibitory Factor Maintains the Developmental Potential of Embryonic Stem-Cells. *Nature* **1988,** 336, (6200), 684-687.

52. Ying, Q. L.; Wray, J.; Nichols, J.; Batlle-Morera, L.; Doble, B.; Woodgett, J.; Cohen, P.; Smith, A., The ground state of embryonic stem cell self-renewal. *Nature* **2008,** 453, (7194), 519-523.

53. Davidson, K. C.; Mason, E. A.; Pera, M. F., The pluripotent state in mouse and human. *Development* **2015,** 142, (18), 3090-9.

54. Honda, A.; Hatori, M.; Hirose, M.; Honda, C.; Izu, H.; Inoue, K.; Hirasawa, R.; Matoba, S.; Togayachi, S.; Miyoshi, H.; Ogura, A., Naive-like Conversion Overcomes the Limited Differentiation Capacity of Induced Pluripotent Stem Cells. *Journal of Biological Chemistry* **2013,** 288, (36), 26157-26166.

55. Jang, H. J.; Kim, J. S.; Choi, H. W.; Jeon, I.; Choi, S.; Kim, M. J.; Song, J.; Do, J. T., Neural stem cells derived from epiblast stem cells display distinctive properties. *Stem Cell Research* **2014,** 12, (2), 506-516.

56. Hanna, J.; Cheng, A. W.; Saha, K.; Kim, J.; Lengner, C. J.; Soldner, F.; Cassady, J. P.; Muffat, J.; Carey, B. W.; Jaenisch, R., Human embryonic stem cells with biological and epigenetic

characteristics similar to those of mouse ESCs. *Proc Natl Acad Sci U S A* **2010,** 107, (20), 9222-7.

57. Ware, C. B.; Nelson, A. M.; Mecham, B.; Hesson, J.; Zhou, W. Y.; Jonlin, E. C.; Jimenez-Caliani, A. J.; Deng, X. X.; Cavanaugh, C.; Cook, S.; Tesar, P. J.; Okada, J.; Margaretha, L.; Sperber, H.; Choi, M.; Blau, C. A.; Treuting, P. M.; Hawkins, R. D.; Cirulli, V.; Ruohola-Baker, H., Derivation of naive human embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **2014,** 111, (12), 4484-4489.

58. Patterson, G. H.; Knobel, S. M.; Sharif, W. D.; Kain, S. R.; Piston, D. W., Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophysical Journal* **1997,** 73, (5), 2782-2790.

59. Lichtman, J. W.; Conchello, J. A., Fluorescence microscopy. *Nature Methods* **2005,** 2, (12), 910-919.

60. Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Robert, K.; Walter, P., Looking at cells in the light microscope. In *Molecular Biology of the cell*, Anderson, M.; Granum, S., Eds. 2008.

61. Zwaka, T. P.; Thomson, J. A., Homologous recombination in human embryonic stem cells. *Nature biotechnology* **2003,** 21, (3), 319-21.

62. Macey, M., Principles of Flow Cytometry. In *Flow cytometry Principles and Applications*, Macey, M., Ed. 2007; pp 1-17.

63. Abcam http://www.abcam.com/index.html?pageconfig=resource&rid=11446.

64. Zhong, J. F.; Weiner, L.; Jin, Y. S.; Lu, W. G.; Taylor, C. R., A Real-Time Pluripotency Reporter for Human Stem Cells. *Stem Cells and Development* **2010,** 19, (1), 47-52.

65. Fischer, Y.; Ganic, E.; Ameri, J.; Xian, X.; Johannesson, M.; Semb, H., NANOG reporter cell lines generated by gene targeting in human embryonic stem cells. *PLoS One* **2010,** 5, (9).

66. Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* **2003,** 422, (6928), 198-207.

67. Baharvand, H.; Hajheidari, M.; Ashtiani, S. K.; Salekdeh, G. H., Proteomic signature of human embryonic stem cells. *Proteomics* **2006,** 6, (12), 3544-3549.

68. Andrews, G. L.; Simons, B. L.; Young, J. B.; Hawkridge, A. M.; Muddiman, D. C., Performance Characteristics of a New Hybrid Quadrupole Time-of-Flight Tandem Mass Spectrometer (TripleTOF 5600). *Analytical Chemistry* **2011,** 83, (13), 5442-5446.

69. Helm, D.; Vissers, J. P. C.; Hughes, C. J.; Hahne, H.; Ruprecht, B.; Pachl, F.; Grzyb, A.; Richardson, K.; Wildgoose, J.; Maier, S. K.; Marx, H.; Wilhelm, M.; Becher, I.; Lemeer, S.; Bantscheff, M.; Langridge, J. I.; Kuster, B., Ion Mobility Tandem Mass Spectrometry Enhances Performance of Bottom-up Proteomics. *Molecular & Cellular Proteomics* **2014,** 13, (12), 3709-3715.

70. Domon, B.; Aebersold, R., Review - Mass spectrometry and protein analysis. *Science* **2006,** 312, (5771), 212-217.

71. Kanu, A. B.; Dwivedi, P.; Tam, M.; Matz, L.; Hill, H. H., Ion mobility-mass spectrometry. *Journal of Mass Spectrometry* **2008,** 43, (1), 1-22.

72. Zhang, Y. Y.; Fonslow, B. R.; Shan, B.; Baek, M. C.; Yates, J. R., Protein Analysis by Shotgun/Bottom-up Proteomics. *Chemical Reviews* **2013,** 113, (4), 2343-2394.

73. Burden, D., Guide to the Disruption of Biological Samples 2012.pdf. *Random Primers* **2012**, (12), 1-25.

74. Bodzon-Kulakowska, A.; Bierczynska-Krzysik, A.; Dylag, T.; Drabik, A.; Suder, P.; Noga, M.; Jarzebinska, J.; Silberring, J., Methods for samples preparation in proteomic research. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **2007,** 849, (1-2), 1-31.

75. Twyman, R. M., *Principles of Proteomics*. 2004.

76. Lin, Y.; Zhou, J.; Bi, D.; Chen, P.; Wang, X. C.; Liang, S. P., Sodium-deoxycholate-assisted tryptic digestion and identification of proteolytically resistant proteins. *Analytical Biochemistry* **2008,** 377, (2), 259-266.

77. Kawasaki, H.; Suzuki, K., Separation of Peptides Dissolved in a Sodium Dodecyl-Sulfate Solution by Reversed-Phase Liquid-Chromatography - Removal of Sodium Dodecyl-Sulfate

from Peptides Using an Ion-Exchange Precolumn. *Analytical Biochemistry* **1990,** 186, (2), 264-268.

78. Loo, R. R. O.; Dales, N.; Andrews, P. C., Surfactant Effects on Protein-Structure Examined by Electrospray-Ionization Mass-Spectrometry. *Protein Science* **1994,** 3, (11), 1975-1983.

79. Proc, J. L.; Kuzyk, M. A.; Hardie, D. B.; Yang, J.; Smith, D. S.; Jackson, A. M.; Parker, C. E.; Borchers, C. H., A Quantitative Study of the Effects of Chaotropic Agents, Surfactants, and Solvents on the Digestion Efficiency of Human Plasma Proteins by Trypsin. *Journal of Proteome Research* **2010,** 9, (10), 5422-5437.

80. Boja, E. S.; Fales, H. M., Overalkylation of a protein digest with iodoacetamide. *Analytical Chemistry* **2001,** 73, (15), 3576-3582.

81. Winter, D.; Steen, H., Optimization of cell lysis and protein digestion protocols for the analysis of HeLa S3 cells by LC-MS/MS. *Proteomics* **2011,** 11, (24), 4726-4730.

82. Klammer, A. A.; MacCoss, M. J., Effects of modified digestion schemes on the identification of proteins from complex mixtures. *Journal of Proteome Research* **2006,** 5, (3), 695-700.

83. Norrgran, J.; Williams, T. L.; Woolfitt, A. R.; Solano, M. I.; Pirkle, J. L.; Barr, J. R., Optimization of digestion parameters for protein quantification. *Analytical Biochemistry* **2009,** 393, (1), 48-55.

84. Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C., Orthogonality of separation in two-dimensional liquid chromatography. *Analytical Chemistry* **2005,** 77, (19), 6426-6434.

85. Capelo, J. L.; Carreira, R.; Diniz, M.; Fernandes, L.; Galesio, M.; Lodeiro, C.; Santos, H. M.; Vale, G., Overview on modern approaches to speed up protein identification workflows relying on enzymatic cleavage and mass spectrometry-based techniques. *Analytica Chimica Acta* **2009,** 650, (2), 151-159.

86. Canas, B.; Pineiro, C.; Calvo, E.; Lopez-Ferrer, D.; Gallardo, J. M., Trends in sample preparation for classical and second generation proteomics. *Journal of Chromatography A* **2007,** 1153, (1-2), 235-258.

87. Hodge, K.; Ten Have, S.; Hutton, L.; Lamond, A. I., Cleaning up the masses: Exclusion lists to reduce contamination with HPLC-MS/MS. *Journal of Proteomics* **2013,** 88, 92-103.

88. Colin, H.; Guiochon, G., Introduction to Reversed-Phase High-Performance Liquid-Chromatography. *Journal of Chromatography* **1977,** 141, (3), 289-312.

89. Washburn, M. P.; Wolters, D.; Yates, J. R., Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology* **2001,** 19, (3), 242-247.

90. Gevaert, K.; Van Damme, J.; Goethals, M.; Thomas, G. R.; Hoorelbeke, B.; Demol, H.; Martens, L.; Puype, M.; Staes, A.; Vandekerckhove, J., Chromatographic isolation of methionine-containing peptides for gel-free proteome analysis - Identification of more than 800 Escherichia coli proteins. *Molecular & Cellular Proteomics* **2002,** 1, (11), 896-903.

91. Neilson, K. A.; Ali, N. A.; Muralidharan, S.; Mirzaei, M.; Mariani, M.; Assadourian, G.; Lee, A.; van Sluyter, S. C.; Haynes, P. A., Less label, more free: Approaches in label-free quantitative mass spectrometry. *Proteomics* **2011,** 11, (4), 535-553.

92. Shliaha, P. V.; Bond, N. J.; Gatto, L.; Lilley, K. S., Effects of Traveling Wave Ion Mobility Separation on Data Independent Acquisition in Proteomics Studies. *Journal of Proteome Research* **2013,** 12, (6), 2323-2339.

93. Wiesner, J.; Premsler, T.; Sickmann, A., Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications. *Proteomics* **2008,** 8, (21), 4466-4483.

94. Levin, Y., Label-free quantification of proteins using data-independent acquisition. In *New developments in mass spectrometry: quantitative proteomics*, Eyers, C. E.; Gaskell, S. J., Eds. 2014; pp 175-184.

95. Geromanos, S. J.; Vissers, J. P. C.; Silva, J. C.; Dorschel, C. A.; Li, G. Z.; Gorenstein, M. V.; Bateman, R. H.; Langridge, J. I., The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics* **2009,** 9, (6), 1683-1695.

96. Michalski, A.; Cox, J.; Mann, M., More than 100,000 Detectable Peptide Species Elute in Single Shotgun Proteomics Runs but the Majority is Inaccessible to Data-Dependent LC-MS/MS. *Journal of Proteome Research* **2011,** 10, (4), 1785-1793.

97. Distler, U.; Kuharev, J.; Navarro, P.; Levin, Y.; Schild, H.; Tenzer, S., Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics. *Nature Methods* **2014,** 11, (2), 167-170.

98. Gillet, L. C.; Navarro, P.; Tate, S.; Rost, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R., Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Molecular & Cellular Proteomics* **2012,** 11, (6).

99. Law, K. P.; Lim, Y. P., Recent advances in mass spectrometry: data independent analysis and hyper reaction monitoring. *Expert Review of Proteomics* **2013,** 10, (6), 551-566.

100. Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999,** 20, (18), 3551-3567.

101. Graves, P. R.; Haystead, T. A. J., Molecular biologist's guide to proteomics. *Microbiology and Molecular Biology Reviews* **2002,** 66, (1), 39-63.

102. Creasy, D. M.; Cottrell, J. S., Error tolerant searching of uninterpreted tandem mass spectrometry data. *Proteomics* **2002,** 2, (10), 1426-1434.

103. Vaudel, M.; Sickmann, A.; Martens, L., Current methods for global proteome identification. *Expert Review of Proteomics* **2012,** 9, (5), 519-532.

104. Elias, J. E.; Haas, W.; Faherty, B. K.; Gygi, S. P., Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nature Methods* **2005,** 2, (9), 667-675.

105. Kall, L.; Canterbury, J. D.; Weston, J.; Noble, W. S.; MacCoss, M. J., Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nature Methods* **2007,** 4, (11), 923-925.

106. NonlinearDynamics http://www.nonlinear.com/progenesis/qi-for-proteomics/how-it-works/.

107. Ludwig, C.; Aebersold, R., Getting absolute: determining absolute protein quantities via selected reaction monitoring mass spectrometry. In *New developments in mass spectrometry: Quantitative proteomics*, Eyers, C. E.; Gaskell, S. J., Eds. 2014; pp 80-109.

108. Ong, S. E.; Mann, M., Mass spectrometry-based proteomics turns quantitative. *Nature Chemical Biology* **2005,** 1, (5), 252-262.

109. Rauniyar, N.; McClatchy, D. B.; Yates, J. R., Stable isotope labeling of mammals (SILAM) for in vivo quantitative proteomic analysis. *Methods* **2013,** 61, (3), 260-268.

110. Kruger, M.; Moser, M.; Ussar, S.; Thievessen, I.; Luber, C. A.; Forner, F.; Schmidt, S.; Zanivan, S.; Fassler, R.; Mann, M., SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. *Cell* **2008,** 134, (2), 353-364.

111. Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics* **2002,** 1, (5), 376-386.

112. Young, V. R., Adult Amino-Acid-Requirements - the Case for a Major Revision in Current Recommendations. *Journal of Nutrition* **1994,** 124, (8), S1517-S1523.

113. Scott, L.; Lamb, J.; Smith, S.; Wheatley, D. N., Single amino acid (arginine) deprivation: rapid and selective death of cultured transformed and malignant cells. *British Journal of Cancer* **2000,** 83, (6), 800-810.

114. Boutilier, J. M.; Warden, H.; Doucette, A. A.; Wentzell, P. D., Chromatographic behaviour of peptides following dimethylation with H-2/D-2-formaldehyde: Implications for comparative proteomics. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **2012,** 908, 59-66.

115. Ong, S. E.; Mann, M., A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nature Protocols* **2006,** 1, (6), 2650-2660.

116. Bendall, S. C.; Hughes, C.; Stewart, M. H.; Doble, B.; Bhatia, M.; Lajoie, G. A., Prevention of amino acid conversion in SILAC experiments with embryonic stem cells. *Molecular & Cellular Proteomics* **2008,** 7, (9), 1587-1597.

117. Doherty, M. K.; Hammond, D. E.; Clagule, M. J.; Gaskell, S. J.; Beynon, R. J., Turnover of the Human Proteome: Determination of Protein Intracellular Stability by Dynamic SILAC. *Journal of Proteome Research* **2009,** 8, (1), 104-112.

118. Schwanhausser, B.; Gossen, M.; Dittmar, G.; Selbach, M., Global analysis of cellular protein translation by pulsed SILAC. *Proteomics* **2009,** 9, (1), 205-209.

119. Geiger, T.; Cox, J.; Ostasiewicz, P.; Wisniewski, J. R.; Mann, M., Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nature Methods* **2010,** 7, (5), 383-385.

120. Hanke, S.; Besir, H.; Oesterhelt, D.; Mann, M., Absolute SILAC for accurate quantitation of proteins in complex mixtures down to the attomole level. *Journal of Proteome Research* **2008,** 7, (3), 1118-1130.

121. Sethuraman, M.; McComb, M. E.; Huang, H.; Huang, S.; Heibeck, T.; Costello, C. E.; Cohen, R. A., Isotope-coded affinity tag (ICAT) approach to redox proteomics: identification and quantitation of oxidant-sensitive cysteine thiols in complex protein mixtures. *J Proteome Res* **2004,** 3, (6), 1228-33.

122. Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J., Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **2004,** 3, (12), 1154-69.

123. Hajkova, D.; Rao, K. C.; Miyagi, M., pH dependency of the carboxyl oxygen exchange reaction catalyzed by lysyl endopeptidase and trypsin. *J Proteome Res* **2006,** 5, (7), 1667-73.

124. Bensadek, D.; Nicolas, A.; Lamond, A. I., Next Generation Proteomics: PTMs in space and time In *New developments in mass spectrometry. Quantitative proteomics*, Eyers, C. E.; Gaskell, S. J., Eds. 2014; pp 233-256.

125. Thompson, A.; Schaefer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R. A. W.; Mohammed, A. K. A.; Hamon, C., Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS (vol 75, pg 1895, 2003). *Analytical Chemistry* **2006,** 78, (12), 4235-4235.

126. Zhou, Y.; Shan, Y. C.; Zhang, L. H.; Zhang, Y. K., Recent advances in stable isotope labeling based techniques for proteome relative quantification. *Journal of Chromatography A* **2014,** 1365, 1-11.

127. Ow, S. Y.; Salim, M.; Noirel, J.; Evans, C.; Rehman, I.; Wright, P. C., iTRAQ Underestimation in Simple and Complex Mixtures: "The Good, the Bad and the Ugly". *Journal of Proteome Research* **2009,** 8, (11), 5347-5355.

128. Zhu, W. H.; Smith, J. W.; Huang, C. M., Mass Spectrometry-Based Label-Free Quantitative Proteomics. *Journal of Biomedicine and Biotechnology* **2010**.

129. NonlinearDynamics http://blog.nonlinear.com/2015/05/11/spectral-counting-not/.

130. NonlinearDynamics http://www.nonlinear.com/progenesis/qi-for-proteomics/v2.0/faq/how-are-peptide-abundances-calculated.aspx.

131. Mountford, J. C., Human embryonic stem cells: origins, characteristics and potential for regenerative therapy. *Transfusion Medicine* **2008,** 18, (1), 1-12.

132. Liu, N.; Lu, M.; Tian, X.; Han, Z., Molecular mechanisms involved in self-renewal and pluripotency of embryonic stem cells. *J Cell Physiol* **2007,** 211, (2), 279-86.

133. Zwaka, T. P.; Thomson, J. A., Differentiation of human embryonic stem cells occurs through symmetric cell division. *Stem Cells* **2005,** 23, (2), 146-9.

134. Yu, J. Y.; Vodyanik, M. A.; He, P.; Slukvin, I. I.; Thomson, J. A., Human embryonic stem cells reprogram myeloid precursors following cell-cell fusion. *Stem Cells* **2006,** 24, (1), 168-176.

135. Aubin, J. E., Autofluorescence of Viable Cultured Mammalian-Cells. *Journal of Histochemistry & Cytochemistry* **1979,** 27, (1), 36-43.

136. O'Leary, T.; Heindryckx, B.; Lierman, S.; Van der Jeught, M.; Menten, B.; Deforce, D.; Cornelissen, R.; Lopes, S. C. D.; De Sutter, P., The Influence of Early Embryo Traits on Human Embryonic Stem Cell Derivation Efficiency. *Stem Cells and Development* **2011,** 20, (5), 785-793.

137. Levenstein, M. E.; Ludwig, T. E.; Xu, R. H.; Llanas, R. A.; VanDenHeuvel-Kramer, K.; Manning, D.; Thomson, J. A., Basic fibroblast growth factor support of human embryonic stem cell self-renewal. *Stem Cells* **2006,** 24, (3), 568-574.

138. *Handbook of Biological Confocal Microscopy, Third edition*. 2006; Vol. 236, p 988.

139. Gerrard, L.; Zhao, D.; Clark, A. J.; Cui, W., Stably transfected human embryonic stem cell clones express OCT4-specific green fluorescent protein and maintain self-renewal and pluripotency. *Stem Cells* **2005,** 23, (1), 124-33.

140. Wei, F.; Scholer, H. R.; Atchison, M. L., Sumoylation of Oct4 enhances its stability, DNA binding, and transactivation. *Journal of Biological Chemistry* **2007,** 282, (29), 21551-21560.

141. Katkov, I. I.; Kim, M. S.; Bajpai, R.; Altman, Y. S.; Mercola, M.; Loring, J. F.; Terskikh, A. V.; Snyder, E. Y.; Levine, F., Cryopreservation by slow cooling with DMSO diminished production of Oct-4 pluripotency marker in human embryonic stem cells. *Cryobiology* **2006,** 53, (2), 194-205.

142. Levin, Y., *New Developments in Mass Spectrometry: Quantitative Proteomics*. Royal Society of Chemistry: 2014; p 175-183.

143. Klammt, C.; Schwarz, D.; Fendler, K.; Haase, W.; Dotsch, V.; Bernhard, F., Evaluation of detergents for the soluble expression of alpha-helical and beta-barrel-type integral membrane proteins by a preparative scale individual cell-free expression system. *Febs Journal* **2005,** 272, (23), 6024-6038.

144. Xu, X. Y.; Mei, H.; Wang, S. Y.; Zhou, O.; Wang, G. F.; Broske, L.; Pena, A.; Korfmacher, W. A., A study of common discovery dosing formulation components and their potential for causing time-dependent matrix effects in high-performance liquid chromatography tandem mass spectrometry assays. *Rapid Communications in Mass Spectrometry* **2005,** 19, (18), 2643-2650.

145. Chamrad, D. C.; Koerting, G.; Gobom, J.; Thiele, H.; Klose, J.; Meyer, H. E.; Blueggel, M., Interpretation of mass spectrometry data for high-throughput proteomics. *Analytical and Bioanalytical Chemistry* **2003,** 376, (7), 1014-1022.

146. Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nature Methods* **2009,** 6, (5), 359-362.

147. Shevchenko, G.; Musunuri, S.; Wetterhall, M.; Bergquist, J., Comparison of Extraction Methods for the Comprehensive Analysis of Mouse Brain Proteome using Shotgun-based Mass Spectrometry. *Journal of Proteome Research* **2012,** 11, (4), 2441-2451.

148. Hustoft, H. K.; Reubsaet, L.; Greibrokk, T.; Lundanes, E.; Malerod, H., Critical assessment of accelerating trypsination methods. *Journal of Pharmaceutical and Biomedical Analysis* **2011,** 56, (5), 1069-1078.

149. Vowinckel, J.; Capuano, F.; Campbell, K.; Deery, M. J.; Lilley, K. S.; Ralser, M., The beauty of being (label)-free: sample preparation methods for SWATH-MS and next-generation targeted proteomics. *F1000Res* **2013,** 2, 272.

150. Zhang, N.; Li, L., Effects of common surfactants on protein digestion and matrix-assisted laser desorption/ionization mass spectrometric analysis of the digested peptides using two-layer sample preparation. *Rapid Communications in Mass Spectrometry* **2004,** 18, (8), 889-896.

151. Lin, Y.; Liu, Y.; Li, J. J.; Zhao, Y.; He, Q. Z.; Han, W. J.; Chen, P.; Wang, X. C.; Liang, S. P., Evaluation and optimization of removal of an acid-insoluble surfactant for shotgun analysis of membrane proteome. *Electrophoresis* **2010,** 31, (16), 2705-2713.

152. Leon, I. R.; Schwammle, V.; Jensen, O. N.; Sprenger, R. R., Quantitative Assessment of In-solution Digestion Efficiency Identifies Optimal Protocols for Unbiased Protein Analysis. *Molecular & Cellular Proteomics* **2013,** 12, (10), 2992-3005.

153. Yu, Y. Q.; Gilar, M.; Lee, P. J.; Bouvier, E. S. P.; Gebler, J. C., Enzyme-friendly, mass spectrometry-compatible surfactant for in-solution enzymatic digestion of proteins. *Analytical Chemistry* **2003,** 75, (21), 6023-6028.

154. Kawashima, Y.; Takahashi, N.; Satoh, M.; Saito, T.; Kado, S.; Nomura, F.; Matsumoto, H.; Kodera, Y., Enhanced recovery of lyophilized peptides in shotgun proteomics by using an LC-ESI-MS compatible surfactant. *Proteomics* **2013,** 13, (5), 751-755.

155. Chen, E. I.; Cociorva, D.; Norris, J. L.; Yates, J. R., Optimization of mass spectrometry-compatible surfactants for shotgun proteomics. *Journal of Proteome Research* **2007,** 6, (7), 2529-2538.

156. Masuda, T.; Tomita, M.; Ishihama, Y., Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *Journal of Proteome Research* **2008,** 7, (2), 731-740.

157. Gasteiger E., H. C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A., *Protein Identification and Analysis Tools on the ExPASy Server*. Humana Press, 2005.

158. Shieh, I. F.; Lee, C. Y.; Shiea, J., Eliminating the interferences from TRIS buffer and SDS in protein analysis by fused-droplet electrospray ionization mass spectrometry. *Journal of Proteome Research* **2005,** 4, (2), 606-612.

159. Ong, S. E.; Kratchmarova, I.; Mann, M., Properties of C-13-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *Journal of Proteome Research* **2003,** 2, (2), 173-181.

160. Hughes, C.; Radan, L.; Chang, W. Y.; Stanford, W. L.; Betts, D. H.; Postovit, L. M.; Lajoie, G. A., Mass Spectrometry-based Proteomic Analysis of the Matrix Microenvironment in Pluripotent Stem Cell Culture. *Molecular & Cellular Proteomics* **2012,** 11, (12), 1924-1936.

161. Liberski, A. R.; Al-Noubi, M. N.; Rahman, Z. H.; Halabi, N. M.; Dib, S. S.; Al-Mismar, R.; Billing, A. M.; Krishnankutty, R.; Ahmad, F. S.; Raynaud, C. M.; Rafii, A.; Engholm-Keller, K.; Graumann, J., Adaptation of a commonly used, chemically defined medium for human embryonic stem cells to stable isotope labeling with amino acids in cell culture. *J Proteome Res* **2013,** 12, (7), 3233-45.

162. Scheerlinck, E.; Van Steendam, K.; Vandewoestyne, M.; Lepez, T.; Gobin, V.; Meert, P.; Vossaert, L.; Van Nieuwerburgh, F.; Van Soom, A.; Peelman, L.; Heindryckx, B.; De Sutter, P.; Dhaenens, M.; Deforce, D., Detailed method description for noninvasive monitoring of differentiation status of human embryonic stem cells. *Anal Biochem* **2014,** 461, 60-6.

163. Scheerlinck, E.; Dhaenens, M.; Van Soom, A.; Peelman, L.; De Sutter, P.; Van Steendam, K.; Deforce, D., Minimizing technical variation during sample preparation prior to label-free quantitative mass spectrometry. *Anal Biochem* **2015**, 490, 14-19

164. Croft, D.; Mundo, A. F.; Haw, R.; Milacic, M.; Weiser, J.; Wu, G.; Caudy, M.; Garapati, P.; Gillespie, M.; Kamdar, M. R.; Jassal, B.; Jupe, S.; Matthews, L.; May, B.; Palatnik, S.; Rothfels, K.; Shamovsky, V.; Song, H.; Williams, M.; Birney, E.; Hermjakob, H.; Stein, L.; D'Eustachio, P., The Reactome pathway knowledgebase. *Nucleic Acids Res* **2014,** 42, (Database issue), D472-7.

165. Milacic, M.; Haw, R.; Rothfels, K.; Wu, G.; Croft, D.; Hermjakob, H.; D'Eustachio, P.; Stein, L., Annotating cancer variants and anti-cancer therapeutics in reactome. *Cancers (Basel)* **2012,** 4, (4), 1180-211.

166. Vizcaino, J. A.; Deutsch, E. W.; Wang, R.; Csordas, A.; Reisinger, F.; Rios, D.; Dianes, J. A.; Sun, Z.; Farrah, T.; Bandeira, N.; Binz, P. A.; Xenarios, I.; Eisenacher, M.; Mayer, G.; Gatto, L.; Campos, A.; Chalkley, R. J.; Kraus, H. J.; Albar, J. P.; Martinez-Bartolome, S.; Apweiler, R.; Omenn, G. S.; Martens, L.; Jones, A. R.; Hermjakob, H., ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nature Biotechnology* **2014,** 32, (3), 223-226.

167. Greenebaum, C. C. from http://www.umgcc.org/research/proteomics_services.htm.

168. Salway, J. G., *Metabolism at a Glance (second edition)*. Blackwell Publishing: 1999; p 111.

169. Tenu, J. P.; Lepoivre, M.; Moali, C.; Brollo, M.; Mansuy, D.; Boucher, J. L., Effects of the new arginase inhibitor N omega-hydroxy-nor-L-arginine on NO synthase activity in murine macrophages. *Nitric Oxide-Biology and Chemistry* **1999,** 3, (6), 427-438.

170. Ash, D. E., Structure and function of arginases. *Journal of Nutrition* **2004,** 134, (10), 2760s-2764s.

171. Napoli, C.; Paolisso, G.; Casamassimi, A.; Al-Omran, M.; Barbieri, M.; Sommese, L.; Infante, T.; Ignarro, L. J., Effects of nitric oxide on cell proliferation: novel insights. *J Am Coll Cardiol* **2013,** 62, (2), 89-95.

172. Becker-Catania, S. G.; Gregory, T. L.; Yang, Y.; Gau, C. L.; de Vellis, J.; Cederbaum, S. D.; Yer, R. K., Loss of arginase I results in increased proliferation of neural stem cells. *Journal of Neuroscience Research* **2006,** 84, (4), 735-746.

173. Rosner, B., Percentage Points for a Generalized Esd Many-Outlier Procedure. *Technometrics* **1983,** 25, (2), 165-172.

174. Menendez, P.; Bueno, C.; Wang, L., Human embryonic stem cells: a journey beyond cell replacement therapies. *Cytotherapy* **2006,** 8, (6), 530-541.

175. ThermoFisher, Alkaline Phosphatase, https://www.thermofisher.com/order/catalog/product/A14353?ICID=search-product.

176. Abcam, Alkaline Phosphatase, http://www.abcam.com/alkaline-phosphatase-assay-kit-fluorometric-ab83371.html.

177. Carey, B. W.; Finley, L. W. S.; Cross, J. R.; Allis, C. D.; Thompson, C. B., Intracellular alpha-ketoglutarate maintains the pluripotency of embryonic stem cells. *Nature* **2015,** 518, (7539).

178. Glibert, P.; Meert, P.; Van Steendam, K.; Van Nieuwerburgh, F.; De Coninck, D.; Martens, L.; Dhaenens, M.; Deforce, D., Phospho-iTRAQ: Assessing Isobaric Labels for the Large-Scale Study Of Phosphopeptide Stoichiometry. *Journal of Proteome Research* **2015,** 14, (2), 839-849.

179. Shechter, D.; Dormann, H. L.; Allis, C. D.; Hake, S. B., Extraction, purification and analysis of histones. *Nature Protocols* **2007,** 2, (6), 1445-1457.

180. Glibert, P.; Van Steendam, K.; Dhaenens, M.; Deforce, D., iTRAQ as a method for optimization: Enhancing peptide recovery after gel fractionation. *Proteomics* **2014,** 14, (6), 680-684.

CURRICULUM VITAE

GENERAL INFORMATION

Name : ELLEN SCHEERLINCK
Residence : Tijl en Nelestraat 9, 8750 Wingene
Date of birth : 20/03/1987
Place of birth : Brugge
Linked in : https://be.linkedin.com/pub/ellen-scheerlinck/a9/718/b62

EDUCATION & ACADEMIC ACTIVITIES

2010-present: PhD Student in Pharmaceutical Science
Laboratory of Pharmaceutical Biotechnology (Prof. Dr. Dieter Deforce)
Ghent University.

Dissertation: *Human embryonic stem cells: from the follow-up of pluripotency to quantitative peptide analysis*.

2008-2010: Master of Drug Development
Ghent University
Great Distinction

Dissertation: *Radiosynthese en evaluatie van [^{11}C]-MC80 voor beeldvorming van P-glycoproteïnen* – Laboratory of Radiofarmacy, Ghent, Belgium

2005-2008: Bachelor Pharmaceutical Sciences
Ghent University
Distinction

1999-2005 Sciences and Mathematics
Sint-Lodewijkscollege, Brugge

Training courses

1. PhD career focus life sciences and medicine, Doctoral Schools Life Sciences & Medicine, UGent, 2014
2. Case studies in quantitative proteome techniques, ASMS, Baltimore, 2014
3. TripleTOF training, Sciex, 2013
4. Human pluripotent stem cell training, Life Technologies, 2012
5. Basic bioinformatics, databases and tools, VIB, 2011
6. Advanced Academic English: Conference Skills – Presentation Skills in English, 14/20, Doctoral Schools Life Sciences & Medicine, UGent, 2011

7. Advanced Academic English: Writing Skills, 16/20, Doctoral Schools Life Sciences & Medicine, UGent, 2011

SCIENTIFIC CURRICULUM

- Articles in peer-reviewed journals

Scheerlinck, E.; Dhaenens, M.; Van Soom, A.; Peelman, L.; De Sutter, P.; Van Steendam, K.; Deforce, D., Minimizing technical variation during sample preparation prior to label-free quantitative mass spectrometry. *Anal Biochem* **2015**.

Paulien Meert, Elisabeth Govaert, Ellen Scheerlinck, Maarten Dhaenens, Dieter Deforce, Pitfalls in histone propionylation during bottom-up mass spectrometry analysis, Proteomics, 0 (2015) 1-6.

E. Scheerlinck, K. Van Steendam, M. Vandewoestyne, T. Lepez, V. Gobin, P. Meert, L. Vossaert, F. Van Nieuwerburgh, A. Van Soom, L. Peelman, B. Heindryckx, P. De Sutter, M. Dhaenens, D. Deforce, Detailed method description for noninvasive monitoring of differentiation status of human embryonic stem cells, Anal Biochem, 461 (2014) 60-66.

L. Vossaert, P. Meert, E. Scheerlinck, P. Glibert, N. Van Roy, B. Heindryckx, P. De Sutter, M. Dhaenens, D. Deforce, Identification of histone H3 clipping activity in human embryonic stem cells, Stem Cell Res, 13 (2014) 123-134.

- Manuscripts in submission/in preparation

Scheerlinck, Ellen; Van Steendam, Katleen; Govaert, Elisabeth; Vossaert, Liesbeth; Meert, Paulien; Van Nieuwerburgh, Filip; van Soom, Ann; Peelman, Luc; De Sutter, Petra; Heindryckx, Björn; Dhaenens, Maarten; Deforce, Dieter. The development of a fully defined SILAC culture medium with minimal arginine conversion in human embryonic stem cells, *submitted to Journal of Proteome Research.*

Govaert, Elisabeth; Van Steendam, Katleen; Scheerlinck, Ellen; Vossaert, Liesbeth; Meert, Paulien; Stella, Martina; Willems, Sander; De Clerck, Laura; Dhaenens, Maarten; Deforce, Dieter. Preparing histones for label-free quantitative mass spectrometry: a comparison of extraction protocols, *in preparation.*

- Poster presentations on national and international conference

- "From cell lysate to peptide analysis: optimization of sample preparation prior to mass spectrometry."
E. Scheerlinck; M. Dhaenens; A. Van Soom; L. Peelman; P. De Sutter; K. Van Steendam; D. Deforce
"*Belgian Proteomics Association" Conference*", Brussels, Belgium, December 18-19, 2014.

- “Pimp your (propionylation) protocol: progenesis QI uncouples protocol comparison from the bias of targeted data analysis”
Meert P, Scheerlinck E, Govaert E, Dhaenens M, Deforce D
“Belgian Proteomics Association” Conference, Brussels, Belgium, December 18-19, 2014.

- “Evaluation of histone extraction protocols for label-free mass spectrometry”
Govaert E, Van Steendam K, Scheerlinck E, Meert P, Vossaert L, Dhaenens M, Deforce D
“Belgian Proteomics Association” Conference, Brussels, Belgium, December 18-19, 2014.

- “Optimization of the use of Stable Isotopic Labeling of Amino aCids (SILAC) for quantifying peptide changes in 2 different human embryonic stem cell conditions.”
E. Scheerlinck; K. Van Steendam; E. Govaert; A. Van Soom; L. Peelman; P. De Sutter; M. Dhaenens; D. Deforce
(1) *“First meeting of Belgian Society for Stem Cell Research”*, Gent, Belgium, September 12, 2014.
(2) “*Belgian Proteomics Association” Conference*”, Brussels, Belgium, December 18-19, 2014.

- “Histone H3 clipping in hESC in relation to OCT4 expression and culture conditions”
Meert P.*, Vossaert L.*, Scheerlinck E., Glibert P., Van Roy N., Heindryckx B., De Sutter P., Dhaenens M. & Deforce D.
“First meeting of Belgian Society for Stem Cell Research”, Gent, Belgium, September 12, 2014.

- “The removal of LC-MS contaminants in shotgun proteomics.”
Ellen Scheerlinck; Katleen Van Steendam; Maarten Dhaenens; Dieter Deforce
“*American Society for Mass Spectrometry*”, Baltimore, USA, June 19, 2014.

- “Validation of a non-destructive method to examine human embryonic stem cell (hESC) (non)-differentiation with a OCT4-eGFP Knock In hESC line.”
Scheerlinck, E; Dhaenens, M; Vandewoestyne, M; Van Steendam, K; Lepez, T; Gobin, V; Meert, P; Vossaert, L; Van Nieuwerburgh, F; Van Soom, A; Peelman, L; Heindryckx, B; De Sutter, P; Deforce, D
“*Knowledge For Growth FlandersBio*”, ICC Ghent, Belgium, May 30, 2013.

- Conference attendances

- Two-day symposium of the Belgian Proteomics Association, Brussels, Belgium, December 18-19, 2014.

- 1st meeting of the Belgian Society for Stem Cell Research, Ghent, Belgium, September 12, 2014.

- American Society for Mass Spectrometry, Baltimore, USA, June 16-19, 2014.

- Knowledge For Growth, Flanders Bio, Ghent, Belgium, Mai 30, 2013.

- Two-day symposium of the Belgian Proteomics Association, Ghent, Belgium, November 29-30, 2012.

- Fourth International Symposium on Proteome Analysis, Antwerp, Belgium, December 16-17, 2010.

➢ Educational experience

Practical courses

Practical course Phytochemistry (2nd Bachelor students).
Practical course Pharmaceutical Biotechnology (3rd Bachelor students).

Supervised dissertations

1st master, Faculty Pharmaceutical Sciences, UGent:

2013-2014: Sarah Callebaut: "Optimalisatie van de staalvoorbereiding voor LC-MS/MS analyse op proteomics niveau."

2013-2014: Simon Daled: "Inhibitie van de arginine conversie bij stable isotope labeling by amino acids in cell culture (SILAC) op humane embryonale stamcellen (hESC)."

2012-2013: Eveline Braet: "Het gebruik van de fluorescentiemicroscoop als tool voor het screenen van muis embryonaal fibroblast geconditioneerd medium."

2011-2012: Karen Depraetere: "Optimalisatie van de in vitro cultuur van humane embryonale stamcellen (hESC)."

ManaMa Master of Science in Industrial Pharmacy, Faculty Pharmaceutical Sciences, UGent:

2013-2014: Soumia Bettioui: "De optimalisatie van SILAC voor het gebruik bij hESC."