Dissecting pulmonary large-cell carcinoma by targeted next generation sequencing of several cancer genes pushes genotypic-phenotypic correlations to emerge

This is the author's manuscript

Original Citation:

Availability:
This version is available http://hdl.handle.net/2318/1532760 since 2016-10-17T17:11:11Z

Published version:
DOI:10.1097/JTO.0000000000000658

Terms of use:
Open Access
Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.
Dissecting pulmonary large cell carcinoma by targeted next generation sequencing of several cancer genes pushes genotypic-phenotypic correlations to emerge

Giuseppe Pelosi, MD, MIAC, 1,2 Alessandra Fabbri, MD, 1 Mauro Papotti, MD, 3 Giulio Rossi, MD, 4 Alberto Cavazza, MD, 5 Luisella Righi, MD, 3 Elena Tamborini, DSc, 1 Federica Perrone, DSc, 1 Giulio Settanni, DSc, 1 Adele Busico, DSc, 1 Maria Adele Testi, DSc, 1 Patrick Maisonneuve, Eng, 6 Filippo De Braud, MD, 7 Marina Garassino, MD, 7 Barbara Valeri, MD, 1 Angelica Sonzogni, MD, 1 and Ugo Pastorino, MD 8

1 Department of Pathology and Laboratory Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy
2 Department of Biomedical and Clinical Sciences “Luigi Sacco”, Università degli Studi, Milan, Italy
3 Department of Oncology, University of Turin at San Luigi Hospital, Orbassano, Torino, Italy
4 Pathologic Anatomy, Azienda Ospedaliero-Universitaria, Policlinico di Modena, Modena, Italy
5 Department of Oncology and Advanced Technology, Operative Unit of Pathologic Anatomy, IRCCS Azienda Arcispedale S. Maria Nuova, Reggio Emilia, Italy
6 Division of Epidemiology and Biostatistics, European Institute of Oncology, Milan, Italy
7 Department of Medical Oncology, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy
8 Division of Thoracic Surgery, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy

Short title: Geno-phenotyping pulmonary large cell carcinoma

Conflict of interest statement: The Authors declare that they have no conflicts of interest.

This work was supported by a finalized grant of Scientific Directorate of the Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy (accession number INT-145/14) to cover reagent costs, by LILT (Lega Italiana per la Lotta contro i Tumori, Sezione di Milano, Milan, Italy) to cover research personnel costs, and by AIRC (Associazione Italiana per la Ricerca
sul Cancro) to cover administrative costs. The Funders 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.

Mailing address for correspondence:

Giuseppe Pelosi, MD, MIAC
Dipartimento di Patologia Diagnostica e Laboratorio
Fondazione IRCCS Istituto Nazionale dei Tumori
Via G. Venezian, 1
I-20133 Milano
ITALY
phone: + 39 02 23902260/2876/3017
fax: + 39 02 23902877
E-mail: giuseppe.pelosi@unimi.it
Abstract

**Hypothesis.** Little is known about genotypic and phenotypic correlations in undifferentiated large cell carcinoma (LCC) of the lung.

**Methods.** Thirty LCC were dissected by unsupervised targeted next generation sequencing (T-NGS) analysis for 50 cancer-associated oncogenes and tumor suppressor genes. Cell differentiation lineages were unveiled by using thyroid transcription factor-1 (TTF1) for adenocarcinoma (ADC) and ΔNp63/p40 for squamous cell carcinoma (SQC), dichotomizing immunohistochemistry (IHC) results for TTF1 as negative or positive (whatever its extent) and for p40 as negative, positive or focal (if <10% of reactive tumor cells).

**Results.** Three LCC were wild type (all TTF1+/p40-), whilst the remaining 27 (90%) tumors had at least one gene mutation. Twenty-four cases featuring TTF1+/p40-, TTF1+/p40±, TTF1-/p40± or TTF1-/p40- phenotypes comprised ATM, BRAF, CDKN2A, EGFR, ERBB4, FBXW7, FLT3, KRAS, NRAS, PIK3CA, PTPN11, RET, SMAD4, SMO, STK11 or TP53 mutations in keeping with ADC lineage while three tumors showing TTF1-/p40+ phenotype harbored TP53 only and no ADC-related mutations in keeping with SQC lineage. Single, double, triple, quadruple and quintuple mutations occurred in 16, 6, 2, 2 and 1 patient, respectively. The occurrence of ≥3 mutations but not any IHC categorization predicted shorter overall (p=0.001) and disease-free survival (p=0.007), independent of age, sex and tumor stage.

**Conclusions.** Albeit preliminary also because of the relatively small number of LCC under evaluation, this T-NGS study, however, revealed gene mutation heterogeneity in LCC with some genotypic-phenotypic correlations. Negativity or focal occurrence of p40 made SQC diagnosis unlikely on molecular grounds, but suggested ADC confirming validity of the axiom “no p40, no squamous”.

**Key words:** Large cell carcinoma; lung; immunohistochemistry; targeted next generation sequencing; p40; TTF-1
Introduction

Large cell carcinoma of the lung is a poorly differentiated tumor lacking cytological, architectural and immunohistochemistry (IHC) features of small cell carcinoma (SCLC), adenocarcinoma (ADC) or squamous cell carcinoma (SQC). The previous 2004 WHO classification included different histological variants under the umbrella term of LCC, namely undifferentiated LCC (U-LCC), large cell neuroendocrine (NE) carcinoma (LCNEC), basaloid carcinoma (BC), lymphoepithelioma-like carcinoma (LELC), LCC with rhabdoid phenotype (LCC-RF) and clear cell carcinoma (CCC). The new 2015 WHO classification has attributed a decisional role to IHC characterization of LCC based on a minimalist panel antibody approach, which now are confined to three tumor categories only (LCC with null IHC features, LCC with unclear IHC features and LCC with no additional stains).

In LCC, IHC has been unveiling the same lineage heterogeneity, as previously demonstrated by means of electron microscopy, indicating that squamous, NE and, especially, glandular cell differentiation underlie most LCC, whilst completely uncommitted tumors are decidedly uncommon. Among the many IHC biomarkers aimed at discovering the hidden face of LCC, thyroid transcription factor-1 (TTF1) and Δ(delta)Np63 (henceforth, simply p40), proved the best diagnostic biomarkers to highlight ADC and SQC lineages, respectively. Although the duet TTF1/p40 remains the most reliable predictor of cell lineage and therefore of the ultimate diagnosis in lung cancer, some uncertainty may arise in undifferentiated tumors when faced with complete absence of both biomarkers (null phenotype: TTF1/p40-) or focal labeling for squamous differentiation biomarkers, such as p40, in otherwise TTF1-negative tumors (unclear phenotype: TTF1-/p40±).

A growing body of information is accumulating about the occurrence of non-random genetic alterations in defined subtypes of lung cancer. New actionable driver genes are emerging from the molecular landscape of lung cancer, especially when using multiplexed or unbiased technologies of next generation sequencing (NGS). NGS/T-NGS data or molecular investigations on LCC are relatively scant for either novelty of these technologies or progressive disappearance of such histological type. Recent surveys on the subject, however, have provided strong evidence that many, if not all, LCC display genetic
profiles (including microRNA expression) mostly aligned with ADC and/or SQC, but the approach has been to correlate a-priori-defined IHC diagnoses with underlying gene alterations rather than to interpret IHC profiles according to the relevant mutation portrait.

This study was aimed at establishing the relationship between genotype and phenotype in LCC also according to tumor categorization of the current 2015 WHO classification by comparing their molecular profile assessed by unsupervised T-NGS with stochastically defined IHC categories according to biomarkers of glandular and squamous cell lineages (TTF1 and p40) and then at a posteriori dissecting tumors on the basis of the preferential distribution of gene mutations between ADC and SQC.
Materials and methods

Design of the study

We designed and conducted a two-phase investigation to test the relationship between genotype of LCC according to T-NGS analysis and phenotype according to TTF1 and p40 IHC, thereby providing a biological rationale to the diverse diagnostic algorithms with particular reference to null and unclear phenotypes. To this regard, in the first phase, we accomplished an unsupervised T-NGS analysis on 30 LCC by using a large panel of 50 oncogenes and tumor suppressor genes recurrently altered in human cancers and compared molecular results with clinico-pathological characteristics of patients, including survival. In the second phase, we attributed the relevant gene mutations to the diverse IHC-prioritized diagnostic algorithms on the basis of their own preferential distribution in the two major categories of lung cancer, i.e. ADC and SQC, to molecularly validate the role of these biomarkers in constructing decisional algorithms even in undifferentiated tumors.

Patients and tumors

A series of 30 consecutive surgical specimens of LCC from 24 males (range 47-87 years; mean±SD 52.0±14.8 years) and 6 females (range 61-72 years; mean±SD 65.5±10.9 years) were identified in the pathology archives of the participant Institutions (Milan, Turin, Modena and Reggio Emilia). The lack of a previous history of cancer elsewhere in the body and the availability of complete clinical information were required for entering the study. Pertinent data regarding the 30 LCC patients as a function of gender, smoking status and tumor stage are summarized in Table 1. Surgical specimens consisted of lobectomy or pneumonectomy along with radical mediastinal lymph node resection to ensure accurate staging. According to the 7th edition of the TNM staging system, there were three tumors staged IA, nine IB, nine IIA, four IIB, three IIIA, one IIIB, and one IV (the latter featuring pT4pN0pM1a). Fifteen patients were current smokers and 15 former smokers.

All surgical specimens had been fixed in 4% buffered formaldehyde solution for 12-24 hours and embedded in paraffin according to standard histopathological methods. All the original hematoxylin and eosin (H&E)-stained sections were jointly reviewed by some Authors
(GP, MP, GR, AC) without knowledge of the patients' identities or original tumor categorization. Solid adenocarcinoma with mucin was excluded by mucin stain, NE differentiation by morphology and NE biomarkers, sarcomatoid carcinoma by morphology and vimentin IHC, and unexpected metastases by clinical history and appropriate IHC work-up. According to 2004 WHO classification on lung cancer primarily based upon morphology, the study comprised 20 U-LCC, 6 LCC-CCC, 3 LCC-RC and 1 LCC-LELC, which were considered poorly differentiated by definition. Out of eight patients with lymph node metastases, five were pN1 and three pN2, with involved lymph node percentage ranging from 5% to 67%. Vascular invasion was seen in 17 tumors and necrosis in 28, the latter ranging from 10% to 70% of the entire tumor mass. Pleural invasion, classified according to updated criteria, was documented in 15 tumors, resulting in PL1, PL2 and PL3 in 11, 2, and 2 cases, respectively. Tumor size ranged from 10 to 120 mm, with a mean value of 41.8±23.8 mm.

Complete clinical follow-up (updated to July 2014) was available for all but one patient lost to follow-up, with overall survival averaging 47±37 months (median 32; range 1-132). During this period, 13 (44.8%) patients had recurrent disease (9 systemic, 3 in the lungs and 1 in the brain) and 12 (41.3%) died of disease, with mean disease-free and overall survivals being 43±38 months (median 31; range 1-132) and 47±37 months (median 32; range 1-132), respectively.

Ethics

The study was notified to and approved by the independent ethics committee of the "Fondazione IRCCS Istituto Nazionale Tumori", Milan, Italy (accession number INT-145/14). All patients gave their written consent for diagnosis and research activities when they were admitted to the hospital.

Sequencing, fluorescence in situ hybridization and immunohistochemistry procedures

A detailed description of molecular and IHC laboratory procedures we pursued in this investigation, including DNA extraction and quantification, fluorescence in situ hybridization (FISH), T-NGS and Sanger analyses and the relevant antibodies, is reported in S1_Molecular&IHC method and Table 2, respectively. Cut-off of 10% for focal expression of
p40 was chosen on the basis of recently refined IHC methods\textsuperscript{1, 3, 27, 58}. The same paraffin tumor blocks were used for every type of analysis, whether sequencing or IHC. In particular, FISH analysis for FGFR1 gene was accomplished in the subsets of tumor patients with null (TTF1-/p40-) or unclear (TTF1-/p40±) phenotype according to its preferential, albeit not exclusive, prevalence in squamous cell carcinoma as compared to adenocarcinoma\textsuperscript{59-61}.

**Statistical analysis**

Qualitative data were compared by Fisher's exact probability test or chi-square test as appropriate. Overall survival (OS) was defined as the time between surgery and the last follow-up or cancer death. If a patient died without cancer recurrence, the patient's survival time was censored at the time of death. Only lung cancer-related deaths or recurrences were considered to be events. Disease-free survival (DFS) was calculated from the date of surgery to the date of progression or the date of last follow-up. Survival estimates were calculated with Kaplan-Meier's method and compared by the log rank test. The Cox proportional hazard regression model was used to evaluate the simultaneous effect of explanatory variables on survival time. All analyses were carried-out using the SAS statistical software (SAS Institute, Inc., Cary, NC, USA). For all tests, only two-sided $p$-values were taken into account (with a threshold of $<0.05$ for statistical significance) and confidence intervals set at the 95% level.
Results

Mutation landscape of large cell carcinoma

Three LCC patients turned out to be wild type, whereas the remaining 27 (90%) patients showed 47 recurring mutations in 16 different genes, accounting for 1.7 mutations per patient on average. Single, double, triple, quadruple and quintuple mutations occurred in 16, 6, 2, 2 and 1 patient, respectively (S3_Table 3 and Table 3). The most frequent alterations affected TP53 gene with 19 mutations in 17 patients, followed by KRAS in 7 patients, ATM in 3 patients, BRAF, EGFR, FBXW7, NRAS and PIK3CA in 2 patients, and CDKN2A, ERBB4, FLT3, PTPN11, RET, SMO, SMAD4 and STK11 in 1 patient for every mutation. Eleven patients let multiple mutations emerge beyond TP53 (ATM, BRAF, CDKN2A, ERBB4, EGFR, FBXW7, KRAS, FLT3, PTPN11, PIK3CA, SMAD4 and STK11) (Table 3 and S1_Table 1). Frequency of mutated allelic DNA ranged from 6% (RET) to 76% (ATM), with the mean±SD threshold being 30.0±19.9% for the 16 mutated genes considered as a whole (S2_Table 2). Apart from two deletions occurring in TP53 and CDKN2A genes, the other mutations comprised 25 transversions and 20 transitions, with no significant difference among the diverse genes. No preferential distribution of mutations was observed according to several clinico-pathologic variables (age, gender, smoking habit, tumor stage or size, histological classification, vascular or pleural invasion or necrosis amount). TP53 mutated tumors significantly over-expressed (p=0.0019) the relevant protein (57.6±42.8%, range 0-100%, median 80%) compared to non-mutated tumors (9.0±18.5%, range 0-60%, median 0%) (Table 3). Moreover, apart from a single patient with TP53 deletion who did not show any p53 expression, there was a trend for stop mutations to down-regulate the relevant protein (8 cases, 46.2±47.5% tumor cells) compared to non-stop mutations (10 cases, 76.5±30.8% tumor cells) (p=0.109) (S1_Table1 and S1_Fig 1). To ensure accuracy, BRAF, EGFR, KRAS, NRAS, PIK3CA, RET, STK11 and TP53 mutations were re-sequenced with Sanger method (see also S1_Molecular&IHC method).

Phenotypic categories of large cell carcinoma
Five different diagnostic algorithms were obtained by stochastically crossing TTF1 and p40 IHC according to positive, negative and focal categories (Table 4). No significant relationship was found between IHC categories and morphologic classification of LCC. Relative to IHC, TTF1 decoration varied from 20% to 100% tumor cells (mean±SD: 61.3±34.5%), while p40 ranged from 2% to 100% tumor cells (mean±SD: 34.0±39.6%). No significant differences were observed between p40 polyclonal and monoclonal antibody. Additional IHC work-up with p63 and CK7 did not provide useful information for better unveiling null or unclear phenotype (see also S2_Result). Representative features of TTF1 and p40 decoration in the diverse phenotype combinations are presented in Fig 1, where p63 and CK7 stains are also included as picture insets.

Genotypic and phenotypic correlations of large cell carcinoma

The distribution of 27 gene-mutated tumor patients among the five categories based on stochastically crossing TTF1 and p40 IHC is shown in Table 4, which includes 6 tumors with TTF1+/p40- phenotype, 1 tumor with TTF1+/p40± phenotype, 3 tumors with TTF1-/p40+ phenotype, 13 tumors with null TTF1-/p40- phenotype and 4 tumors with unclear TTF1-/p40± phenotype.

Taking advantage of the molecular distribution, all tumors negative or only focally positive for p40, regardless of TTF1 IHC results (categories TTF1+/p40-, TTF1+/p40±, TTF1-/p40± and TTF1-/p40-) were molecularly akin to ADC (LCC-favor ADC), whereas TTF1-/p40+ category was molecularly in keeping with SQC for presenting TP53 mutations only along with FGFR1 amplification in two of three cases (Fig 2 A,B), (LCC-favor SQC). FGFR1 FISH analysis was extended to the 13 tumors with null (TTF1-/p40-) and the 4 tumors with unclear (TTF1-/p40±), finding that there was only one out of 17 (6%) tumor samples to exhibit FGFR1 gene amplification (p=0.046) (Fig 2 C,D).

The 11 tumors harboring multiple mutations corresponded to ADC patients with TTF1+/p40-, TTF1-/p40- or TTF1+/p40± phenotype in 5, 5 and 1 tumor, respectively.

Patient survival correlations
General survival of the 30 LCC patients under evaluation for both OS and DFS is shown in S2_Fig 2. Tumor patients expressing 3 mutations or more experienced a shorter OS (p=0.001) and DFS (p=0.007) independent of tumor stage (p=0.827) in comparison with wild type or up-to-2-gene-mutated patients (Fig 3). Other factors affecting OS and DFS were tumor stage, tumor size and pleural invasion, but none of them or the number of mutations per patient emerged as independent prognostic factors at multivariate analysis. Categorization of LCC according to the diverse IHC diagnostic algorithms or morphologic subclasses failed to demonstrate any survival implication.
Discussion

In this study, we performed a mutation-based classification of LCC by taking advantage of T-NGS for a large number of tumor suppressor genes and oncogenes repeatedly altered in human cancers. This was compared with a morphologic classification and unsupervised IHC diagnostic algorithms of tumors resulting from stochastically crossing two highly specific biomarkers of glandular and squamous cell differentiation, namely TTF1 and p40, respectively. At variance with similar recent works on LCC where tumor diagnostic categories were assigned before genotyping according to a relatively small number of gene mutations\textsuperscript{16-18, 48, 49}, we designed this two-phase investigation to minimize not only an \textit{a priori} interpretation of undifferentiated tumors according to decisional IHC algorithms suited for conventional NSCLC but also to rule out “aberrant/illegitimate” acquisition of biomarkers by undifferentiated tumor cells due to loss of relevant cell fate niches or unfaithful expression of nuclear transcription factors. As there were some limitations to the study due to the relatively small number of LCC under evaluation, the lack of SQC-specific mutations in the 50-gene panel we used for the analysis (such as DDR2, KEAP1 and NFE2L2/NRF2), some overlap in the mutation scenario between ADC and SQC and the uncertain meaning of some uncommon mutations when referring to a certain tumor subtype, the results should be better considered as preliminary and worthwhile being further confirmed in future validation tumor series. The five possible phenotypic categories we encountered in our tumor series (TTF1+/p40+ was the only missing phenotype) mirrored to some extent the underlying molecular portrait, which was preferentially relative to ADC or not excluding SQC when considered as a whole thereby allowing a robust biomarker-based classification of LCC to be afforded by T-NGS, independent of gender, age and tumor stage. All phenotypic categories were then \textit{a posteriori} dissected on the basis of the preferential distribution of underlying gene mutations between ADC and SQC (Table 4 and S3_Table3), in the premise that DNA mutations were more straightforward and robust than phenotype (potentially affected by epigenetic changes). Accordingly, mutations variably described to cluster in lung ADC, such as \textit{ATM}, \textit{BRAF}, \textit{EGFR}, \textit{ERBB4}, \textit{FBXW7}, \textit{FLT3}, \textit{KRAS}, \textit{NRAF}, \textit{PTPN11}, \textit{RET}, \textit{STK11}, \textit{SMAD4} and \textit{SMO}, albeit not completely specific for this tumor type and possibly with meaning of passenger or transitory mutations (destined to be fixed or disappear over time),
were distributed as a whole into IHC categories unified by absent or only focal (<10%) expression of p40 irrespective of TTF1 status thereby supporting the ultimate diagnosis of serendipitous ADC or non-squamous NSCLC (a questionable terminology that is however largely used by oncologists for patient treatment), whereas the complete lack of equipollent ADC-related mutations in the three cases featuring TTF1-/p40+ phenotype along with the presence of TP53 mutation (the most frequently mutated gene in SQC) and the co-occurrence of FGFR1 amplification in two out of these three (66%) cases reinforced the likelihood of really facing with undifferentiated SQC. Furthermore, we noted that FGFR1 gene amplification was uncommon (6%) in non-squamous-phenotype tumors of our series with null (TTF1-/p40-) or unclear (TTF1-/p40±) phenotype as compared to LCC-squamous (TTF1-/p40+) phenotype (p=0.046) and/or bearing mutations credited to be not so strictly specific for ADC, such as PIK3CA, CDKN2A, FBXW7, FLT3 or SMO (S3_Table 3 and Fig 2), indicating that these tumors were really more akin to ADC than SQC differentiation lineage, in keeping with reported prevalence data of low FGFR1 gene amplification in ADC. In other words, even if certain mutations or molecular events are not per se so strictly specific for a given tumor subtype as to be considered transitory or passively passenger tumor determinants, the same mutations or molecular events may become meaningful and coherent with that tumor type as a whole according to the proper IHC context. This managerial approach was proven to be particularly useful in null (13 cases) or unclear (4 cases) phenotypes, which were hard to classify because of the simultaneous absence or unclear expression of decisional biomarkers and the general prudence on the diagnostic interpretation of negative IHC algorithms. On the basis of these findings, we can reasonably exclude aberrant/illegitimate expression of either biomarker in our series of LCC, confirming molecularly the previously proposed IHC axiom “no p40, no squamous” in keeping to which, whenever there is no expression of p40 in the tumor cell population, the diagnosis of SQC is extremely unlikely.

Notably, our paper indicated that gene mutations were frequent (90%) events in LCC, and that about 40% of patients exhibited multiple mutations (S2_Table 2, Table 3). The high mutation rate of LCC correlated with a high frequency of mutated allelic DNA and a tobacco exposure-related signature (our patients were current or former smokers, Table 1) in keeping
with other lung carcinomas, such as SCLC\textsuperscript{69,70} or ADC\textsuperscript{35,36,43}. Multiplicity of mutations (S1_Table 1, S2_Table 2 and Table 3), while offering alternative options of targeted therapy for these life-threatening tumors (13 of our patients had recurrent disease and 12 died of disease even if 25 (83%) of whom were in stage I-II, see Table 1 and S2_Fig 2), may be responsible for secondary resistance development upon target therapy by selecting tumor clones and for dismal prognosis (Fig 3) likely due to the involvement of tumor suppressor genes and/or accumulating numbers of mutations, as recently suggested by multi-region sequencing studies.\textsuperscript{71}

Robustness of our T-NGS results was confirmed by means of Sanger re-sequencing for common mutations or IHC correlations for selected protein expression such as p53 (S1_Fig 1). The number of genes in our tumor series was greater than previously assessed in other similar works\textsuperscript{16,17,48,49}, revealing 16 different gene mutations underlying diverse biological mechanisms (S3_Table 3). Most of these mutations preferentially split into ADC or SQC, but some of the less frequently described mutations in lung cancer and in pulmonary ADC, such as \textit{ERBB4}, \textit{FBXW7}, \textit{FLT3}, \textit{NRAS}, \textit{PTPN11}, \textit{RET}, \textit{SMAD4} and \textit{SMO}, were interpreted as belonging to the ADC category not only on the basis of available gene mutation catalogues and the absence of FGFR1 amplification in these tumors, but also of the relevant IHC profiles all basically dominated by p40 negativity or focal (<10%) occurrence irrespective of TTF-1 status.

It could be speculated that LCC null/unclear phenotypes represent variant of TTF-1 negative ADC, which may account up to 20% of morphological ADC\textsuperscript{20}, whereas absence or focal presence of p40 is unlikely to occur in SQC\textsuperscript{27,31}, inasmuch as p40 acts as a master regulator of epidermal cell fate trans-activating genes of basal keratins\textsuperscript{66-68}. In keeping with our findings, other studies dealing with specific microRNA prevalence\textsuperscript{18} or different mutation profiles\textsuperscript{16,17} supported the notion that TTF1/p40 double negative LCC would be more akin to ADC, once NE tumors, sarcoma, melanoma or unexpected metastasis was ruled out. Focal p40 decoration in otherwise TTF1-negative LCC showing mutation traits of ADC could even underline focal squamous/basal-like (co)-lineage within individual tumor cells of ADC\textsuperscript{31,63,72-74} rather than an aberrant/illegitimate expression of these non-TA isoforms in ADC-differentiated LCC.
Tumor categories unveiled by IHC did not change if different antibodies to p40 were used, whether polyclonal or monoclonal, this indicating that both reagents pinpointed the same tumor cells in keeping with previously reported findings. Gene associations beyond TP53 were seen for KRAS with ATM, FLT3, PTPN11, PIK3CA, STK11 and SMAD4, and for EGFR with FBXW7 (Table 3 and S1_Table1), suggesting different activation pathways in the development of LCC (S2_Table2), but further investigation on a larger number of LCC is clearly warranted to confirm these preliminary data.

Conclusions

This is a preliminary molecular study based on T-NGS, which suggests - in keeping with IHC findings - that most LCC basically pinpoint ADC with a minority of SQC, and that p40 and TTF1-based IHC profiling is likely to mirror underlying differentiation lineages. This conclusion paves the way to an innovative and clinically oriented classification of LCC resulting from integration of gene mutation and IHC profiles to successfully align with precision medicine requirements (the right drug, to the right patient, at the right time).

Bullet points

1) LCC basically underlie dual cell lineages based on T-NGS and p40/TTF1 profiling
2) ΔNp63-p40 positivity unveils SQC while negativity underpins ADC in keeping with T-NGS to sustain the axiom “no p40, no squamous”
3) Null or unclear phenotype corresponds to ADC by merging molecular and IHC grounds
4) Integrating T-NGS and IHC may result in a clinically oriented classification on LCC
Acknowledgments

This work was dedicated to the memory of Carlotta, an extraordinarily lively girl who untimely died of cancer in the prime of life.
Table titles

**Table 1.** Clinico-pathological characteristics of the 30 patients with large cell carcinoma of the lung according to gender, smoking habit and tumor stage

**Table 2.** Antibody panel and immunostain conditions used in the current study

**Table 3.** Distribution of gene mutation and correlation with p53 protein immunohistochemical expression in the 30 large cell carcinoma patients under evaluation

**Table 4.** Morphologic, immunohistochemistry and molecular correlations in 30 large cell carcinomas

**S1_Table 1. Distribution of gene mutations.** Data on the distribution of gene mutation and the correlation with p53 protein immunohistochemical expression are provided in the 30 large cell carcinoma patients under evaluation.

**S2_Table 2.** Next generation sequencing results on 30 LCC patients.

**S3_Table 3.** Gene function. List of mutated genes found in 27 cases of LCC is shown with functional information.
Figure legends

Fig 1 A-O. Phenotypic and genotypic correlations of large cell carcinoma. Phenotypic and genotypic correlations of morphologically defined LCC (A: LCC-CCC; D: LCC-RC; G: LCC-LELC; J,M: LCC-U) between TTF1 (B,E,H,K,N) or p40 (C,F,I,L,O) protein expression (along with CK7 and p63 decoration in the insets, respectively), and the relevant gene mutation as derived from T-NGS analysis. The phenotypes TTF1+/p40- (A,B,C: case #04), TTF1+/p40± (D,E,F: case #05), TTF1-/p40- (J,K,L: case #10) and TTF1-/p40± (M,N,O: case #24) underpinned molecular alterations consistent with adenocarcinoma, whereas the phenotype TTF1-/p40+ (G,H,I: case: #7) showed TP53 mutation only with no other ADC-related gene mutations in keeping with the diagnosis of SQC. LCC-CCC stands for large cell carcinoma-clear cell carcinoma, LCC-RC for large cell carcinoma with rhabdoid cell phenotype, LCC-LELC for large cell carcinoma-lymphoepitelioma-like carcinoma, and LCC-U for large cell carcinoma-unclassified type.

Fig 2 A-D. Fluorescence in situ hybridization analysis for FGFR1 gene showed gene amplification (A) in a p40-positive large cell carcinoma (B) fulfilling criteria for squamous differentiation lineage (the same tumor also presented TP53 mutation but no other adenocarcinoma-related gene aberrations). By comparison, no FGFR1 amplification (C) was observed in a LCC sample exhibiting null (TTF1-/p40-) phenotype (D, only p40 IHC is shown), which fulfilled criteria for adenocarcinoma differentiation lineage.

Fig 3. Survival analysis according to mutation status. Survival curves for OS (left side) and DFS (right side) according to the number of gene mutations per patient, which were not affected by age, gender and tumor stage (stage I comprised one tumor with no mutation, 10 tumors with 1-2 mutations and one tumor with ≥3 mutations vs. stage II-IV showing 2, 12 and 4 tumors in the relevant categories, respectively, p=0.827).

S1_Fig 1. Phenotypic and genotypic correlations for p53. Phenotypic and genotypic correlations between p53 protein expression and the relevant gene alterations as derived from T-NGS analyses in two cases of lung adenocarcinoma. Upper panel: accumulation of p53 in 100% tumor cells was present in the case #01 bearing inactivating V173A mutation (48% of allelic DNA). Bottom panel: complete lack of the relevant protein was seen in the case #10 bearing G266Stop mutation (38% of allelic DNA).
S2_Fig 2. Survival analysis on the entire tumor series. Overall and recurrence-free survival in the whole series of 30 LCC under evaluation is shown.
Supporting Information

S1_Molecular&IHC method. Molecular and immunohistochemistry method. Details regarding the NGS, fluorescence in situ hybridization and immunohistochemistry procedures set up for the study.

S2_Result File. Additional Immunohistochemistry information. Detailed results are provided on the immunohistochemistry for p63 and cytokeratin 7 in 30 large cell carcinoma patients.

S1_Table 1. Distribution of gene mutations. Data on the distribution of gene mutation and the correlation with p53 protein immunohistochemical expression are provided in the 30 large cell carcinoma patients under evaluation.

S2_Table 2. Next generation sequencing results on 30 LCC patients.

S3_Table 3. Gene function. List of mutated genes found in 27 cases of LCC is shown with functional information.

S1_Fig 1. Phenotypic and genotypic correlations for p53. Phenotypic and genotypic correlations between p53 protein expression and the relevant gene alterations as derived from T-NGS analyses in two cases of lung adenocarcinoma. Upper panel: accumulation of p53 in 100% tumor cells was present in the case #01 bearing inactivating V173A mutation (48% of allelic DNA). Bottom panel: complete lack of the relevant protein was seen in the case #10 bearing G266Stop mutation (38% of allelic DNA). Phenotypic and genotypic correlations for p53.

S2_Fig 2. Survival analysis on the entire tumor series. Overall and recurrence-free survival in the whole series of 30 LCC under evaluation is shown.
References


48. Cardarella S, Ortiz TM, Joshi VA, et al. The introduction of systematic genomic


<table>
<thead>
<tr>
<th>Variable</th>
<th>Patient number</th>
<th>Gender</th>
<th>Smoking habit</th>
<th>Tumor stage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Men</td>
<td>Women</td>
<td>Current</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>30</td>
<td>24</td>
<td>6</td>
<td>15</td>
</tr>
<tr>
<td>0-59</td>
<td>7</td>
<td>7</td>
<td>.</td>
<td>3</td>
</tr>
<tr>
<td>60-69</td>
<td>12</td>
<td>7</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>70+</td>
<td>11</td>
<td>10</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>6</td>
<td>-</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Men</td>
<td>24</td>
<td>24</td>
<td>-</td>
<td>10</td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 1</td>
<td>12</td>
<td>10</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Stage 2</td>
<td>13</td>
<td>12</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Stage 3-4</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>LN status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pN0</td>
<td>22</td>
<td>19</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>pN+</td>
<td>8</td>
<td>5</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Smoking status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current</td>
<td>15</td>
<td>10</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>Former</td>
<td>15</td>
<td>14</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>Tumor size</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;2 cm</td>
<td>4</td>
<td>4</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>2-5 cm</td>
<td>17</td>
<td>14</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>&gt;5 cm</td>
<td>9</td>
<td>6</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Vascular invasion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>13</td>
<td>11</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Yes</td>
<td>17</td>
<td>13</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>Necrosis (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>absent</td>
<td>2</td>
<td>2</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>&lt;20</td>
<td>7</td>
<td>7</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>&gt;20</td>
<td>21</td>
<td>15</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>Pleural invasion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PL0</td>
<td>15</td>
<td>11</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>PL1/2</td>
<td>13</td>
<td>12</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>PL3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Histology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>clear cell</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>rhabdoid cell</td>
<td>3</td>
<td>3</td>
<td>-</td>
<td>9</td>
</tr>
<tr>
<td>LELC</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>undifferentiated</td>
<td>20</td>
<td>16</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

LELC: lymphoepithelioma-like carcinoma
<table>
<thead>
<tr>
<th>Antibodies</th>
<th>Clone</th>
<th>Source</th>
<th>Dilution</th>
<th>Pretreatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytokeratin 7</td>
<td>OV-TL 12/30</td>
<td>DAKO, Glostrup, Denmark</td>
<td>1:200</td>
<td>PTLink-EDTA</td>
</tr>
<tr>
<td>p40</td>
<td>polyclonal</td>
<td>Calbiochem, Darmstadt, Germany</td>
<td>1:3000</td>
<td>PTLink-EDTA</td>
</tr>
<tr>
<td>p40</td>
<td>BC28</td>
<td>Biocare, Concord, CA, USA</td>
<td>1:400</td>
<td>PTLink-EDTA</td>
</tr>
<tr>
<td>p53</td>
<td>DO7</td>
<td>DAKO, Glostrup, Denmark</td>
<td>1:400</td>
<td>PTLink-EDTA</td>
</tr>
<tr>
<td>p63</td>
<td>DAKp63</td>
<td>DAKO, Glostrup, Denmark</td>
<td>1:50</td>
<td>PTLink-EDTA</td>
</tr>
<tr>
<td>Thyroid transcription factor-1</td>
<td>8G7G3/1</td>
<td>DAKO, Glostrup, Denmark</td>
<td>1:2000</td>
<td>PTLink-EDTA</td>
</tr>
<tr>
<td>Synaptophysin</td>
<td>DAK-SYNAP</td>
<td>DAKO, Glostrup, Denmark</td>
<td>1:200</td>
<td>PTLink-EDTA</td>
</tr>
<tr>
<td>Vimentin</td>
<td>V9</td>
<td>DAKO, Glostrup, Denmark</td>
<td>1:50</td>
<td>PTLink-EDTA</td>
</tr>
</tbody>
</table>

PTLink-EDTA: boiling bath for 30 min in EDTA buffer pH8; PTLink-CB: boiling bath for 30 min in citrate buffer pH6
Table 3. Distribution of gene mutation and correlation with p53 protein immunohistochemical expression in the 30 large cell carcinoma patients under evaluation

| Gene/Patients | #01 | #03 | #04 | #07 | #09 | #10 | #13 | #14 | #18 | #19 | #22 | #24 | #27 | #28 | #29 | #30 | #02 | #05 | #06 | #11 | #12 | #15 | #16 | #17 | #20 | #21 | #23 | #25 | #26 |
|---------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Multiple mutations |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| TP53 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| % p53 protein | 100 | 100 | 0   | 100 | 5   | 100 | 0   | 45  | 15  | 100 | 80  | 0   | 100 | 45  | 80  | 10  | 100 | 0   | 0   | 2   | 0   | 40  | 0   | 15  | 0   | 60  | 0   | 0   | 0   | 0   |
| KRAS |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| NRAS |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| ATM |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| BRAF |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| EGFR |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| FBXW7 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| PIK3CA |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| CDKN2A |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| ERBB4 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| FLT3 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| PTPN11 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| RET |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| SMO |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| STK11 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| SMAD4 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Wild Type |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

Samples are displayed and columns and arranged to emphasize mutually exclusive mutations (red boxes). A total of 27 (90%) tumor patients were found to have at least one mutation. Blue boxes stand for wild type samples. In the second row, the distribution of p53 protein expression according to the presence or absence of the relevant mutations is shown.

The patients #19 (exon 6 and 8) and #28 (exon 10) had double mutations for TP53 (see Table 3 for details); red boxes: mutations; blue boxes: wild type tumors.
Table 4. Morphologic, immunohistochemistry and molecular correlations in 30 large cell carcinomas

<table>
<thead>
<tr>
<th>Category</th>
<th>Total</th>
<th>Histologic classification</th>
<th>Number of mutated genes</th>
<th>Wild type</th>
<th>TTF1 % (mean±SD)</th>
<th>p40 % (mean±SD)</th>
<th>IHC phenotype interpretation</th>
<th>Gene alterations *</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTF1+/p40-</td>
<td>6 (9)</td>
<td>LCC-U 5, LCC-CCC 4, - 1, -</td>
<td>12</td>
<td>3</td>
<td>29.2±38.9</td>
<td>-</td>
<td>ADC</td>
<td>ATM, ERBB4, FLT3, KRAS (2), BRAF, NRAS, TP53 (4), FBXW7</td>
</tr>
<tr>
<td>TTF1+/p40±</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>0</td>
<td>98</td>
<td>5</td>
<td>ADC</td>
<td>NRAS</td>
</tr>
<tr>
<td>TTF1-/p40+ **</td>
<td>3</td>
<td>LCC-U 1, LCC-CCC 1, LCC-LELC 1, -</td>
<td>3</td>
<td>0</td>
<td>-</td>
<td>83.3±20.8</td>
<td>SQC</td>
<td>TP53 (3)</td>
</tr>
<tr>
<td>TTF1-/p40-</td>
<td>13</td>
<td>LCC-U 11, LCC-CCC 1, - 1, -</td>
<td>25</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>ADC</td>
<td>ATM (2), BRAF, CDKN2A, EGFR (2), FBXW7, KRAS (4), PIK3CA (2), PTPN11, RET, STK11, TP53 (9)</td>
</tr>
<tr>
<td>TTF1-/p40±</td>
<td>4</td>
<td>-</td>
<td>1</td>
<td>6</td>
<td>-</td>
<td>4.2±1.9</td>
<td>ADC</td>
<td>KRAS, SMAD4, SMO, TP53 (3)</td>
</tr>
</tbody>
</table>

No cases with TTF1+/p40+ phenotype were seen. *: in brackets the number of multiple mutated genes; #: three patients with TTF1+/p40- phenotype were wild type for the gene mutations being assessed; LCC-U: large cell carcinoma-undifferentiated type; LCC-CCC: large cell carcinoma - clear cell carcinoma type; LCC-LELC: large cell carcinoma - lymphoepithelioma-like carcinoma type; LCC-RC: large cell carcinoma - rhabdoid cell type; ADC: adenocarcinoma; SQC: squamous cell carcinoma; IHC: immunohistochemistry; SD: standard deviation

** These tumors also presented with FGFR1 gene amplification by fluorescence in situ hybridization analysis in two out of three cases
Case #04, LCC-CCC
Category: TTF1+/p40-
Mutation: KRAS, FLT3, TP53
Mutations in the same category: ATM, ERBB4, FLT3, KRAS, BRAF, NRAS, TP53, FBXW7

Case #05, LCC-RC
Category: TTF1+/p40±
Mutation: NRAS
Mutations in the same category: NRAS

Case #07, LCC-LELC
Category: TTF1-/p40+
Mutation: TP53
Mutations in the same category: TP53

Case #10, LCC-U
Category: TTF1-/p40-
Mutation: CDKN2A, TP53
Mutations in the same category: ATM, BRAF, CDKN2A, EGFR, FBXW7, KRAS, PIK3CA, PTPN11, RET, STK11, TP53

Case #24, LCC-U
Category: TTF1-/p40±
Mutation: TP53
Mutations in the same category: KRAS, SMAD4, SMO, TP53