

Pulmonary adenocarcinoma with mucin production modulates phenotype according to common genetic traits: a reappraisal of mucinous adenocarcinoma and colloid adenocarcinoma

Angelica Sonzogni,¹ Fabrizio Bianchi,² Alessandra Fabbri,¹ Mara Cossa,¹ Giulio Rossi,³ Alberto Cavazza,⁴ Elena Tamborini,¹ Federica Perrone,¹ Adele Busico,¹ Iolanda Capone,¹ Benedetta Picciani,¹ Barbara Valeri,¹ Ugo Pastorino⁵ and Giuseppe Pelosi^{6,7*}

¹ Department of Pathology and Laboratory Medicine, Fondazione IRCCS Istituto Nazionale Tumori, Milan, Italy

² Institute for Stem-Cell Biology, Regenerative Medicine and Innovative Therapies (ISBreMIT), IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy

³ Division of Anatomic Pathology, Regional Hospital Umberto Parini, Aosta, Italy

⁴ Department of Oncology and Advanced Technology, Operative Unit of Pathologic Anatomy, IRCCS Azienda Arcispedale S. Maria Nuova, Reggio Emilia, Italy

⁵ Division of Thoracic Surgery, Fondazione IRCCS Istituto Nazionale Tumori, Milan, Italy

⁶ Department of Oncology and Hemato-Oncology, Università degli Studi, Milan, Italy

⁷ Inter-Hospital Pathology Division, Science & Technology Park, IRCCS MultiMedica Group, Milan, Italy

*Correspondence to: Giuseppe Pelosi, Servizio Interaziendale di Anatomia Patologica, Polo Scientifico e Tecnologico, Via Gaudenzio Fantoli 16/15, 20138 Milan, Italy. E-mail: giuseppe.pelosi@unimi.it

Abstract

Whether invasive mucinous adenocarcinoma (IMA) and colloid adenocarcinoma (ICA) of the lung represent separate tumour entities, or simply lie within a spectrum of phenotypic variability, is worth investigating. Fifteen ICA, 12 IMA, 9 *ALK*-rearranged adenocarcinomas (ALKA), 8 non-mucinous *KRAS*-mutated adenocarcinomas (KRASA) and 9 mucinous breast adenocarcinomas (MBA) were assessed by immunohistochemistry for alveolar (TTF1, cytoplasmic MUC1), intestinal (CDX-2, MUC2), gastric (membrane MUC1, MUC6), bronchial (MUC5AC), mesenchymal (vimentin), neuroendocrine (chromogranin A, synaptophysin), sex steroid hormone-related (oestrogen and progesterone receptors), pan-mucinous (HNF4A) and pan-epithelial (keratin 7) lineage biomarkers and by targeted next generation sequencing (TNGS) for 50 recurrently altered cancer genes. Unsupervised clustering analysis using molecular features identified cluster 1 (IMA and ICA), cluster 2 (ALKA and KRASA) and cluster 3 (MBA) ($p < 0.0001$). Cluster 1 showed four histology-independent sub-clusters (S1 to S4) pooled by HNF4A and MUC5AC but diversely reacting for TTF1, MUC1, MUC2, MUC6 and CDX2. Sub-cluster S1 predominantly featured intestinal-alveolar, S2 gastrointestinal, S3 gastric and S4 alveolar differentiation. In turn, KRASA and ALKA shared alveolar lineage alongside residual MUC5AC expression, with additional focal CDX2 and diffuse vimentin, respectively. A proximal-to-distal scheme extending from terminal (TB) and respiratory (RB) bronchioles to alveolar cells was devised, where S3 originated from distal TB (cellular mucinous adenocarcinoma), S2 from proximal RB (secreting mucinous adenocarcinoma), S1 from intermediate RB (mucin lake-forming colloid adenocarcinoma), S4 from distal RB (colloid alveolar adenocarcinoma), KRASA from juxta-alveolar RB (KRAS-mutated non-mucinous adenocarcinoma) and ALKA from juxta-bronchial alveolar cells (ALK-translocated adenocarcinoma). TNGS analysis showed *KRAS*, *LKB1*, *TP53*, *APC* and *CDKN2A* mutation predominance. In conclusion, IMA and ICA are basket categories, which likely originate from distinct domains of stem/progenitor cells spatially distributed along bronchioles upon common molecular features and genetic alterations.

Keywords: adenocarcinoma; lung; mucinous; colloid; immunohistochemistry; next generation sequencing; reappraisal; cluster analysis

Received 17 November 2016; Accepted 20 February 2017

No conflicts of interest were declared.

Introduction

Among mucin-laden adenocarcinomas (MLA) of the lung, the 2015 WHO classification identifies mucinous adenocarcinoma *in situ*, minimally invasive mucinous adenocarcinoma, invasive mucinous adenocarcinoma (IMA) and invasive colloid adenocarcinoma (ICA) [1]. Invasive enteric adenocarcinoma (IEA) resembling the homologous intestinal tumours has been included as a variant [1]. Other MLA types are solid adenocarcinoma with mucin [1] and signet ring cell adenocarcinoma (SRCA) [2], although SRCA has been downgraded to mere cytological change of any pattern's adenocarcinoma [1]. Lastly, there are non-mucinous adenocarcinomas accumulating extracellular mucin, which have been reappraised as histological patterns of adenocarcinoma with mucin extravasation (HPAME) [3]. IMA, ICA and IEA share many phenotypic traits, which are opposed to conventional non-mucinous adenocarcinomas and HPAME. Expression of NK2 homeobox 1/thyroid-specific transcription factor-1 (NKX2-1/TTF-1) [4–7], napsin A aspartic peptidase (NAPSA) [4,8,9], surfactant protein A1 (SFTPA1) [5,8], caudal type homeobox 2 (CDX2) [5,6,8,10,11], keratin 20 (KRT20) [5,6,8,12], mucin 1 cell surface associated (MUC1) [13], mucin 2 oligomeric mucus/gel-forming (MUC2) [5,11,13], mucin 5AC oligomeric mucus/gel-forming (MUC5AC) [5,10,13], mucin 6 oligomeric mucus/gel-forming (MUC6) [10,13,14] and hepatocyte nuclear factor 4 alpha (HFN4A) [15], have been variably reported onto emphasize the inherent heterogeneity of these tumours. Furthermore, IMA may even express neuroendocrine (NE) markers [16] and sex steroid hormone receptors [17,18]. The molecular landscape of IMA, ICA, IEA and, to lesser extent, HPAME is dominated by *KRAS* [3,6,10,19,20], *TP53* and *LKB1/STK11* [21,22] over Epidermal Growth Factor Receptor (EGFR) mutations [3,6,19], while the *CD74-NRG1* fusion genes has been documented in 13–27% *KRAS*-wild type tumours [23,24]. Most SRCA harbour *ALK* rearrangement [2,25].

Recent data have indicated that synchronous *KRAS* mutation and *Nkx2-1/TTF1* repression cause IMA in the lung [26], with napsin-A downregulation [10] and gastric over intestinal differentiation [10,27]. These *Nkx2-1/TTF1*-deleted IMA with *KRAS* mutation grew in well-differentiated multiple foci upon reduced apoptosis from mucinous precursors in the distal bronchiole epithelium [27], developing metastases only when loss-of-function *TP53* mutation occurred [28].

According to these assumptions, it is tempting to speculate that IMA, ICA and IEA lie with in a

spectrum of relatively uniform tumours on molecular grounds [6,19,20,23], which differ upon morphology and immunohistochemistry (IHC) traits [4–8,10]. They would therefore embody prototypical lesions resembling gastro-entero-pancreatic or gynaecological tumours [29–32], and even primary *ALK*-translocated adenocarcinoma (ALKA) when abundant extracellular mucin is accumulated [6,11,33,34]. Upon IHC and molecular investigation, we herein suggest that IMA, ICA and IEA make up a heterogeneous family of tumours encompassing four different subgroups, which likely arise from different cancer progenitor/stem cell niches in the distal airways.

Materials and methods

Study design

We designed and conducted a proof-of-concept study to challenge whether pulmonary IMA, ICA and IEA encompass separate tumour entities with distinct phenotypic profiles depending upon varying interplay of architectural complexity and lineage changes (cellular phase) or mucin production and extravasation (secretion phase), under the pooling of a common mucin-encoding gene activation programme, or simply made up a spectrum of histological heterogeneity.

Patients and tumours

A series of 44 consecutive resection specimens of mucin-laden lung adenocarcinoma patients from 25 males averaging 63.8 years (range 28–85 years) and 19 females averaging 61.4 years (range 38–80) were retrospectively identified in the pathology archives of the participant Institutions among diagnoses of lung adenocarcinoma, inasmuch as ICA, IMA and ALKA are less likely to be documented in surgical specimens for being either quite uncommon tumours or inoperable lesions in most instances at the time of the clinical diagnosis. All tumours fulfilling diagnostic criteria for mucin-laden adenocarcinoma made up different comparable groups in terms of confounding factors, such as age, gender and tumour stage as detailed in supplementary material A. They included 15 ICA (5 pure and 10 mixed with non-colloid component), 12 IMA (9 pure and 3 mixed with non-mucinous component), 9 ALKA and 8 non-mucinous *KRAS*-mutated adenocarcinomas (KRASA). The occurrence of *ALK* rearrangement and *KRAS* mutation had been verified by fluorescence *in situ* hybridization [35,36] and Sanger direct sequencing of exon 2 to 4 [37,38], respectively. As negative control

group of non-pulmonary and unrelated MLA, nine mucinous breast adenocarcinomas (MBA) entered the study, inasmuch as mucin-laden adenocarcinomas of the lung are prototypical lesions with similarities to homologous digestive and gynaecological mucinous tumours likely due to commonalities in the embryological derivation and pathogenesis.

All tumours were classified according to the criteria of WHO classifications on lung [1] and breast [39] cancer. There were 30 specimens from (bi)lobectomy, 12 from segmentectomy and 2 from pneumonectomy corresponding to 21 tumours staged IA, 5 staged IB, 2 staged IIA, 11 staged IIB, 3 staged IIIA and 1 staged IV (one patient was lost to pathological staging). Information on smoking habit was not available at the time of the present study and survival analysis was not a specific endpoint of the study. All surgical specimens had been fixed in 4% buffered formaldehyde solution for 12–24 hr and embedded in paraffin according to standard histopathological methods. All the original haematoxylin and eosin sections were jointly reviewed by four (GP, MC, AF, AS) of the authors for consistency, without knowledge of the patients' identities or original tumour categorization. Pleural invasion was classified as PL0, PL1, PL2 and PL3 according to updated criteria [40].

The group of IMA, ICA, ALKA and KRASA was comparatively studied for age, gender, type of surgery, tumour stage, tumour diameter, number of excised lymph nodes, percentage of metastatic lymph nodes, pT factor, pN factor, pleural invasion, architectural patterns and variants, and occurrence of signet ring, goblet or floating cells (each by 10% cutoff for presence vs. absence), extracellular mucin (cutoff 50% for minor vs. prevalent) and type of mucin (basophilic, amphophilic or both), as detailed in supplementary material A.

In particular, non-mucinous adenocarcinoma components were expressed as percentages of patterns of growth and variants in 5% numerical increments according to current classification guidelines [1]. No cases of adenocarcinoma *in situ* or minimally invasive adenocarcinoma were documented in this study.

Immunohistochemistry

The primary antibody list entering the study and technical information on IHC procedures is available in supplementary material B and supplementary material C, respectively. The antibodies included biomarkers of alveolar (TTF-1, MUC1), gastrointestinal (CDX-2, keratin 20, MUC2, MUC6), bronchial (MUC5AC), NE (chromogranin A, synaptophysin), mesenchymal (vimentin), sex steroid hormone

(oestrogen and progesterone receptors), pan-mucinous (HNF4A) and pan-epithelial (keratin 7) differentiation lineages, along with the ALK rearrangement-associated protein. Normal lung tissue was investigated for the same IHC biomarkers.

Molecular study

The group of IMA/ICA/IEA tumours underwent targeted-next generation sequencing (NGS). Normal lung tissue was used as negative controls as appropriate for consistency. Detailed descriptions of the molecular procedures pursued in this investigation, including DNA extraction, quantification and targeted next generation sequencing (TNGS) analysis, have been reported previously [35,37,38]. In brief, NGS analysis was carried out using small DNA samples (10 ng/ μ l) from tumour paraffin sections accounting for at least 80% cellularity, with the Ion AmpliSeq Cancer Hotspot Panel v.2 (Life Technologies-Thermo Fisher Scientific, Waltham, MA, USA) running on the Ion-Torrent™ Personal Genome Machine platform comprising 50 oncogenes and tumour suppressor genes recurrently mutated in human cancers (Life Technologies-Thermo Fisher Scientific), as previously detailed [37,38].

Ethics

The study was approved by the independent ethics committee of the National Tumour Institute IRCCS Foundation, Milan, Italy (accession number INT-188/15). All patients gave their written consent for diagnosis and research activities when they were admitted to the hospital.

Statistics

Statistical analyses were performed using JMP IN software (SAS Institute, Inc., Cary, NC) and VassarStats for Statistical Computation at <http://vassarstats.net> website. Qualitative or quantitative data were compared by Fisher exact *t* test, chi-squared test, Mann–Whitney test and Kruskal–Wallis test as appropriate. Unsupervised and supervised hierarchical clustering analyses were performed using Cluster 3.0 software (<http://www.eisenlab.org/eisen/>) and visualized using Java TreeView (<http://jtreeview.sourceforge.net>). The defining features of the diverse clusters corresponded to the different IHC percentages used for the relevant markers. Similarity was measured using Spearman rank correlation metric and clustering was performed by centroid linkage. Contingency analysis and relative mosaic plots were performed by means of JMP IN software and

calculating *P* values through the likelihood ratio test. Mosaic plots indicated graphically the fraction (expressed from 0.0 to 1.0) of patients belonging to the different clusters after stratifying for the different tumour subtypes. For all tests, two-sided *P* values were taken into account, with a threshold <0.05 as being statistically significant.

Results

Analysis by histology

We initially searched for either morphology or relevant IHC findings, which stratified the cohort of 44 lung adenocarcinomas. In KRASA and IMA, we found only predominant acinar-cribriform or solid with mucin and lepidic/acinar/enteric patterns, respectively, and basophilic mucin in IMA, while in ICA, ALKA and KRASA we found the presence of extracellular mucin and SRC and the lack of goblet and freely floating cells, respectively (supplementary material A). The IHC analysis revealed that the alveolar biomarkers prevailed in mixed IMA/ICA, ALKA and KRASA, whereas gastrointestinal biomarkers were preferentially expressed in IMA and ICA alongside MUC5AC and HNF4A (supplementary material D; supplementary material, Figure S1A–F). Vimentin and ALK protein were limited to ALKA, whereas all tumours exhibited diffuse labelling for keratin 7 and at least focal NE biomarkers (supplementary material D; supplementary material, Figure S1A–F). Only the staining for sex steroid hormone receptors was erratic. For secreted mucin, MUC1 was only noted in 12/15 (80%) ICA, 5/12 (42%) IMA, 7/9 (78%) ALKA and 1/8 (13%) KRASA (supplementary material D). Further details of IHC findings, including normal lung tissue, are reported on supplementary material C, with representative pictures in supplementary material, Figure S1A–F.

Analysis by clustering

We then used hierarchical clustering analysis to perform a molecular classification of this cohort of 53 MLA, using a set of known IHC biomarkers. Such analysis revealed three main clusters and a non-random distribution of these IHC biomarkers across the 53 MLA under evaluation. Upon histology, *cluster 1* included a mixture of IMA and ICA and, unexpectedly, even one KRASA case; *cluster 2* comprised all ALKA and the remaining KRASA cases; *cluster 3* was limited to MBA ($p < 0.0001$) (Figure 1A and 1B). Upon IHC, gastrointestinal (CDX-2, keratin 20,

membrane MUC1, MUC6), bronchial (MUC5AC) and pan-mucinous (HNF4A) biomarkers prevailed in *cluster 1*, whilst alveolar (TTF1, cytoplasmic MUC1), ALKA-related (ALK protein, vimentin) and sex steroid hormone along with MUC2 were dominant in *clusters 2* and *3*, respectively. The remaining biomarkers were not differentially distributed, even though NE biomarkers prevailed in the *cluster 3* (Table 1).

Analysis by sub-clustering

Next, we further characterized the composition of *cluster 1*, which revealed the presence of four distinct *sub-clusters*, named *S1* to *S4* (Figure 2). Upon histology, a mixture of colloid and mucinous/non-mucinous patterns was seen in *S4*, mucinous pattern in *S2* and *S3* and colloid variant in *S1*. Moreover, floating cells with amphophilic extracellular mucin clustered in *S1* only (supplementary material E; supplementary material, Figure S2). Upon IHC, differential distribution was found among *sub-clusters* for TTF1, MUC1, CDX-2, MUC2, MUC6 and HNF4A but no other biomarkers, which were thus excluded from further analysis (supplementary material F). Of note, MUC1 labelled cell membranes in *sub-cluster 3*, while was confined to cytoplasm in *sub-clusters S1, S2* and *S4*, as in the normal lung tissue (Table 2; supplementary material F).

The six biomarkers were then used for re-classifying IMA and ICA, regardless of histology. According to the distribution of alveolar and gastrointestinal biomarkers, *S3* and *S4 sub-clusters* were placed at both ends and *S1* and *S2 sub-clusters* in the middle of a physio-anatomical unit extending from bronchioles up to alveolar cells (supplementary material, Figure S3). Briefly, *sub-cluster S3* featured exclusive gastric differentiation (membrane MUC1, MUC6), *S2* admixed gastric and intestinal lineage, *S1* admixed intestinal and alveolar biomarkers and *S4* escalation of alveolar traits with striking downregulation of intestinal biomarkers (Table 2). Pan-mucinous HNF4A tracer significantly decreased from *S3/S2/S1* to *S4 sub-cluster*, whereas MUC1-reactive extracellular mucin accumulation increased from *S3/S2 sub-clusters* (4/11 cases, 36.4%) to *S1/S4 sub-clusters* (14/17 cases, 82.4%) (Table 2). HNF4A correlated inversely with TTF1 ($r = -0.839$, $p < 0.0001$) and MUC1 ($r = -0.504$, $p = 0.006$) and directly with CDX-2 ($r = 0.462$, $p = 0.013$) and MUC6 ($r = 0.523$, $p = 0.004$), whilst HNF4A exceeded MUC5AC median values (90.5% vs. 72%, $p = 0.026$). In turn, CDX-2 associated directly to MUC2 expression ($r = 0.876$, $p < 0.001$) and inversely with TTF1

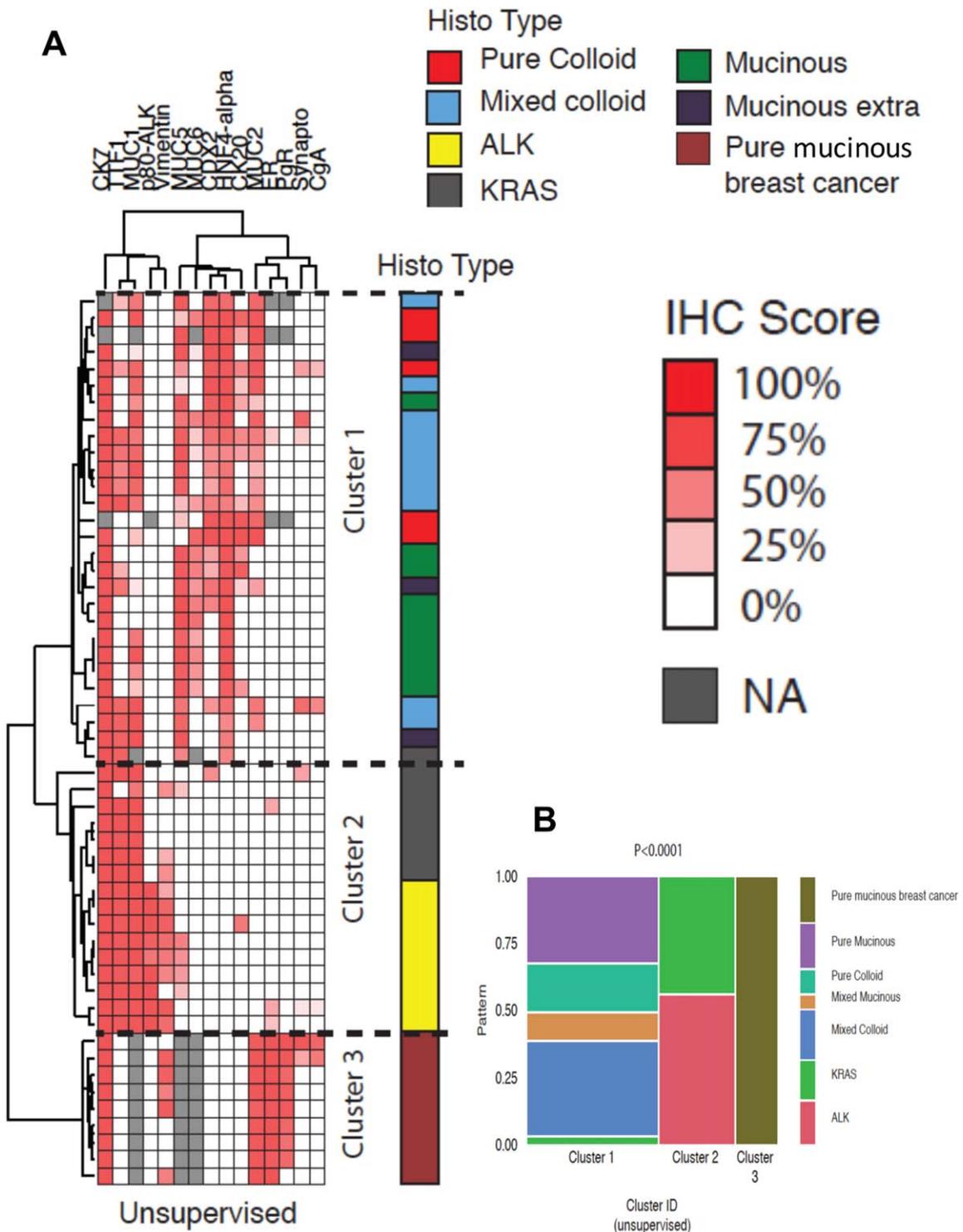


Figure 1. Unsupervised hierarchical clustering analysis of the IHC biomarkers used in this study. (A) Three distinct clusters were identified corresponding to the main tree branches: *cluster 1* which includes all mucinous and colloid adenocarcinomas with intermingling histological subtypes, either pure or mixed; *cluster 2* which comprises all *ALK*-translocated and all but one *KRAS*-mutated adenocarcinoma; and *cluster 3* which groups all MBA. (B) A mosaic plot (the greater the surface of the different coloured rectangles, the greater the number of cases in the different tumour subsets) confirms the significance ($p < 0.0001$) of the differential distribution of tumours according to phenotype, and the inherent heterogeneity of tumour composition, especially in *cluster 1*. NA, IHC score not available.

Table 1. Immunohistochemistry data according to unsupervised clustering 1 to 3

Type of marker	Variable	Cluster 1	Cluster 2	Cluster 3	P value
Alveolar marker	TTF1	20.5 ± 28.6	88.1 ± 23.9	0	<0.0001
	MUC1	39.7 ± 36.4	93.6 ± 15.9	n.a.	<0.0001
Gastro-intestinal marker	CDX2	37.8 ± 37.3	1.6 ± 5.0	0	<0.0001
	CK20	16.1 ± 28.5	2.3 ± 9.0	0	0.002
	MUC2	30.5 ± 34.8	0	90.0 ± 7.5	<0.0001
Bronchial marker	MUC6	13.5 ± 16.5	0	n.a.	0.0001
	MUC5AC	60.8 ± 34.1	4.4 ± 8.4	n.a.	<0.0001
Neuroendocrine marker	Synaptophysin	4.5 ± 14.3	0.8 ± 2.8	11.7 ± 31.4	0.645
	Chromogranin A	1.1 ± 14.5	0.2 ± 0.5	10.6 ± 21.9	0.223
Sex steroid hormone marker	ER	0.2 ± 0.8	1.6 ± 3.5	97.7 ± 6.6	<0.0001
	PgR	0	0	64.4 ± 30.9	<0.0001
ALK-related marker	ALK	0	50.6 ± 48.6	0	<0.0001
	Vimentin	0	27.4 ± 28.8	31.7 ± 34.8	<0.0001
Pan-mucinous marker	HNF4- α	74.5 ± 29.3	0	0	<0.0001
Pan-epithelial marker	CK7	99.0 ± 4.1	100.0 ± 0.0	99.4 ± 1.7	0.458

n.a., not available.
P values in bold are considered statistically significant.

($r = -0.368$, $p = 0.054$). Regarding tumour staging, there was a tendency for *S1* and *S4* sub-clusters to show pT1a early tumours (9/17 cases: 53%) as

opposite to *S1* and *S4* sub-clusters (2/11 cases: 18%) ($p = 0.115$). Representative pictures of sub-clusters are shown in supplementary material, Figures S4–S7.

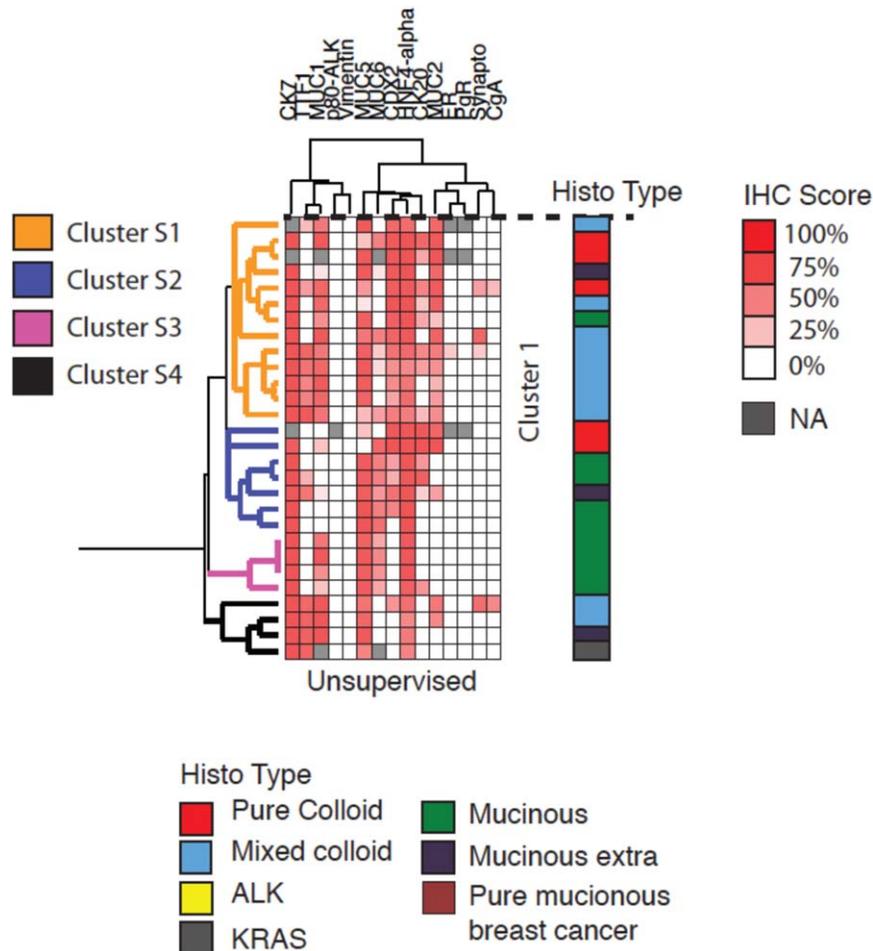


Figure 2. Detailed unsupervised clustering analysis of cluster 1 tumours shows four distinct and separate sub-clusters (highlighted in different colours), namely *S1*, *S2*, *S3* and *S4*, according to the IHC results. NA, IHC score not available.

Table 2. Histogenetic analysis by immunohistochemistry and stratification by sub-clusters S1 to S4

Type of marker	Variable	S3 (n = 4)	P value S3 vs. S2	S2 (n = 7)	P value S2 vs. S1	S1 (n = 13)	P value S1 vs. S4	S4 (n = 4)	P value among all
Alveolar marker	TTF1	0	0.327	6.8 ± 14.9	0.259	20.6 ± 28.7	0.021	64.5 ± 32.5	0.006
	MUC1 c	0	n.a.	0	0.0001	53.1 ± 27.5 c/m	0.052	97.5 ± 39.6 c	0.0002
Intestinal marker	CDX2	0	0.023	40.0 ± 40.1	0.284	58.6 ± 31.2	0.004	4.2 ± 8.3	0.002
	MUC2	0	0.186	24.7 ± 39.8	0.073	47.8 ± 33.4	0.186	14.9 ± 17.9	0.012
Gastric marker	MUC1 m	38.2 ± 37.0	0.009	1.0 ± 1.9	n.a.	0	n.a.	0	0.0004
	MUC6	15.3 ± 6.7	0.154	3.0 ± 19.3	0.013	6.6 ± 11.1	0.042	0	0.009
Pan-mucinous marker	HFN4-alpha	88.7 ± 11.8	0.214	96.8 ± 3.4	0.152	71.2 ± 30.2	0.030	32.3 ± 10.4	0.018
	MUC1 positive	2	0.576	2	0.062	10	0.541	4	0.054
MUC IHC in mucin	MUC1 positive	2		2		10		4	
	MUC1 negative	2		5		3		0	

m, membrane labelling; c, cytoplasmic labelling; c/m, cytoplasmic and membranous; n.a., not applicable. P values in bold are considered statistically significant.

For the sake of comparison, pictures from KRASA and ALKA morphology and IHC findings are also provided (supplementary material, Figures S8 and S9).

Analysis by TNGS

Finally, we performed next-generation sequencing of the same cohort of tumours to further look at the genetic alterations underpinning the different histological subtypes and molecular sub-clusters. There were 22 mutated out of 28 analysed tumours (78%) totalling 32 mutations (1.45 mutation/tumour patient). The most frequent mutation turned out KRAS (16 cases) followed by LKB1 (4 cases), TP53 (4 cases), APC (2 cases), CDKN2A (2 cases) and HFN1A, c-KIT, SMAD4 and SMO in 1 case each, regardless of histology (supplementary material, Figure S10).

There were no preferential types of KRAS mutation. The rate for any mutation was 50% (2/4) in S3, 57% (4/7) in S2, 54% (7/13) in S1, 100% (4/4) in S4 and 100% (7/7) in KRASA (p = 0.021) (supplementary material F). No differences emerged between trans-versions (17/32 mutations) and transitions (15/32 mutations), irrespective of genes. ALKA were not subjected to TNGS analysis.

Discussion

In our study, we noticed that the visceral endoderm HFN4A master gene in Nkx2-1/TTF1-downregulating and KRAS-mutated tumours [10,26,41] could be placed at the top of a developmental scheme (Figure 3), where HFN4A acted as inducer of mucinous differentiation via MUC5AC trans-activation [26,41,42]. MUC5AC was in turn significantly lower than HNF4A [41], while is normally limited to bronchiolar goblet cells. The lack of recognizable precursor lesions (bronchial columnar cell dysplasia [43] or mucinogenic growth clusters [44,45]), suggested that alternative non-linear or de novo mechanisms were likely to play some role in tumour development since the reported genotypic and phenotypic profile of these precursors [44,45] adapted to only a subset of lesions in our study. The subsequent tumour evolution depended on the relative activation state of Nkx2-1/TTF1, which basically causes mucin repression in terminal respiratory unit cells [26,46]. The inverse relationship of HNF4A with Nkx2-1/TTF1 and its direct correlation with CDX-2 [47] and MUC6 were likely to induce either mucinous gastric differentiation of antral gland/mucopeptic cell type

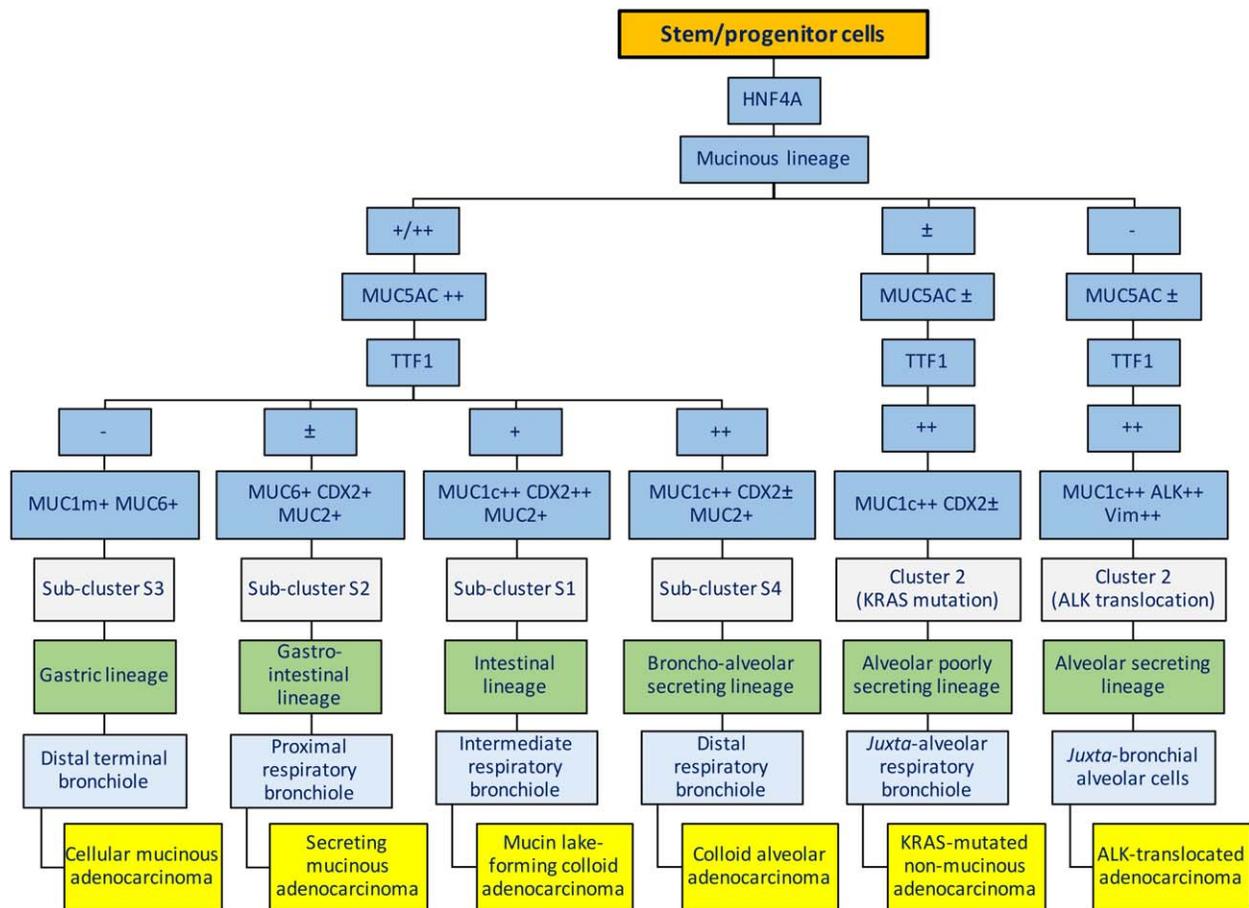


Figure 3. Operative flow-chart showing the development of lung adenocarcinomas with mucin production from different stem/reserve cell niches distributed along the terminal and respiratory bronchioles up to alveolar cells. A new terminology was then devised on the basis of differential phenotypic combinations of the *S1* → *S4* sub-clusters of *cluster 1* and *cluster 2* tumours. Profiles were considered negative if staining was completely absent in the relevant cells; focally (±) immunoreactive cases exhibited immunostaining in 1–10% neoplastic cells; + cases in 11–50% neoplastic cells; ++ cases in 50% or more neoplastic cells. The prefix 'juxta' stands for 'near to'.

[10,26,48] or secretory intestinal lineage [49]. CDX-2 in turn closely associated to MUC2 expression [50] and, marginally, was alternative to *Nkx2-1/TTF1*, in keeping with the finding of TTF1 and CDX-2 co-expression even at the level of individual tumour cells. Depolarized MUC1 decoration we documented in varying proportions has been linked to cell detachment [51], apoptosis impairment [52] and cancer cell proliferation [53].

This peculiar and non-random distribution of master regulators and activation products made us hypothesize that diversely regulated niches or domains of mucinous-committed stem/progenitor cancer cells existed in bronchioles, in keeping with lung ontogenesis, as suggested by the congeries of tumours in cluster 1. These tumours recapitulated

two alternative biological properties: increasing architectural complexity (cellular phase) and mucin escalation (secretion phase), both phenomena being under the control of common genetic traits. When the cellular phase predominated, multifocal tumours with more advanced stage developed, which featured antral gland/foveolar gastric (membrane MUC1, MUC6) and/or intestinal (CDX-2, MUC2) but nil or negligible alveolar (TTF1, cytoplasmic MUC1) cell lineage. When gastric differentiation prevailed upon complete TTF1 silencing (*S3* sub-cluster), there were multiple microscopic neoplastic foci of well-differentiated and organoid cells with membrane MUC1-related cell detachment, lepidic growth, tufting, stromal invasion and spread through air spaces. These cells preserved the secretion polarity and

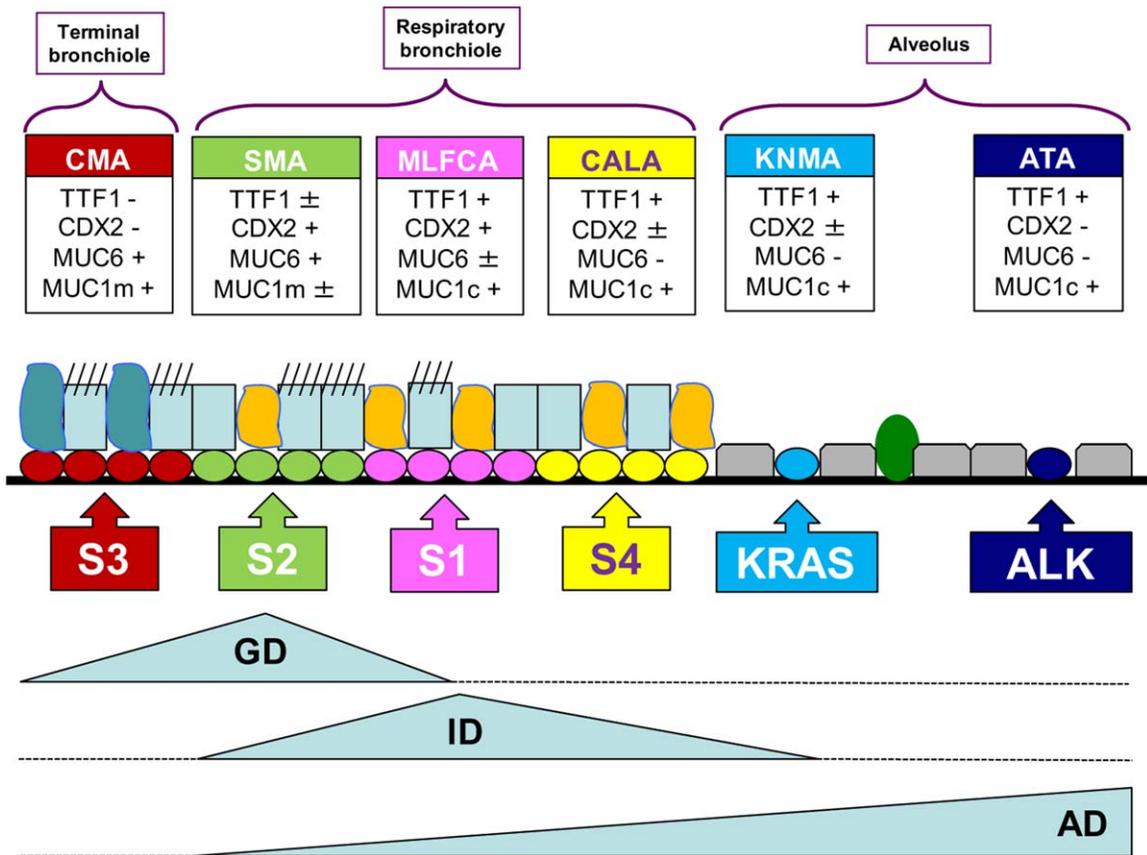


Figure 4. Spatial distribution of different stem/progenitor cell domains along bronchiolar and alveolar structures according to clustering analysis results. Cellular mucinous adenocarcinoma (CMA) belonging to *sub-cluster S3* showed exclusive gastric differentiation (GD) and developed from distal terminal bronchioles (red cells), while secreting mucinous adenocarcinoma (SMA) corresponding to *sub-cluster S3* with gastro-intestinal differentiation originated from proximal respiratory bronchioles (light green cells). Intestinal differentiation (ID) peaked in mucin lake-forming colloid adenocarcinoma (MLFCA) of intermediate respiratory bronchioles (pink cells), whereas alveolar differentiation (AD) steadily increased from *S2/S1* to *S4* *sub-clusters*, with colloid alveolar adenocarcinoma (CALA) arising from distal respiratory bronchioles (yellow cells), and peaked in *KRAS*-mutated non-mucinous adenocarcinoma (KNMA) and *ALK*-translocated adenocarcinoma (ATA). KNMA and ATA were thought to derive from stem/progenitor cells located in alveolar structures near terminal bronchioles (light blue cells) or inside alveolar cells (dark blue cells), respectively. In turn, ID was only residual in CALA and KNMA and completely disappeared in ATA. The normal bronchial epithelium comprises ciliated and non-ciliated cells (light grey), goblet cells (emerald green) and Clara cells (orange cells), while alveolar structures are covered by type I (flat grey cells) and type II pneumocytes (forest green).

exhibited basophilic mucin, negligible extravasation and uncommon freely floating elements. When combined gastrointestinal trans-differentiation occurred (*S2 sub-cluster*) upon HNF4A-induced CDX-2 and MUC2 activation [50,54] while persisting *Nkx2-1/TTF1* downregulation, multiple microscopic neoplastic foci of less differentiated tumour cells developed alongside a greater architectural complexity revealed by incipient enteric features, less preserved secretion polarity, amphiphilic mucin extravasation and freely floating tumour cells. Mitoses and apoptotic bodies were undetectable in *S3* but not enteric foci of *S2 sub-cluster*, indicating that apoptosis blockage [27] and cell proliferation diversely accounted for the

growth of these tumours (data not shown). *S3* and *S2 sub-clusters* were strikingly homologous to the gastric and pancreatobiliary subtypes of intra-ductal papillary mucinous neoplasms (IPMN) of the pancreas, respectively, inasmuch as the gastric subtype is often multifocal and the pancreatobiliary subtype shows CDX2 expression [55], in keeping also with a common embryologic origin from the foregut.

When the secretion phase predominated, mass-forming and localized tumours developed with either an exorbitant and grossly appreciable extracellular mucin pouring or microscopic mucin remnants encasing mucinous epithelium. If intestinal differentiation prevailed upon HNF4A-induced CDX-2 upregulation (*S1 sub-*

cluster), there were MUC1-positive amphophilic mucin lakes encasing freely floating elements, less architectural complexity, colloid-type stromal dissection and negligible tumour cell proliferation if not in less differentiated enteric foci (data not shown). These tumours sometimes co-expressed CDX-2 and TTF1 at the level of individual cells. This developmental scheme was strikingly similar to the intestinal subtype of pancreatic IPMN, which usually forms large cystic lesions involving the main duct [55]. If alveolar differentiation prevailed (*S4 sub-cluster*), mass-forming lung-type tumours were encountered with solid appearance, less differentiated enteric foci and MUC1-positive amphophilic mucin remnants encasing freely floating elements. NE cell component was virtually limited to these two *sub-clusters* only intestinal differentiation as hallmarked by garland-like architecture, eosinophilic cytoplasm, terminal bars and dirty necrosis crossed *S1*, *S2*, *S4* but not *S3 sub-cluster*. These observations would suggest that IEA are phenotypic changes only in lung IMA and ICA in keeping with current defining criteria [1], which are likely to be related to pancreatic rather than colorectal cancer [56].

Overall, we organized our results stepwise according to six different categories as highlighted in Figure 4 (MBA were excluded because they were completely unrelated tumours).

S3 sub-cluster tumours showed gastric-type differentiation [16] and more advanced stage likely related to impaired apoptosis. These tumours were thought to derive from distal terminal bronchioles and termed cellular mucinous adenocarcinoma.

S2 sub-cluster tumours maintained antral gland-type MUC6 as seen in incomplete intestinal metaplasia [57], whilst CDX-2 recruitment alongside negligible *Nkx2-1/TTF1* activity caused a more complete intestinal differentiation to appear [58] in less differentiated and more advanced tumours. This tumour category was thought to derive from proximal respiratory bronchioles and was termed secreting mucinous adenocarcinoma.

S1 sub-cluster tumours showed predominance of secreting intestinal (CDX-2/MUC2) differentiation responsible for inverted secretion polarity with freely floating cells, simplified architecture and less advanced tumours. These tumours were thought to derive from intermediate respiratory bronchioles and were termed mucin lake-forming colloid adenocarcinoma.

S4 sub-cluster tumours showed overwhelming alveolar lineage recruitment, with minor mucin secretion, less floating cells and less advanced tumour stage, mucinous epithelium or enteric features were seen. These tumours were thought to derive from distal

respiratory bronchioles and were termed colloid alveolar adenocarcinoma.

Non-mucinous KRASA featured *Nkx2-1/TTF1* and MUC1 expression [59], while repressed any secreting gastro-enteric lineage [26,27]. These KRASA with poor secretion were thought to derive from the most distal or juxta-alveolar region of the respiratory bronchiole and were termed KRAS-mutated non-mucinous adenocarcinoma.

ALK-translocated tumours silenced intestinal-type mucinous differentiation and activated depolarized MUC1-dependant cell disaggregation and vimentin-related epithelial-mesenchymal transition vimentin accumulation via *Nkx2-1/TTF1* and oncogenic ALK recruitment, respectively [59–61]. These tumours were thought to derive from proximal, juxta-bronchial progenitors of alveolar structures and were simply defined as ALK-translocated adenocarcinoma.

No synchronous or metachronous metastatic tumours were available for analysis in terms of inter-tumour heterogeneity, but the relative homogeneity of sub-clusters we observed made marked differences in genotypic or phenotypic traits quite unlikely to occur in this tumour setting. The figure of 79% for any identified gene mutation was in keeping with the literature [3,6,15,19,20], but studies on *CD74-NRG1* fusion of KRAS-negative tumours have been planned. Limitations of this the study were the relatively small number in each tumour category and the lack of validation sets, yet these lung tumours are in absolute terms quite rare and ALKA, KRASA and MBA were indeed validation control groups. The lack of survival analysis and the use of targeted NGS rather than whole exome/transcriptome analysis may have limited drawing definite conclusions. However, strength points were the identification of interpretative biomarkers in accordance with normal lung tissue distribution, multiple and non-random significant differences across the diverse tumour categories, a close genotype-phenotype relationship in the relevant tumour categories and coherence of results with current concepts on the development of MLA and tissue-specific adult stem cells in human lung [62].

Conclusion

Despite improvements in lung cancer classification, IMA, ICA, IEA and, to some extent, ALKA and non-mucinous KRASA continue to blur with each other for concurrent unanswered phenomena. In this scenario, six IHC biomarkers and simple molecular findings allowed an effective reappraisal of IMA,

ICA, IEA, ALKA and non-mucinous KRASA according to spatially organized and histology-independent tumour categories.

Acknowledgements

This work was supported by Novartis Farma Italia, Origgio (VA), Italy. The Funder had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript, which are responsibilities of the Authors only.

This work is dedicated with special mention to the memory of Carlotta, an extraordinarily lively girl who ultimately died of cancer in the prime of life.

Author contributions

GP conceived, designed and wrote the manuscript, reviewed all cases and performed immunohistochemistry analysis; AS collected materials and clinical data and contributed to write the manuscript; FB carried out cluster analysis, supervised all statistical procedures and contributed to write the manuscript; AF, MC, GR, AC, BV and UP collected materials and clinical data; ET, FP, AB and IC micro-dissected samples and performed molecular investigations; GP, AS, FB, AF and BV drafted the manuscript; GP, AS, FB, AF, MC, GR, AC, ET, FP, AB, IC, BV and UP revised critically the manuscript. GP finalized the manuscript. All authors approved the submitted version.

References

1. Travis WD, Brambilla E, Müller-Hermelink HK, Harris CC (Eds). World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of the Lung, Pleura, Thymus and Heart 4th Edn. IARC Press, Lyon, 2004.
2. Boland JM, Wampfler JA, Jang JS, *et al.* Pulmonary adenocarcinoma with signet ring cell features: a comprehensive study from 3 distinct patient cohorts. *Am J Surg Pathol* 2014; **38**: 1681–1688.
3. Kadota K, Yeh YC, D'angelo SP, *et al.* Associations between mutations and histologic patterns of mucin in lung adenocarcinoma: invasive mucinous pattern and extracellular mucin are associated with KRAS mutation. *Am J Surg Pathol* 2014; **38**: 1118–1127.
4. Rossi G, Cavazza A, Righi L, *et al.* Napsin-A, TTF-1, EGFR, and ALK status determination in lung primary and metastatic mucin-producing adenocarcinomas. *Int J Surg Pathol* 2014; **22**: 401–407.
5. Rossi G, Murer B, Cavazza A, *et al.* Primary mucinous (so-called colloid) carcinomas of the lung: a clinicopathologic and immunohistochemical study with special reference to CDX-2 homeobox gene and MUC2 expression. *Am J Surg Pathol* 2004; **28**: 442–452.
6. Righi L, Vatrano S, Di Nicolantonio F, *et al.* Retrospective multi-center study investigating the role of targeted next-generation sequencing of selected cancer genes in mucinous adenocarcinoma of the lung. *J Thorac Oncol* 2016; **11**: 504–515.
7. Wislez M, Antoine M, Baudrin L, *et al.* Non-mucinous and mucinous subtypes of adenocarcinoma with bronchioloalveolar carcinoma features differ by biomarker expression and in the response to gefitinib. *Lung Cancer* 2010; **68**: 185–191.
8. Zenali MJ, Weissferdt A, Solis LM, *et al.* An update on clinicopathological, immunohistochemical, and molecular profiles of colloid carcinoma of the lung. *Hum Pathol* 2015; **46**: 836–842.
9. Wu J, Chu PG, Jiang Z, *et al.* Napsin A expression in primary mucin-producing adenocarcinomas of the lung: an immunohistochemical study. *Am J Clin Pathol* 2013; **139**: 160–166.
10. Hwang DH, Sholl LM, Rojas-Rudilla V, *et al.* KRAS and NKX2-1 mutations in invasive mucinous adenocarcinoma of the lung. *J Thorac Oncol* 2016; **11**: 496–503.
11. Geles A, Gruber-Moesenbacher U, Quehenberger F, *et al.* Pulmonary mucinous adenocarcinomas: architectural patterns in correlation with genetic changes, prognosis and survival. *Virchows Arch* 2015; **467**: 675–686.
12. Yatabe Y, Koga T, Mitsudomi T, *et al.* CK20 expression, CDX2 expression, K-ras mutation, and goblet cell morphology in a subset of lung adenocarcinomas. *J Pathol* 2004; **203**: 645–652.
13. Awaya H, Takeshima Y, Yamasaki M, *et al.* Expression of MUC1, MUC2, MUC5AC, and MUC6 in atypical adenomatous hyperplasia, bronchioloalveolar carcinoma, adenocarcinoma with mixed subtypes, and mucinous bronchioloalveolar carcinoma of the lung. *Am J Clin Pathol* 2004; **121**: 644–653.
14. Tsuta K, Ishii G, Nitadori J, *et al.* Comparison of the immunophenotypes of signet-ring cell carcinoma, solid adenocarcinoma with mucin production, and mucinous bronchioloalveolar carcinoma of the lung characterized by the presence of cytoplasmic mucin. *J Pathol* 2006; **209**: 78–87.
15. Sugano M, Nagasaka T, Sasaki E, *et al.* HNF4alpha as a marker for invasive mucinous adenocarcinoma of the lung. *Am J Surg Pathol* 2013; **37**: 211–218.
16. Honda T, Ota H, Ishii K, *et al.* Mucinous bronchioloalveolar carcinoma with organoid differentiation simulating the pyloric mucosa of the stomach: clinicopathologic, histochemical, and immunohistochemical analysis. *Am J Clin Pathol* 1998; **109**: 423–430.
17. Tam A, Morrish D, Wadsworth S, *et al.* The role of female hormones on lung function in chronic lung diseases. *BMC Womens Health* 2011; **11**: 24.
18. Tam A, Wadsworth S, Dorscheid D, *et al.* Estradiol increases mucus synthesis in bronchial epithelial cells. *PLoS One* 2014; **9**: e100633.
19. Lee B, Lee T, Lee SH, *et al.* Clinicopathologic characteristics of EGFR, KRAS, and ALK alterations in 6,595 lung cancers. *Oncotarget* 2016; **7**: 23874–23884.
20. Finberg KE, Sequist LV, Joshi VA, *et al.* Mucinous differentiation correlates with absence of EGFR mutation and presence of KRAS mutation in lung adenocarcinomas with bronchioloalveolar features. *J Mol Diagn* 2007; **9**: 320–326.
21. Osoegawa A, Kometani T, Nosaki K, *et al.* LKB1 mutations frequently detected in mucinous bronchioloalveolar carcinoma. *Jpn J Clin Oncol* 2011; **41**: 1132–1137.

22. Skoulidis F, Byers LA, Diao L, et al. Co-occurring genomic alterations define major subsets of KRAS-mutant lung adenocarcinoma with distinct biology, immune profiles, and therapeutic vulnerabilities. *Cancer Discov* 2015; **5**: 860–877.
23. Fernandez-Cuesta L, Plenker D, Osada H, et al. CD74-NRG1 fusions in lung adenocarcinoma. *Cancer Discov* 2014; **4**: 415–422.
24. Nakaoku T, Tsuta K, Ichikawa H, et al. Druggable oncogene fusions in invasive mucinous lung adenocarcinoma. *Clin Cancer Res* 2014; **20**: 3087–3093.
25. Rodig SJ, Mino-Kenudson M, Dacic S, et al. Unique clinicopathologic features characterize ALK-rearranged lung adenocarcinoma in the western population. *Clin Cancer Res* 2009; **15**: 5216–5223.
26. Maeda Y, Tsuchiya T, Hao H, et al. Kras(G12D) and Nkx2-1 haplo insufficiency induce mucinous adenocarcinoma of the lung. *J Clin Invest* 2012; **122**: 4388–4400.
27. Snyder EL, Watanabe H, Magendanz M, et al. Nkx2-1 represses a latent gastric differentiation program in lung adenocarcinoma. *Mol Cell* 2013; **50**: 185–199.
28. Winslow MM, Dayton TL, Verhaak RG, et al. Suppression of lung adenocarcinoma progression by Nkx2-1. *Nature* 2011; **473**: 101–104.
29. Chu PG, Chung L, Weiss LM, et al. Determining the site of origin of mucinous adenocarcinoma: an immunohistochemical study of 175 cases. *Am J Surg Pathol* 2011; **35**: 1830–1836.
30. Krasinskas AM, Chiosea SI, Pal T, et al. KRAS mutational analysis and immunohistochemical studies can help distinguish pancreatic metastases from primary lung adenocarcinomas. *Mod Pathol* 2014; **27**: 262–270.
31. Laszlo T, Lacza A, Toth D, et al. Pulmonary enteric adenocarcinoma indistinguishable morphologically and immunohistologically from metastatic colorectal carcinoma. *Histopathology* 2014; **65**: 283–287.
32. Yousem SA. Pulmonary intestinal-type adenocarcinoma does not show enteric differentiation by immunohistochemical study. *Mod Pathol* 2005; **18**: 816–821.
33. Cha YJ, Han J, Hwang SH, et al. ALK-rearranged adenocarcinoma with extensive mucin production can mimic mucinous adenocarcinoma: clinicopathological analysis and comprehensive histological comparison with KRAS-mutated mucinous adenocarcinoma. *Pathology* 2016; **48**: 325–329.
34. Cai D, Li H, Wang R, et al. Comparison of clinical features, molecular alterations, and prognosis in morphological subgroups of lung invasive mucinous adenocarcinoma. *Onco Targets Ther* 2014; **7**: 2127–2132.
35. Pelosi G, Gasparini P, Cavazza A, et al. Multiparametric molecular characterization of pulmonary sarcomatoid carcinoma reveals a nonrandom amplification of anaplastic lymphoma kinase (ALK) gene. *Lung Cancer* 2012; **77**: 507–514.
36. Pelosi G, Gasparini P, Conte D, et al. Synergistic activation upon MET and ALK coamplification sustains targeted therapy in sarcomatoid carcinoma, a deadly subtype of lung cancer. *J Thorac Oncol* 2016; **11**: 718–728.
37. Pelosi G, Pellegrinelli A, Fabbri A, et al. Deciphering intra-tumor heterogeneity of lung adenocarcinoma confirms that dominant, branching, and private gene mutations occur within individual tumor nodules. *Virchows Arch* 2016; **468**: 651–662.
38. Pelosi G, Fabbri A, Papotti M, et al. Dissecting pulmonary large-cell carcinoma by targeted next generation sequencing of several cancer genes pushes genotypic-phenotypic correlations to emerge. *J Thorac Oncol* 2015; **10**: 1560–1569.
39. Lakhani SR, Ellis IO, Schnitt SJ, Tan PH, van de Vijver MJ (Eds). WHO Classification of Tumours of the Breast. WHO/IARC Classification of Tumours, 4th Edition, Volume 4. Lyon 2012.
40. Travis WD, Giroux DJ, Chansky K, et al. The IASLC Lung Cancer Staging Project: proposals for the inclusion of broncho-pulmonary carcinoid tumors in the forthcoming (seventh) edition of the TNM classification for lung cancer. *J Thorac Oncol* 2008; **3**: 1213–1223.
41. Kunii R, Jiang S, Hasegawa G, et al. The predominant expression of hepatocyte nuclear factor 4 alpha (HNF4 alpha) in thyroid transcription factor-1 (TTF-1)-negative pulmonary adenocarcinoma. *Histopathology* 2011; **58**: 467–476.
42. Sugai M, Umezue H, Yamamoto T, et al. Expression of hepatocyte nuclear factor 4 alpha in primary ovarian mucinous tumors. *Pathol Int* 2008; **58**: 681–686.
43. Ullmann R, Bongiovanni M, Halbwedl I, et al. Bronchiolar columnar cell dysplasia–genetic analysis of a novel preneoplastic lesion of peripheral lung. *Virchows Arch* 2003; **442**: 429–436.
44. Rossi G, Gasser B, Sartori G, et al. MUC5AC, cytokeratin 20 and HER2 expression and K-RAS mutations within mucinogenic growth in congenital pulmonary airway malformations. *Histopathology* 2012; **60**: 1133–1143.
45. Lantuejoul S, Nicholson AG, Sartori G, et al. Mucinous cells in type 1 pulmonary congenital cystic adenomatoid malformation as mucinous bronchioloalveolar carcinoma precursors. *Am J Surg Pathol* 2007; **31**: 961–969.
46. Maeda Y, Chen G, Xu Y, et al. Airway epithelial transcription factor NK2 homeobox 1 inhibits mucous cell metaplasia and Th2 inflammation. *Am J Respir Crit Care Med* 2011; **184**: 421–429.
47. Boyd M, Bressendorff S, Moller J, et al. Mapping of HNF4alpha target genes in intestinal epithelial cells. *BMC Gastroenterol* 2009; **9**: 68.
48. Kim DH, Shin N, Kim GH, et al. Mucin expression in gastric cancer: reappraisal of its clinicopathologic and prognostic significance. *Arch Pathol Lab Med* 2013; **137**: 1047–1053.
49. Saandi T, Baraille F, Derbal-Wolfrom L, et al. Regulation of the tumor suppressor homeogene Cdx2 by HNF4alpha in intestinal cancer. *Oncogene* 2013; **32**: 3782–3788.
50. Mesquita P, Almeida R, Van Seuning I, et al. Coordinated expression of MUC2 and CDX-2 in mucinous carcinomas of the lung can be explained by the role of CDX-2 as transcriptional regulator of MUC2. *Am J Surg Pathol* 2004; **28**: 1254–1255.
51. Guddo F, Giatromanolaki A, Koukourakis MI, et al. MUC1 (episialin) expression in non-small cell lung cancer is independent of EGFR and c-erbB-2 expression and correlates with poor survival in node positive patients. *J Clin Pathol* 1998; **51**: 667–671.
52. McAuley JL, Linden SK, Png CW, et al. MUC1 cell surface mucin is a critical element of the mucosal barrier to infection. *J Clin Invest* 2007; **117**: 2313–2324.
53. Lakshmanan I, Ponnusamy MP, Macha MA, et al. Mucins in lung cancer: diagnostic, prognostic, and therapeutic implications. *J Thorac Oncol* 2015; **10**: 19–27.
54. Boudreau F, Rings EH, van Wering HM, et al. Hepatocyte nuclear factor-1 alpha, GATA-4, and caudal related homeodomain protein Cdx2 interact functionally to modulate intestinal

- gene transcription. Implication for the developmental regulation of the sucrase-isomaltase gene. *J Biol Chem* 2002; **277**: 31909–31917.
55. Kloppel G, Basturk O, Schlitter AM, *et al.* Intraductal neoplasms of the pancreas. *Semin Diagn Pathol* 2014; **31**: 452–466.
 56. Garajova I, Funel N, Fiorentino M, *et al.* MicroRNA profiling of primary pulmonary enteric adenocarcinoma in members from the same family reveals some similarities to pancreatic adenocarcinoma—a step towards personalized therapy. *Clin Epigenetics* 2015; **7**: 129.
 57. Reis CA, David L, Correa P, *et al.* Intestinal metaplasia of human stomach displays distinct patterns of mucin (MUC1, MUC2, MUC5AC, and MUC6) expression. *Cancer Res* 1999; **59**: 1003–1007.
 58. Silberg DG, Sullivan J, Kang E, *et al.* Cdx2 ectopic expression induces gastric intestinal metaplasia in transgenic mice. *Gastroenterology* 2002; **122**: 689–696.
 59. Kolla V, Gonzales LW, Gonzales J, *et al.* Thyroid transcription factor in differentiating type II cells: regulation, isoforms, and target genes. *Am J Respir Cell Mol Biol* 2007; **36**: 213–225.
 60. ten Berge RL, Snijdwint FG, von Mensdorff-Pouilly S, *et al.* MUC1 (EMA) is preferentially expressed by ALK positive anaplastic large cell lymphoma, in the normally glycosylated or only partly hypoglycosylated form. *J Clin Pathol* 2001; **54**: 933–939.
 61. Voena C, Varesio LM, Zhang L, *et al.* Oncogenic ALK regulates EMT in non-small cell lung carcinoma through repression of the epithelial splicing regulatory protein 1. *Oncotarget* 2016; **7**: 33316–33330.
 62. Kajstura J, Rota M, Hall SR, *et al.* Evidence for human lung stem cells. *N Engl J Med* 2011; **364**: 1795–1806.

SUPPLEMENTARY MATERIAL ONLINE

Supplementary material A. Clinico-pathological details for the entire series of 44 lung adenocarcinoma samples under evaluation

Supplementary material B. Monoclonal and polyclonal antibody list used in the current study

Supplementary material C. Immunohistochemistry procedures and distribution of immunohistochemistry biomarkers in normal lung tissue

Supplementary material D. Immunohistochemistry details according to conventional subtyping of adenocarcinoma

Supplementary material E. Clinico-pathological data by unsupervised sub-clustering S1 to S4

Supplementary material F. Immunohistochemistry and molecular data by unsupervised sub-clustering S1 to S4

Supplementary figure legends

Figure S1. Distribution of TTF1, MUC5AC, MUC1 and HNF4A in normal lung tissue

Figure S2. Relationship between sub-clusters S1 → S4 and TTF1, MUC1, CDX-2, MUC2 and MUC6 immunostaining

Figure S3. Relationship between sub-clusters S1 → S4 as defined by immunohistochemistry findings and histological classification of mucinous and colloid adenocarcinoma, either pure or mixed

Figure S4. Mucin lake-forming colloid adenocarcinoma (*sub-cluster S1*)

Figure S5. Secreting mucinous adenocarcinoma (*sub-cluster S2*)

Figure S6. Cellular mucinous adenocarcinoma (*sub-cluster S3*)

Figure S7. Alveolar colloid adenocarcinoma (*sub-cluster S4*)

Figure S8. Non mucinous *KRAS*-mutated adenocarcinoma

Figure S9. *ALK*-rearranged adenocarcinoma

Figure S10. Mutation analysis by targeted next generation sequencing and direct Sanger sequencing