

RESEARCH ARTICLE

Open Access

# Attempting to distinguish between endogenous and contaminating cytokeratins in a corneal proteomic study

Mikkel Lyngholm<sup>1\*</sup>, Henrik Vorum<sup>2,3</sup>, Kim Nielsen<sup>1</sup>, Niels Ehlers<sup>1</sup>, Bent Honoré<sup>2</sup>

## Abstract

**Background:** The observation of cytokeratins (CK's) in mass spectrometry based studies raises the question of whether the identified CK is a true endogenous protein from the sample or simply represents a contaminant. This issue is especially important in proteomic studies of the corneal epithelium where several CK's have previously been reported to mark the stages of differentiation from corneal epithelial stem cell to the differentiated cell.

**Methods:** Here we describe a method to distinguish very likely endogenous from uncertain endogenous CK's in a mass spectrometry based proteomic study. In this study the CK identifications from 102 human corneal samples were compared with the number of human CK identifications found in 102 murine thymic lymphoma samples.

**Results:** It was anticipated that the CK's that were identified with a frequency of <5%, *i.e.* in less than one spot for every 20 spots analysed, are very likely to be endogenous and thereby represent a 'biologically significant' identification. CK's observed with a frequency >5% are uncertain endogenous since they may represent true endogenous CK's but the probability of contamination is high and therefore needs careful consideration. This was confirmed by comparison with a study of mouse samples where all identified human CK's are contaminants.

**Conclusions:** CK's 3, 4, 7, 8, 11, 12, 13, 15, 17, 18, 19, 20 and 23 are very likely to be endogenous proteins if identified in a corneal study, whilst CK's 1, 2e, 5, 6A, 9, 10, 14 and 16 may be endogenous although some are likely to be contaminants in a proteomic study. Further immunohistochemical analysis and a search of the current literature largely supported the distinction.

## Background

Cytokeratins (CK's) belong to the family of intermediate filaments, and are expressed in a variety of different cell types, including those of the eye. CK's can be classified into type I (acidic) and type II (basic), and they often appear together as pairs of these two types of proteins [1,2]. Within the eye in particular, CK's have been shown to be important proteins with regard to cellular development, proliferation and differentiation [3-7].

Proteomics using two-dimensional gel electrophoresis (2D-PAGE) in combination with mass spectrometry (MS) is often used as a screening tool in the search for differentially expressed proteins [8]. If CK's are the subject of such an experiment, an important issue to

determine is whether the CK originates from the investigated tissue or whether its presence is a consequence of contamination from the laboratory environment.

As a consequence of the very high cell turn-over of the surface epithelia the environment is contaminated with cells from hair, skin, nails, eyebrows, eyelashes, airways, etc. Furthermore, CK's remain ubiquitous contaminants even in laboratories with very high cleaning standards being extremely difficult to eliminate [9], though their presence in MS laboratories should be reduced as much as possible [10]. One way to diminish the environmental contribution of CK's is to filter all the liquids used for 2D-PAGE, including the reducing agents [11] and to have efficient laboratory protocols to maintain clean conditions in general [12].

In the search for stem cell markers, we previously investigated the differences in protein expression between the central corneal epithelium and the limbal

\* Correspondence: mikkel.lyngholm@ki.au.dk

<sup>1</sup>Department of Ophthalmology, Aarhus University Hospital, Nørrebrogade 44, 8000 Aarhus C, Denmark

Full list of author information is available at the end of the article

epithelium by a proteomic approach, in an effort to identify proteins either highly expressed or exclusively present in limbal epithelium [13]. The limbus, which is located between the conjunctiva and the cornea, is a niche for corneal epithelial stem cells. Several CK's have been proposed as markers of stem cells, transient amplifying cells or differentiated cells [5,7,14]. Therefore, we wanted to include biologically significant CKs in the investigation in the search for limbal stem cell markers.

This study attempts to discuss whether it is possible by a simple procedure to distinguish between very likely endogenous and uncertain endogenous CK's by counting and evaluating all the human CK's identified by mass spectrometry in a human and murine study. These results are compared with previous reports on location and distribution of CK's in the ocular surface tissue. Further examination was performed by evaluating the specific mass spectra of the cytokeratins in conjunction with immunohistochemistry.

## Methods

### Human sample collection and preparation

Seven human eyes were obtained from The Institute of Anatomy, Aarhus University within 48 hours post mortem. The study adhered with the guidelines from the local ethical committee and the tenets of the Declaration of Helsinki. The eyes were carefully rinsed in sterile isotonic saline and the epithelium was marked with 8 and 10 mm trephines prior to loosening the epithelium by dabbing with 70% alcohol. Under clean conditions, the corneal epithelium was gently scraped to separate the central 8 mm epithelium. The intermediate epithelium (8-10 mm ring) was discarded and the limbal fractions were considered to be the epithelium scraped outside the 10 mm ring since, the conjunctival epithelium was not able to be as easily loosened by this method. Both fractions were transferred to lysis buffer, pH 3-10 NL [13].

### Mouse sample collection and preparation

Two spontaneously developed thymic lymphomas, SM5 and SM7, from C57BL/6J-Trp53tm1Tyj mice deficient for the p53 gene were explanted, *in vitro* cultured, and established as cell lines growing in RPMI-1640 culture medium supplemented with 10% fetal calf serum and 50  $\mu$ M 2-mercaptoethanol. Freshly prepared thymocytes obtained from pools of five mice (either C57BL/6J-Trp53tm1Tyj or normal C57BL/6) and cultured SM5 and SM7 cells were washed extensively in PBS and subsequently freeze dried. The cell pellets were dissolved in lysis buffer, pH 3-10 NL [8].

### 2D-PAGE and MS identification

2D-PAGE and silver staining was performed as previously described [8,13]. The gels were re-hydrated and

the cellophane sheets peeled off prior to protein gel spot excision. Gel pieces were dehydrated in acetonitrile, dried and the proteins reduced for 1 h at 56°C in 10 mM dithiothreitol (DTT) and 100 mM  $\text{NH}_4\text{HCO}_3$ . The solution was exchanged with 55 mM iodoacetamide in 100 mM  $\text{NH}_4\text{HCO}_3$  for 45 min. Gel pieces were then washed in 100 mM  $\text{NH}_4\text{HCO}_3$ , dehydrated in acetonitrile, rehydrated in 100 mM  $\text{NH}_4\text{HCO}_3$ , dehydrated in acetonitrile, dried and swelled in digestion buffer (50 mM  $\text{NH}_4\text{HCO}_3$ , 5 mM  $\text{CaCl}_2$  and 12.5 ng/ $\mu$ l trypsin Gold (mass spectrometry grade; Promega, Madison, WI, USA). The digestion was performed overnight at 37°C prior to peptide extraction by 1 change of 20 mM  $\text{NH}_4\text{HCO}_3$  and 3 changes of 5% formic acid in 50% acetonitrile. The sample was dried and peptides resuspended in buffer A (water/acetonitrile/formic acid, 97.7/2/0.3, V/V/V). The peptides were separated on an inert nano LC system of a Famos micro autosampler, a Switchos micro column switching module and an Ultimate micro pump from LC Packings (San Francisco, CA) before MS analysis. Samples were concentrated and desalted on a 300  $\mu$ m inner diameter x 5 mm precolumn (LC Packings) packed with 5  $\mu$ m C18 PepMap100 material. A 75  $\mu$ m inner diameter x 15 cm Nano column packed with 3  $\mu$ m C18 PepMap100 material was used to separate the peptides. Gradient elution from the column was performed by mixing decreasing volumes of buffer A with increasing volumes of buffer B (water/acetonitrile/formic acid, 9.7/90/0.3, V/V/V). The peptides were eluted into the nano electrospray ion source of the quadrupole time-of-flight mass spectrometer (Micromass, Manchester, UK). MS survey scans were acquired using MassLynx 4 SP4 (Waters). The instrument was operated in a data-dependent MS to MS/MS switching mode. Doubly, triply and quadruply charged peptide ions detected in MS survey scans triggered a switch to MS/MS for obtaining peptide fragmentation spectra. The processed data were used to search in the Swiss-Prot Database (version 56.9) using the on-line version of the Mascot MS/MS Ion Search facility (Matrix Science, Ltd., <http://www.matrixscience.com>) [15]. Searching was performed with doubly and triply charged ions with 2 missed cleavages, a peptide mass tolerance of 50 ppm, one variable modification, Carbamidomethyl-C and an MS/MS tolerance of  $\pm$  0.02 Da. Only human proteins identified by bold red peptides were regarded as significant and reported (excluding duplicate homologous proteins).

### Evaluation of the LC-MS/MS identifications

All significant human CK hits from the Mascot search were counted in each sample of the human and mouse study (Table 1) and the mass spectra were evaluated. The CK's were grouped as very likely endogenous

**Table 1 Frequency of human cytokeratins identified in spots from 2D-PAGE in a human corneal study and mouse thymic lymphoma study**

CKs	Type I(acidic)/II(neutral or basic)	Previously described in human non-tumour-tissue	Previously described in ocular epithelium	Location <sup>1</sup>	Total count in 102 mouse spots - %
<i>Very likely endogenous (observed in &lt;5% of human samples)</i>					
3	II		Cornea <sup>19,17</sup>	B	0%
4	II	Sebaceous glands, airway epithelium <sup>19</sup>	Conjunctiva <sup>17</sup> , cornea <sup>5</sup>	B	4%
7	II	Glands <sup>19</sup>	Central basal cornea <sup>17</sup> No found <sup>20</sup>	-	0%
8	II	Glands <sup>19</sup>	Cornea <sup>21</sup> , conjunctiva <sup>17</sup>	-	0%
11	I	Epidermis <sup>19</sup>		-	0%
12	I		Cornea <sup>19,2</sup>	C	0%
13	I	Glands, airway epithelium <sup>19,2</sup>	Conjunctiva <sup>25</sup> , cornea <sup>5</sup>	B	2%
15	I	Basal keratinocytes <sup>2</sup>	Limbal basal <sup>7</sup>	L	2%
17	I	Hair follicle, nails <sup>2</sup>	Cornea <sup>17</sup>	-	0%
18	I	Glands, simple epithelium <sup>19,2</sup>	Cornea <sup>21</sup>	-	0%
19	I	Glands, airway epithelium <sup>19</sup>	Conjunctiva <sup>4</sup> , basal limbus <sup>2</sup>	L	0%
20	I	Gastrointestinal tract epithelium <sup>2</sup>		-	0%
23	I	Non (found in pancreatic tumours) <sup>22</sup>		-	0%
<i>Uncertain endogenous (observed in &gt;5% of human samples)</i>					
1	II	Epidermis <sup>19</sup>		B	36%
2e	II	Epidermis <sup>19</sup>		B	78%
5	II	Epidermis, glands, airway epithelium, hair follicles <sup>19</sup>	Conjunctiva <sup>25</sup> , cornea <sup>19, 24</sup>	B	3%
6A	II	Glands, airway epithelium, hair follicles <sup>19</sup>		B	27%
9	I	Palmoplantar epidermis <sup>2</sup>		B	99%
10	I	Epidermis <sup>19</sup>		B	75%
14	I	Epidermis, glands, airway epithelium, hair follicles <sup>19</sup>	Conjunctiva <sup>25</sup> , cornea <sup>17</sup>	B	18%
16	I	Epidermis <sup>19</sup>	Cornea <sup>4</sup>	B	6%

<sup>1</sup>Location: C, central fraction, L, limbal fraction, B, both limbal and central fraction.

(observed in less than 5% of the spots) or uncertain endogenous (observed in more than 5% of the spots) on the basis of the observed frequency in the human study.

### Immunohistochemistry

Antibodies against CK 3/12 and 19 were obtained from Chemicon (Billerica, MA), and antibodies for CK 15 were obtained from Santa Cruz Biotechnology (Santa Cruz, CA). The immunohistochemical protocol has previously been published [13].

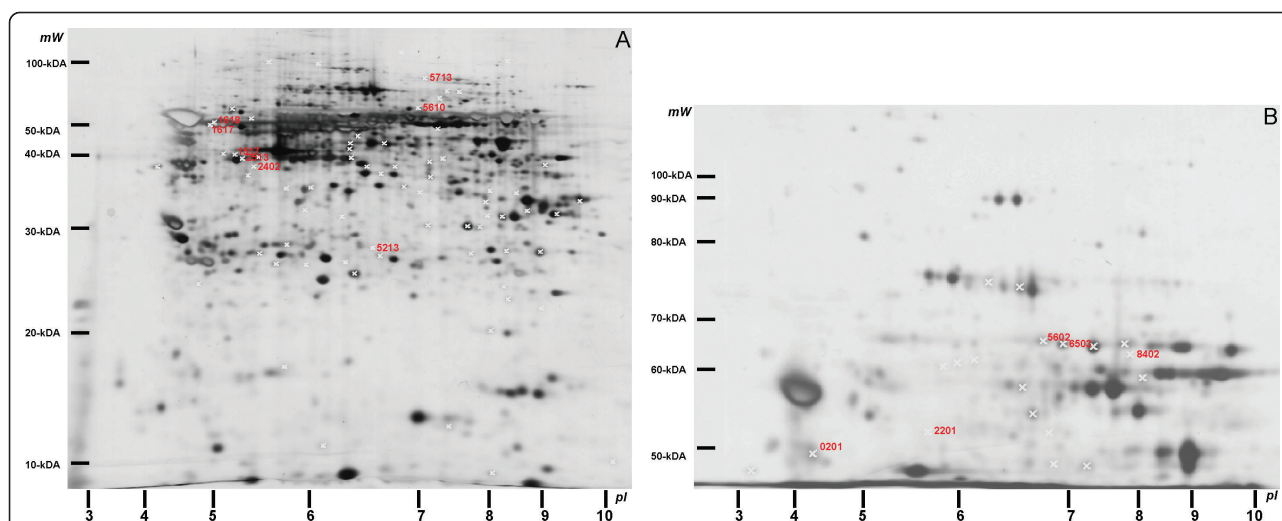
### Statistical analyses

An unpaired non-parametric test (Mann-Whitney) was performed to test for significant differences between CK expression in the human and mouse studies.

### Results and Discussion

We hypothesised that the higher the frequency that a certain CK is identified by MS from a group of 2D-PAGE spots the more likely it is to be a contaminating protein. A set of 102 spots from the human study [13]

(Figure 1) was included in the investigation in conjunction with a set of 102 mouse spots [16]. In Table 1, CKs are listed in two groups based on the observed frequency of the CK in the human study; very likely endogenous (observed in less than 5% of the spots) and uncertain endogenous (observed in more than 5% of the spots) together with previously published information on human tissue expression of each CK. The corresponding frequencies in the mouse study are also shown. Thirteen CK's were either not detected at all in any spot or found in less than 5% of the spots in the human study; CK's 3, 4, 7, 8, 11, 12, 13, 15, 17, 18, 19, 20 and 23 correlating with the mouse study (<5%). Eight CK's were identified in more than 5% of the human samples; CK's 1, 2e, 5, 6A, 9, 10, 14 and 16. These were also seen often in the mouse study except for CK5, which was observed with a frequency of 3%. There were no differences between the CK expression in the human and mouse study in the combined groups ( $p = 0.51$ ), the group of very likely endogenous ( $p = 0.34$ ) or the group of uncertain endogenous ( $p = 0.72$ ). We expected to



**Figure 1** 2D-gels (A:12%; B:6%) from the limbal fraction shows the significant spots (marked with the SSP number). The approximate pI and molecular mass is indicated on the axes. SSP No. 0103 and 1205 were not expressed on this gel (B:6%).

find some of the very likely endogenous CK's in the mouse study since several of these proteins, in addition to the expression in the human corneal epithelium, also are found in skin, hair and other tissues, largely explaining the source of contamination from the laboratory environment. Table 2 shows the theoretical and observed molecular mass (M) and pI of each very likely endogenous CK that was identified in the human study. In general, a higher observed molecular mass than theoretical molecular mass may indicate post-translational

modification, the protein is a precursor or alternatively a contaminant. A lower observed molecular mass than theoretical molecular mass is likely to be a cleavage product. The observed molecular masses listed in Table 2 are largely equal to or below the theoretical masses. The coverage of each observed CK in the likely endogenous group is shown in Table 2 together with the coverage from the first amino acid in the most N-terminal identified peptide to the last amino acid in the most C-terminal identified peptide in the protein. In none of the cases do we find

**Table 2** Characterization of some very likely endogenous cytokeratins

CKs	Spot number in gels (6%=',12%='*)	UniProtKB/Swiss-Prot	Observed M (kDA)	Theoretical M (kDA)	Observed pI	Theoretical pI	Coverage (%)	Fraction of protein covered (%) <sup>1</sup>
3	6503'	K2C3_HUMAN	64	64.5	6.90	6.12	12	56
	5213*	P12035	28		6.60		3	34
	5610*		60		7.00		31	88
	5713*		80		7.05		9	56
4	8402'	K2C4_HUMAN	62	57.3	7.80	6.25	6	21
	5602'	P19013	65		6.70		4	57
	5610*		60		7.00		3	24
12	0201'	K1C12_HUMAN Q99456	51	53.4	4.30	4.70	9	40
13	0301'	K1C13_HUMAN	56	49.6	4.40	4.91	25	71
	1205'	P13646	56		4.80		35	71
	1617*		50		5.05		16	64
	1618*		50		5.10		34	90
	2402*		38		5.50		5	19
15	2201'	K1C15_HUMAN P19012	52	49.2	5.60	4.71	16	78
19	2513*	K1C19_HUMAN	39	44.1	5.40	5.04	23	81
	1517*	P08727	40		5.30		17	81

<sup>1</sup>This was determined by calculating the fraction of the protein covered from the first amino acid in the most N-terminal peptide to the last amino acid in the most C-terminal peptide observed.

the fraction of protein covered to be significantly larger than in accordance with the observed molecular mass, in keeping with the identifications being very likely endogenous. Identifications tend to be more reliable when the sequence coverage is large or based on a few high quality spectra. Only identifications “in bold red” were included (see above).

#### Very likely endogenous CK's

Among the CK's observed in less than 5% of the human samples, CK 15 (49 kDa) was identified in one spot from the limbal epithelial fraction in the 50-kDa area. The proteomic identification was confirmed by immunohistochemical staining (Figure 2A). Indeed, CK 15 has recently been found in the basal limbal epithelium [7]. CK 19 (44 kDa) was identified in 2% of the spots, all focused in the 40-kDa area in the limbal epithelial fractions (Figure 1). CK 19 is a well-characterized marker for basal limbal epithelial cells (Figure 2B), and is furthermore found in simple epithelia and epidermal appendages [2,5,17]. CK 15 and CK 19 were regarded as very likely endogenous because of the distinct expression in only the limbal fraction, and also because they were identified in only a few spots (<2%). In addition, both proteins were identified by peptides covering around 80% of the protein as defined by the most N-terminal to the most C-terminal peptides observed (Table 2, Figure 3A). Human CK 15 was also observed in two mouse samples. CK 15 is known to be expressed in basal epithelia and hair follicle cells [18], which thus are likely to be rare contaminants in the laboratory environment.

CK 3 and CK 12 were identified in 4% and 1% of the human samples respectively and are both specific markers for differentiated corneal epithelial cells (Figure 2C) [5,17,19]. A mass spectrum from a 28-kDa spot identified CK3 by two peptides from the C-terminal of the protein (64.5 kDa) and thereby could be an identification of an endogenous protein. However, we cannot exclude that this is a cleaved contaminating protein introduced before electrophoresis (Figure 3B).

CK 4 and CK 13 were identified in 3% and 5% of the spots in the human set. Both proteins have previously been identified in the ocular surface (immunohistochemically), in addition to dermal glands and airway epithelium [2,5,19]. A 60-kDa spot was identified as CK 4 by two peptides in the C-terminal part of the protein (Figure 3C). This could be a cleaved endogenous protein, however, contamination cannot be excluded. CK 4 and CK13 were observed in 4 and 2 samples in the mouse study respectively, indicating that they may be rare contaminants.

Some of the keratins were rarely identified, some were not observed and among these some have to our

knowledge not been reported to be localized in the cornea, i.e. CKs 11, 20 and 23 (table 1). Thus, if identified, they would be very likely endogenous CK's [20-22].

#### Uncertain endogenous CK's

CK 16 appeared in 6% of the spots in both the human and the mouse study. It is expressed in palmoplantar epidermis, in epidermal appendage and in mucosa [2,19,23]. However, CK 16 has also been shown to be expressed in corneal epithelium to a minor extent [4]. None of the mass spectra proved useful in determining the origin of the contamination (data not shown).

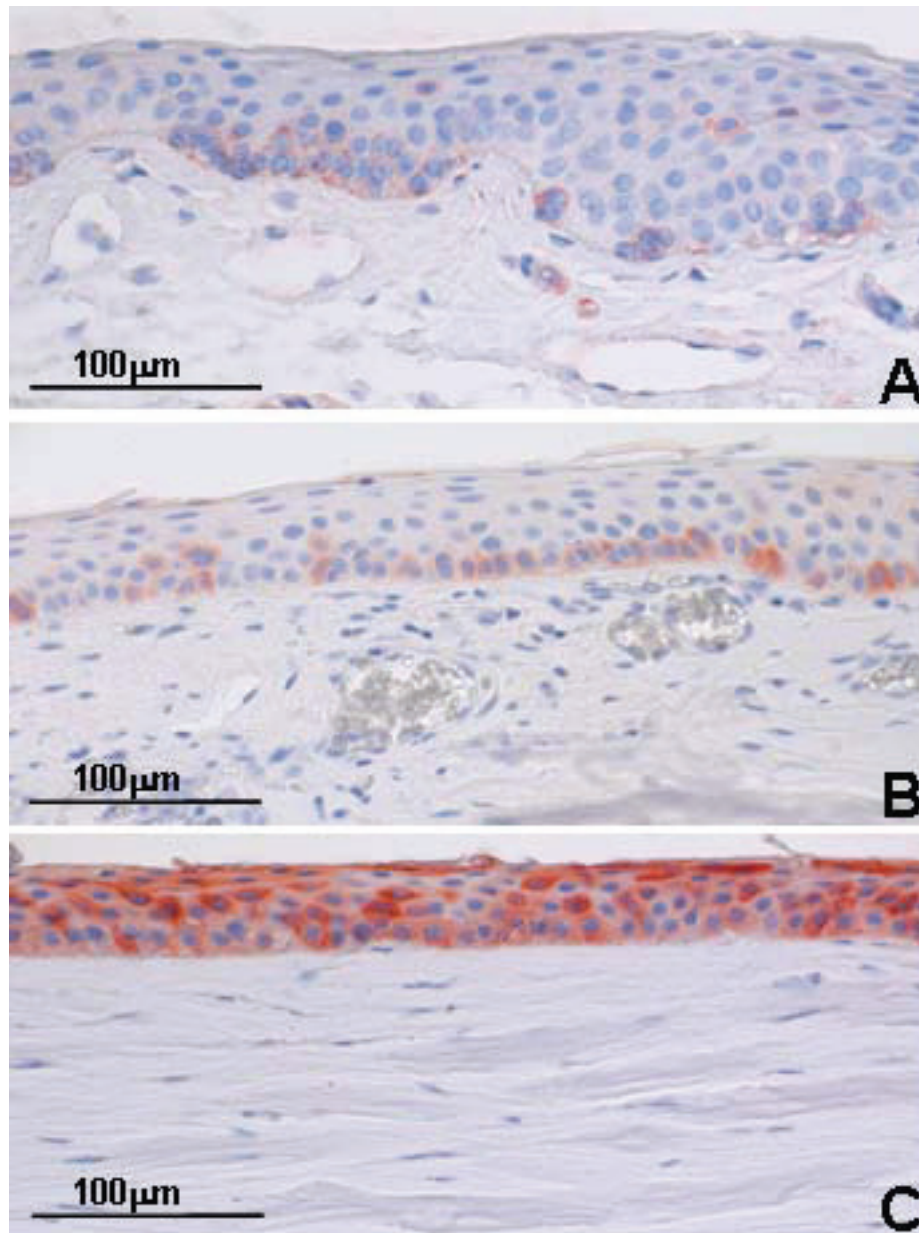
CK 5 and CK 14 are paired CK's and were present in 12% and 19% of the samples respectively. CK 5 and CK 14 have previously been found in the corneal epithelium [17,24], however, they are also present in basal keratinocytes in epidermis [2,19,23]. Therefore, most of the identified CK 5 and CK 14 are considered as contaminants, though a few of the identifications could be endogenous.

CK 6A was identified in 22% of the spots and was most likely a contaminant. CK 6A is expressed in palmoplantar epidermis, epidermal appendage and mucosa [2,19,23]. Again, none of the mass spectra aided in determining the origin of the contamination (data not shown).

CK's 1, 2e, 9 and 10 were identified in more than 73% of the human spots. Figure 3D shows the peptide-sequence leading to identification of CK 9 in a spot focused at around 10 kDa in the gel. Fragments of peptides were identified throughout the whole protein (62 kDa). It entails that the identification was based not only on a 10-kDa fragment, but rather on the whole protein. This can only be explained by contamination with an un-cleaved protein located on the surface of the excised spot or contamination introduced into the sample prior to enzymatic digestion. All 4 keratins are abundantly expressed in suprabasal cells of stratified and cornified epithelia or palmoplantar epidermis [2,19,23], but not in the corneal epithelium [4,17,25]. CK's 1, 2e, 9 and 10 are all considered to be contaminating. This is consistent with previous reports [26,27].

#### Aspects of contamination

Contamination of CK's occur principally during sample collection or gel preparation and excision. Indeed, by our scraping method there is a risk of contaminating the limbal epithelial fraction by conjunctival epithelium. Immunohistochemistry is especially effective in revealing this type of contamination [13]. If contamination occurs during sample collection and preparation prior to loading on the gels, the contaminating CK's may be focused as spots in the gels. If the contamination occurs during the preparation of the gel solutions, gel casting, spot excision or



**Figure 2** Immunohistochemical staining against various antigens. (A) CK 15 and (B) CK 19 appear in the limbal epithelium. (C) CK 3/CK 12 stainings in central corneal epithelium.

in the preparation of the spots prior to digestion, the contaminants may appear as identified proteins with molecular masses not in accordance with the spot position.

#### Impact of the number of investigated spots

The total number of investigated spots is important in this set-up. The more frequent a CK is identified the more likely it is to be a contaminant. However, it should also be considered that endogenous CK's that have several isoforms or have undergone degradation could

erroneously be interpreted as contamination and be excluded from the investigation.

#### Optimization of the method

In a proteomic screening experiment, it may be desirable to obtain as many protein candidates as possible if specific validation techniques are to be subsequently undertaken. An important issue is how much of the sample should be loaded onto the gel. Overloaded gels may lead to increased spot size that can merge with

<b>A)</b>					
1	MTSYSYR <b>QSS</b>	<b>ATSSFGGLGG</b>	<b>GSVR</b> FGPGVA	<b>FRAPSIHGGG</b>	<b>GGR</b> GVSVSSA
51	RFVSSSSSGG	YGGYGGVLT	ASDGLLAGNE	KLTMQNLNDR	<b>LASYLDKVRA</b>
101	<b>LEAANGELEV</b>	<b>KIRDWYQKQG</b>	PGPSRDYSHY	YTTIQDLRDK	ILGATIENSR
151	IVLQIDNARL	AADFRTKFE	TEQALRMSVE	ADINGLRR <b>VL</b>	<b>DELTLAR</b> TDL
201	EMQIEGLKEE	LAYLKNHEE	EISTLR <b>QVG</b>	<b>GQVSVEVDSA</b>	<b>PGTDLAK</b> ILS
251	DMRSQYEVMA	EQNRKDAEAW	FTSRTEELNR	EVAGHTEQLQ	MSRSEVTDLR
301	RTLQGLEIEL	QSQLSM <b>KAA</b> L	<b>EDTLAETEAR</b>	FGAQLAHTQA	LISGTEAQLG
351	DVRADSERQN	QEYQRLMDIK	SRLEQEIATY	RSLLEGQEDH	YNNLSASKVL
401					
<b>B)</b>					
1	MSRQASKTSG	GGSQGFSGRS	AVVSGSSRMS	CVAHSGGAGG	GAYGFRSAGG
51	GFGSRSLYNL	GGNKSISISV	AAGGSRAGGF	GGGRSSCAFA	GGYGGGFGSG
101	YGGGFGGGFG	GGRGMGGGFG	GAGGFGGAGG	FGGAGGFGGP	GGFGGSGGFG
151	GPGSLGSPGG	FGPGFPGGI	QEVITINQSL	QPLNVEIDPQ	IGQVKAQERE
201	QIKTLNNKFA	SFIDKVRFL	QQNKVLETKW	NLLQQQGTSS	ISGTNLEPL
251	FENHINYLR	YLDNILGERG	RLDSELKNME	DLVEDFKKKY	EDEINKRTAA
301	ENEFVTLKDK	VDSAYMNKVE	LQAKVDALID	EIDFLRLTYD	AELSQMOSHI
351	SDTSVVLSDM	NRSLDLDSI	IAEVRAQYED	IAQRSKAEAE	ALYQTK <b>L</b> GEL
401	<b>QTTAGR</b> HGDD	LRNTKSEIIE	LNRMIQRLRA	EIEGVKKQNA	NLQTAIAEAE
451	QHGENALKDA	NAKLQELQAA	LQQAKDDLAR	LLRDYQELMN	VKLALDVEIA
501	TYRKLLEGEE	YSRMSGECPS	AVSISVVSSS	TTSASAGGYG	GGYGGGMMGG
551	LGGGFSAGGG	SGSGFGRGGG	GGIGGGFGGG	SSGFSGGSGF	GSISGAR <b>Y</b> GV
601	<b>SGGGFSSASN</b>	<b>R</b> GGSIKFSQS	<b>S</b> QSSQRYSR		
<b>C)</b>					
1	MIARQQCVRG	GPRGFSCGSA	IVGGGKRGAF	SSVSMGGAG	RCSSGGFGSR
51	SLYNLRGNKS	ISMSVAGSRQ	GACFGGAGGF	GTGGFGAGGF	GAGFGTGGFG
101	GGFGGFSFGK	GGPGFPVCPA	GGIQEVTINQ	SLLTPLHVEI	DPEIQKVRTE
151	EREQIKLLNN	KFASFIDKVQ	FLEQQNKVLE	TKWNLLQQQT	TTTSSKNLEP
201	LFETYLSVLR	KQLDTLGNDK	GRLQSELKTM	QDSVEDFKTK	YEEETNKRTA
251	AENDFVVLKK	DVDAAYLNKV	ELEAKVDSL	DEINFLKVLV	DAELSQMOTH
301	VSDTSVVLSDM	DNNRNLDDLS	IIEVRAQYED	EIAQR <b>SKAEA</b>	<b>EALYQTKV</b> QQ
351	LQISVDQHG	NLRNTKSEIA	ELNRMIQRLR	AEIENIKKQC	QTLQVSVADA
401	EQRGENALKD	AHSCRVELEA	ALQQAKEELA	RMLREYQELM	SVKLALDIEI
451	ATYR <b>KLLEGE</b>	<b>EYR</b> MSGECQS	AVSISVVSGS	TSTGGISGGL	GGSGGFGLS
501	GFSGSGSGGF	FGGGSVSGSS	SSKIISTTTL	NKRR	
<b>D)</b>					
1	MSCRQFSSSY	LSR <b>SGGGGGG</b>	<b>GLGGGGSIRS</b>	SYSRFSSSGG	GGGGGR <b>FSSS</b>
51	<b>SGYGGGSSRV</b>	CGRGGGGSFG	YSYGGGSGGG	FSASSLGGGF	GGGSRGFGGA
101	SGGGYSSSGG	FGGGFGGSGG	GGFGGGYSGG	FGGFGGFGGG	AGGGDGGILT
151	ANEKSTMQEL	NSRLASYLDK	VQALEEANN	LENKIQDWYD	KKGPAAIQKN
201	YSPYYNTIDD	LKDQIVDLTV	GNNK <b>TL</b> LDID	<b>NTR</b> MTLDDFR	IKFEMEQLNR
251	<b>QGV</b> DADINGL	<b>R</b> QVLDNLTME	KSDLEMQYET	LQEELMALKK	NHKEEMSQLT
301	GQNSGDVNVE	INVAPGKDLT	KTLNDRMQEY	EQLIAKNRDK	IENQYETQIT
351	QIEHEVSSSQ	QEVQSSAKEV	TQLRHGVQEL	EIELQSLSK	KAALEKSLED
401	TKNRYCGQLQ	MIQEQISNLE	AQITDVRQEI	ECQNQEYSLL	LSIKMRLEKE
451	IETYHNLLEG	GQEDFESSGA	GKIGLGGG	SGGSYGRGSR	<b>GGSGGSYGGG</b>
501	<b>SGGGYGGGS</b>	<b>GSR</b> GGSGGSY	GGGSGSGGGS	GGGYGGGSGG	GHSGGSGGGH
551	SGGSGGNYGG	SGSGGSGSGG	GYGGGSGSRG	SGGSHGGGS	FGGESGGSY
601	GGGEEASGSG	GGYGGGSGKS	SHS		

**Figure 3 Peptide sequences from the Mascot search results.** **A:** CK 19 identified in a 39-kDa spot by 7 peptides (Matched peptides shown in bold red). It is most likely an endogenous protein although a contaminating protein cannot be excluded. **B:** Two peptides (24 amino acids) identify CK 3 from a spot focused around 28 kDa. This could be an endogenous protein or it could be due to contamination of the samples. **C:** Two peptides (20 amino acids) identify CK 4 from a spot focused around 60 kDa. This could be an endogenous protein or it could be a contamination of the samples. **D:** CK 9 identified by the matched peptides throughout the whole protein. Since the sample originates from a spot focused around 10 kDa in the gel the protein is regarded as a contaminant.

neighbouring spots making further analysis potentially difficult and inaccurate. Conversely, if an inadequate amount of sample is loaded onto the gel, the relative amount of contaminating proteins will tend to increase. In the present study, a relatively small amount of protein was loaded, in an effort to preserve the resolution of proteins in the 40-60 kDa range where gel spots have a tendency to merge together. However, despite the gels produced well-focused spots with low background

staining the number of successful identifications was lower than expected, probably as a result of the relatively small amount of protein loaded and the relatively high frequency of contaminating CK's.

### Conclusions

It is very important to consider possible contamination of CK's when undertaking MS based proteomic screening experiments, especially with regard to studies of the

cornea. Using proteomics as a screening tool in the search for differences in protein expression, including CK's, careful consideration should be applied to the sample amount chosen. Generally one should load as much sample as possible, to the point of not compromising 2D-gel spot resolution. We have shown that if several 2D-gel spots are being processed, CK's represented in less than 5% percent of the spots are very likely to be endogenous identifications (CK's 3, 4, 7, 8, 11, 12, 13, 15, 17, 18, 19, 20 and 23). If a CK is identified in a large part of the samples it is an uncertain endogenous CK (CK's 1, 2e, 5, 6A, 9, 10, 14 and 16). Confirmation of the CK identification by other biochemical methods such as immunohistochemistry is also very important.

#### Acknowledgements

We are grateful to Nakul Mandal MRCOphth, for reading and commenting on the manuscript. We would like to thank Inge Kjærgaard for excellent technical assistance. This work was supported by The Danish Medical Research Council, Aarhus University Research Foundation, The John and Birthe Meyer Foundation, The Velux Foundation, The Synoptic Foundation and The Beckett Foundation.

#### Author details

<sup>1</sup>Department of Ophthalmology, Aarhus University Hospital, Nørrebrogade 44, 8000 Aarhus C, Denmark. <sup>2</sup>Department of Medical Biochemistry, Aarhus University, Ole Worms Allé 3, Bldg. 1170, 8000 Aarhus C, Denmark. <sup>3</sup>Department of Ophthalmology, Aalborg Hospital, Aarhus University Hospital, Hobrovej 18-22, 9100 Aalborg, Denmark.

#### Authors' contributions

ML carried out the 2D PAGE analysis and the immunohistochemistry and contributed substantially to the writing. HV carried out the 2D PAGE and contributed to the writing. KN contributed substantially to the immunohistochemistry and the writing. NE participated in its design and coordination and helped to draft the manuscript. BH carried out the mass spectrometry and contributed substantially to the writing. All authors read and approved the final manuscript.

#### Competing interests

The authors declare that they have no competing interests.

Received: 13 April 2010 Accepted: 27 January 2011

Published: 27 January 2011

#### References

1. Linder S: Cytokeratin markers come of age. *Tumour Biol* 2007, **28**:189-195.
2. Irvine AD, McLean WH: Human keratin diseases: the increasing spectrum of disease and subtlety of the phenotype-genotype correlation. *Br J Dermatol* 1999, **140**:815-828.
3. Uusitalo M, Kivela T: Development of cytoskeleton in neuroectodermally derived epithelial and muscle cells of the human eye. *Invest Ophthalmol Vis Sci* 1995, **36**:2584-2591.
4. Kivela T, Uusitalo M: Structure, development and function of cytoskeletal elements in non-neuronal cells of the human eye. *Prog Retin Eye Res* 1998, **17**:385-428.
5. Kasper M, Moll R, Stosiek P, Karsten U: Patterns of cytokeratin and vimentin expression in the human eye. *Histochemistry* 1988, **89**:369-377.
6. Kasper M, Viebahn C: Cytokeratin expression and early lens development. *Anat Embryol (Berl)* 1992, **186**:285-290.
7. Figueira EC, Di Girolamo N, Coroneo MT, Wakefield D: The phenotype of limbal epithelial stem cells. *Invest Ophthalmol Vis Sci* 2007, **48**:144-156.
8. Honoré B, Vorum H, Pedersen AE, Buus S, Claesson MH: Changes in protein expression in p53 deleted spontaneous thymic lymphomas. *Exp Cell Res* 2004, **295**:91-101.

9. Ochs D: Protein contaminants of sodium dodecyl sulfate-polyacrylamide gels. *Anal Biochem* 1983, **135**:470-474.
10. Mann M, Hendrickson RC, Pandey A: Analysis of proteins and proteomes by mass spectrometry. *Annu Rev Biochem* 2001, **70**:437-473.
11. Paul-Pletzer K, Parness J: Elimination of keratin contaminant from 2-mercaptoethanol. *Anal Biochem* 2001, **289**:98-99.
12. Lubec G, Afjehi-Sadat L: Limitations and pitfalls in protein identification by mass spectrometry. *Chem Rev* 2007, **107**:3568-3584.
13. Lyngholm M, Vorum H, Nielsen K, Ostergaard M, Honore B, Ehlers N: Differences in the protein expression in limbal versus central human corneal epithelium—a search for stem cell markers. *Exp Eye Res* 2008, **87**:96-105.
14. Schlotzer-Schrehardt U, Kruse FE: Identification and characterization of limbal stem cells. *Exp Eye Res* 2005, **81**:247-264.
15. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS: Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999, **20**:3551-3567.
16. Honore B, Buus S, Claesson MH: Identification of differentially expressed proteins in spontaneous thymic lymphomas from knockout mice with deletion of p53. *Proteome Sci* 2008, **6**:18.
17. Elder MJ, Hiscott P, Dart JK: Intermediate filament expression by normal and diseased human corneal epithelium. *Hum Pathol* 1997, **28**:1348-1354.
18. Whitbread LA, Powell BC: Expression of the intermediate filament keratin gene, K15, in the basal cell layers of epithelia and the hair follicle. *Exp Cell Res* 1998, **244**:448-459.
19. Moll R, Franke WW, Schiller DL, Geiger B, Krepler R: The catalog of human cytokeratins: patterns of expression in normal epithelia, tumors and cultured cells. *Cell* 1982, **31**:11-24.
20. Cockerham GC, Laver NV, Hidayat AA, McCoy DL: An immunohistochemical analysis and comparison of posterior polymorphous dystrophy with congenital hereditary endothelial dystrophy. *Cornea* 2002, **21**:787-791.
21. Merjava S, Neuwirth A, Mandys V, Jirsova K: Cytokeratins 8 and 18 in adult human corneal endothelium. *Exp Eye Res* 2009, **89**:426-431.
22. Zhang JS, Wang L, Huang H, Nelson M, Smith DI: Keratin 23 (K23), a novel acidic keratin, is highly induced by histone deacetylase inhibitors during differentiation of pancreatic cancer cells. *Genes Chromosomes Cancer* 2001, **30**:123-135.
23. Kurokawa I, Mizutani H, Kusumoto K, Nishijima S, Tsujita-Kyutoku M, Shikata N, Tsubura A: Cytokeratin, filaggrin, and p63 expression in reepithelialization during human cutaneous wound healing. *Wound Repair Regen* 2006, **14**:38-45.
24. Endo K, Kawasaki S, Nakamura T, Kinoshita S: The presence of keratin 5 as an IgG Fc binding protein in human corneal epithelium. *Exp Eye Res* 2004, **78**:1137-1141.
25. Pitz S, Moll R: Intermediate-filament expression in ocular tissue. *Prog Retin Eye Res* 2002, **21**:241-262.
26. Parker KC, Garrels JL, Hines W, Butler EM, McKee AH, Patterson D, Martin S: Identification of yeast proteins from two-dimensional gels: working out spot cross-contamination. *Electrophoresis* 1998, **19**:1920-1932.
27. Ding Q, Xiao L, Xiong S, Jia Y, Que H, Guo Y, Liu S: Unmatched masses in peptide mass fingerprints caused by cross-contamination: an updated statistical result. *Proteomics* 2003, **3**:1313-1317.

#### Pre-publication history

The pre-publication history for this paper can be accessed here:  
<http://www.biomedcentral.com/1471-2415/11/3/prepub>

doi:10.1186/1471-2415-11-3

Cite this article as: Lyngholm et al.: Attempting to distinguish between endogenous and contaminating cytokeratins in a corneal proteomic study. *BMC Ophthalmology* 2011 **11**:3.