Journal of Pathology
J Pathol November 2022; **258:** 250–263
Published online 23 September 2022 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/path.5996

ORIGINAL ARTICLE

Disseminated cancer cells detected by immunocytology in lymph nodes of NSCLC patients are highly prognostic and undergo parallel molecular evolution

Felix Elsner[1,2,3], Martin Hoffmann[4], Rezan Fahrioglu-Yamaci[1], Zbigniew Czyz[1], Giancarlo Feliciello[4], Tobias Mederer[1], Bernhard Polzer[4], Steffi Treitschke[4], Petra Rümmele[2,3], Florian Weber[2], Hellmuth Wiesinger[5], Tobias Robold[6], Zsolt Sziklavari[7,8], Wulf Sienel[9], Hans-Stefan Hofmann[6,7] and Christoph A Klein[1,4]*

[1] *Experimental Medicine and Therapy Research, University of Regensburg, Regensburg, Germany*
[2] *Institute of Pathology, University of Regensburg, Regensburg, Germany*
[3] *Institute of Pathology, University Hospital Erlangen, Erlangen, Germany*
[4] *Division of Personalized Tumour Therapy, Fraunhofer ITEM-R, Regensburg, Germany*
[5] *Gemeinschaftspraxis für Pathologie Regensburg, Regensburg, Germany*
[6] *Department of Thoracic Surgery, University Hospital Regensburg, Regensburg, Germany*
[7] *Department of Thoracic Surgery, Krankenhaus Barmherzige Brüder Regensburg, Regensburg, Germany*
[8] *Department of Thoracic Surgery, Klinikum Coburg, Coburg, Germany*
[9] *Department of Thoracic Surgery, University of Munich, Munich, Germany*

**Correspondence to: CA Klein, Experimental Medicine and Therapy Research, University of Regensburg, Franz-Josef-Strauss-Allee 11, 93053, Regensburg, Germany. E-mail: christoph.klein@ukr.de.*

Abstract

In melanoma, immunocytology (IC) after sentinel lymph node disaggregation not only enables better quantification of disseminated cancer cells (DCCs) than routine histopathology (HP) but also provides a unique opportunity to detect, isolate, and analyse these earliest harbingers of metachronous metastasis. Here, we explored lymph node IC in non-small cell lung cancer (NSCLC). For 122 NSCLC patients, 220 lymph nodes (LNs) were split in half and prepared for IC and HP. When both methods were compared, IC identified 22% positive patients as opposed to 4.5% by HP, revealing a much higher sensitivity of IC ($p < 0.001$). Assessment of all available 2,952 LNs of the same patients by HP uncovered additional patients escaping detection of lymphatic tumour spread by IC alone, consistent with the concept of skip metastasis. A combined lymph node status of IC and complete HP on a larger cohort of patients outperformed all risk factors in multivariable analysis for prognosis ($p < 0.001$; RR = 2.290; CI 1.407–3.728). Moreover, isolation of DCCs and single-cell molecular characterization revealed that (1) LN-DCCs differ from primary tumours in terms of copy number alterations and selected mutations and (2) critical alterations are acquired during colony formation within LNs. We conclude that LN-IC in NSCLC patients when combined with HP improves diagnostic precision, has the potential to reduce total workload, and facilitates molecular characterization of lymphatically spread cancer cells, which may become key for the selection and development of novel systemic therapies.

Keywords: early metastasis; disseminated cancer cell; prognosis; molecular evolution of NSCLC

Received 29 April 2022; Revised 20 July 2022; Accepted 28 July 2022

No conflicts of interest were declared.

Introduction

Clinical data for non-small cell lung cancer (NSCLC) suggest that aggressive cancer cells disseminate often in early stages (UICC I, i.e. T≤4 cm, N0, M0) [1–3] and can be found either as circulating tumour cells (CTCs) within the blood stream or as disseminated cancer cells (DCCs) within distant tissues, where they first form micrometastases and later grow out into macroscopically overt distant metastases [4]. Since the 1990s, several studies have been conducted focusing on the detection of occult cancer spread in the lymph nodes (LNs) of NSCLC patients. Immunohistochemistry (IHC) studies, applied either on cryostat sections or on sections from formalin-fixed, paraffin-embedded (FFPE) tissue to detect lymphatic spread [5–14], have suggested that NSCLC patients with occult lymphatic cancer spread experience a poor outcome, consistent with the known impact of LN status in the TNM staging system.

However, the lymphatic spread of DCCs in NSCLC has not been quantified in absolute terms. Radical LN dissection combined with histopathology (HP) is the

recommended standard procedure for operable NSCLC that results in the most accurate nodal staging [15,16], superior to non-invasive, imaging-based staging procedures. Although international standards have not been defined, national guidelines recommend HP screens of all available LNs of a surgical specimen and at least two cutting levels for all macroscopically tumour-free LNs [17]. In contrast, quantitative immunocytology (IC) after disaggregation of sentinel LNs (SLNs) was shown to be one of the most important prognostic factors for predicting outcome in patients with malignant melanoma, outperforming standard HP plus IHC [18]. The molecular analysis of SLN-DCCs of melanoma patients and of their corresponding primary tumours (PTs) revealed that dissemination to the SLN occurs early at a median thickness of the PT of about 0.5 mm. Moreover, driver changes crucial for colony formation and metastatic outgrowth were acquired within the LN [19] showing transition to LN metastasis detectable by HP at a DCC density (DCCD) of approximately 100.

We therefore explored the suitability of IC in NSCLC patients. Unlike melanoma, SLNs are unavailable in NSCLC and the occurrence of skip metastases often necessitates the analysis of several dozens of regional LNs. We therefore asked whether the analysis of 2×10^6 LN cells in total from the first and second LN station improves outcome prediction compared with standard HP. We then examined the genomes of early and late DCCs to obtain a first glimpse into the evolutionary dynamics of lymphatic metastasis in NSCLC and to clarify whether DCCs acquire their malignant traits within the PT or in parallel [4,20–22].

Materials and methods

Patients

Between March 2011 and November 2018, we collected LN samples from 263 consecutive patients who were diagnosed with NSCLC and underwent curative surgery ($n = 261$) or diagnostic mediastinoscopy ($n = 2$) at the Department of Thoracic Surgery either at the University Hospital Regensburg or at the Krankenhaus der Barmherzigen Brüder Regensburg. The study was approved by the local Ethics Committee (ethics vote number 07-079) and all patients gave written informed consent before entering the study. The baseline characteristics of the complete patient cohort and a subset for comparative LN sampling are summarized in Table 1.

Table 1. Characteristics of the total patient cohort and subset.

	$n = 263$		$n = 122$	
Parameter	*n*	%	*n*	%
Sex				
Female	79	30.0	36	30.0
Male	184	70.0	86	70.0
Histology				
ADC	146	55.5	66	54.1
SCC	105	39.9	50	41.0
ADS	5	1.9	4	1.6
LCC	6	2.3	2	3.3
Rare subtypes	1	0.4	0	0
T category*,†				
pT1	100	38.0	49	40.2
pT2	71	27.0	30	24.6
pT3	42	16.0	18	14.8
pT4	30	11.4	9	7.3
Not determined	20	7.6	16	13.1
Tumour size (mm)‡				
≤30			57	46.7
30 < *x* ≤ 70			39	32.0
>70			8	6.6
Not determined			18	14.8
N category*,§				
pN0	163	62.0	87	71.3
pN1	55	20.9	21	17.2
pN2	44	16.7	13	10.2
pN3	1	0.4	1	0.8
UICC stage*,‖				
I	118	44.9	66	54.1
II	61	23.2	27	22.1
III	72	27.4	21	17.2
IV	12	4.6	8	6.6
Residual tumour				
R0	247	93.9	114	93.4
R1	10	3.8	6	4.9
R2	4	1.5	0	0
No resection	2	0.8	2	1.6
Neoadjuvant treatment				
None	245	93.2	108	88.5
Neoadjuvant therapy	18	6.8	14	11.5
Grading				
G1	5	1.9	2	1.6
G2	126	47.9	49	40.2
G3	132	50.2	71	58.2

ADC, adenocarcinoma; ADS, adenosquamous carcinoma; LCC, large-cell carcinoma; SCC, squamous cell carcinoma.
*According to TNM Classification of Malignant Tumours, 8th edn, 2016 [3].
†pT status not available for patients with neoadjuvant treatment and patients without resection of primary tumour.
‡Measured after formalin fixation; patients with multiple tumour nodules in the same lobe of the lung are excluded; only determined for the subset of 122 patients.
§According to histopathological examination only.
‖cTNM data were used for determination of UICC stage in cases of missing pTNM data.

Comparative LN sample processing

We aimed to obtain half an N1-LN and half an N2-LN from each patient for IC analysis. For the total cohort, 249 N1-LNs and 240 N2-LNs were bisected by the surgeon directly after removal, resulting in 489 half LNs that were analysed by IC. In a subset of 122 patients, comparative LN sample processing was performed, i.e. the corresponding half of the LN could be followed up through HP processing so that IC and HP results on the same LN could be directly compared. Altogether, 2,952 LNs from 122 patients were surgically removed (supplementary material, Figure S1). Half a lymph node was formalin-fixed and subjected to routine HP; the other half was disaggregated into single cells as previously described [18,19,23]. IC staining was performed essentially as described by Passlick *et al* [5] using the monoclonal mouse antibody clone Ber-EP4 (Dako, Hamburg, Germany) directed against EpCAM/

CD326 [24] (see Supplementary materials and methods for details). Detection criteria for DCCs were strong membranous staining for Ber-EP4 and typical cytomorphologic and phenotypic features as reported by Fehm *et al* [25]. LNs containing at least one Ber-EP4-positive cell were considered as positive. The staining result was reported as DCCD, i.e. the number of Ber-EP4-positive cells per 10^6 examined mononuclear cells. Ber-EP4-positive single cells were isolated using a micromanipulator and submitted to whole genome amplification (WGA). WGA was performed as previously described [26,27].

Histopathology

Of all surgically removed LNs, at least one H&E-stained slide with up to three cutting levels was examined by an experienced pathologist (FW or HW). For eight IC positive LNs, step sectioning of the corresponding paraffin-embedded half was performed along with IHC (see Supplementary materials and methods for details).

DNA extraction from FFPE tissue

Genomic DNA from FFPE tissue of PTs was isolated and purified using the QIAamp DNA FFPE Tissue Kit (Qiagen, Hilden, Germany) (see Supplementary materials and methods for details).

Single-cell array comparative genome hybridization

Array comparative genomic hybridization (aCGH) analysis of single cells and FFPE samples was performed as previously described [28] (see Supplementary materials and methods for details). Chromosomal CGH was performed as published elsewhere [27], with modifications described in ref 26.

Mutation analysis

Mutations in *EGFR* exons 18–21, *HER2* exon 20, *KRAS* exon 2, *PIK3CA* exons 9 and 20, and *TP53* exons 5–8 were detected using Sanger sequencing (see Supplementary materials and methods).

Statistical analysis

We used SPSS version 28 (SPSS Inc., Chicago, IL, USA) for all statistical analyses (see Supplementary materials and methods for details).

Results

Relationship of IC positivity and DCCD to different clinicopathologic parameters

We used quantitative IC to examine 220 LN halves from 122 different patients. Ber-EP4-positive cells (Figure 1A) were detected in 48 out of 220 LNs. The LNs examined had DCCD values ranging from 0 to 750,000. The majority of LNs positive in IC displayed DCCD values less than or equal to 10 (Figure 1B), with a median DCCD of 3.0. The detection rate of DCCs only marginally increased from 28.1% in patients with PTs less than 30 mm to 37.5% in those with PTs more than 70 mm (Figure 1C). No significant correlation between the number of LN-DCCs and PT size was observed (Figure 1D; $p = 0.404$; $r = 0.061$, Spearman's correlation). However, the detection of LN metastases by HP (p+) was associated with significantly higher DCCD values when compared with metastasis-free (p−) LNs (median 17,500 versus 2; $p < 0.001$; Mann–Whitney *U*-test) (Figure 1E). The only significant association between any clinicopathologic variable and the presence of DCCs in the LN was noted for L status (supplementary material, Table S2), which is a known surrogate parameter for metastatic LN involvement in NSCLC. For all other variables, no significant correlation was found. We specifically tested patient sex, age, histologic subtype, T status, PT size, N status, UICC stage, R status, and histopathological grading (Table 2 and supplementary material, Figure S3).

Comparison of IC and routine HP

The results of IC and routine HP were directly compared for corresponding LN halves analysed by both methods and samples were grouped into four categories: LNs positive in both IC and HP (i+/p+), LNs positive in IC and negative in HP (i+/p−), LNs negative in both IC and HP (i−/p−), and LNs negative in IC despite histopathological detection of an LN metastasis (i−/p+) (Table 3). DCCs detected in LNs of the i+/p+ category were termed i+/p+ DCCs and considered representative of metastatic colonies as evidenced in H&E sections. On the other hand, DCCs originating from i+/p− LNs were categorized as i+/p− DCCs that mostly had not yet formed metastatic colonies.

IC showed higher detection rates than routine HP. LNs positive in IC made up 22% of samples, whereas only 4.5% showed LN metastases in routine HP, all of which were also detected by IC. Both LN halves were negative in 78.2% of cases. Moreover, in 17.3% of LNs, only IC detected tumour cells, whereas in no case did histology detect tumour cells that were not detected by IC ($p < 0.0001$; χ^2 test), resulting in a five-fold higher detection rate of IC compared with routine HP. When evaluated for the 122 patients instead of the 220 LNs (i.e. a patient was called 'positive' if at least one sample was positive), the results were very similar (Table 3).

Since the histopathological routine diagnosis was based only on one H&E-stained slide per LN, we asked whether a more accurate examination of the paraffin-embedded LN tissue would increase histopathological detection of minimal lymphatic spread and thereby support the IC results. Thus, eight LNs positive in IC but negative in routine HP (i+/p−) showing low DCCD values (mean 4.7; range 0.5–12) were randomly selected and thoroughly re-examined by step sectioning and IHC

for Ber-EP4 and PanCK. In three of these eight cases, we identified small tumour nests (Figure 1F). Ber-EP4-positive single cells and small aggregates were found in one out of eight LNs and three out of eight LNs harboured PanCK-positive cells (supplementary material, Table S3). Morphologically, the

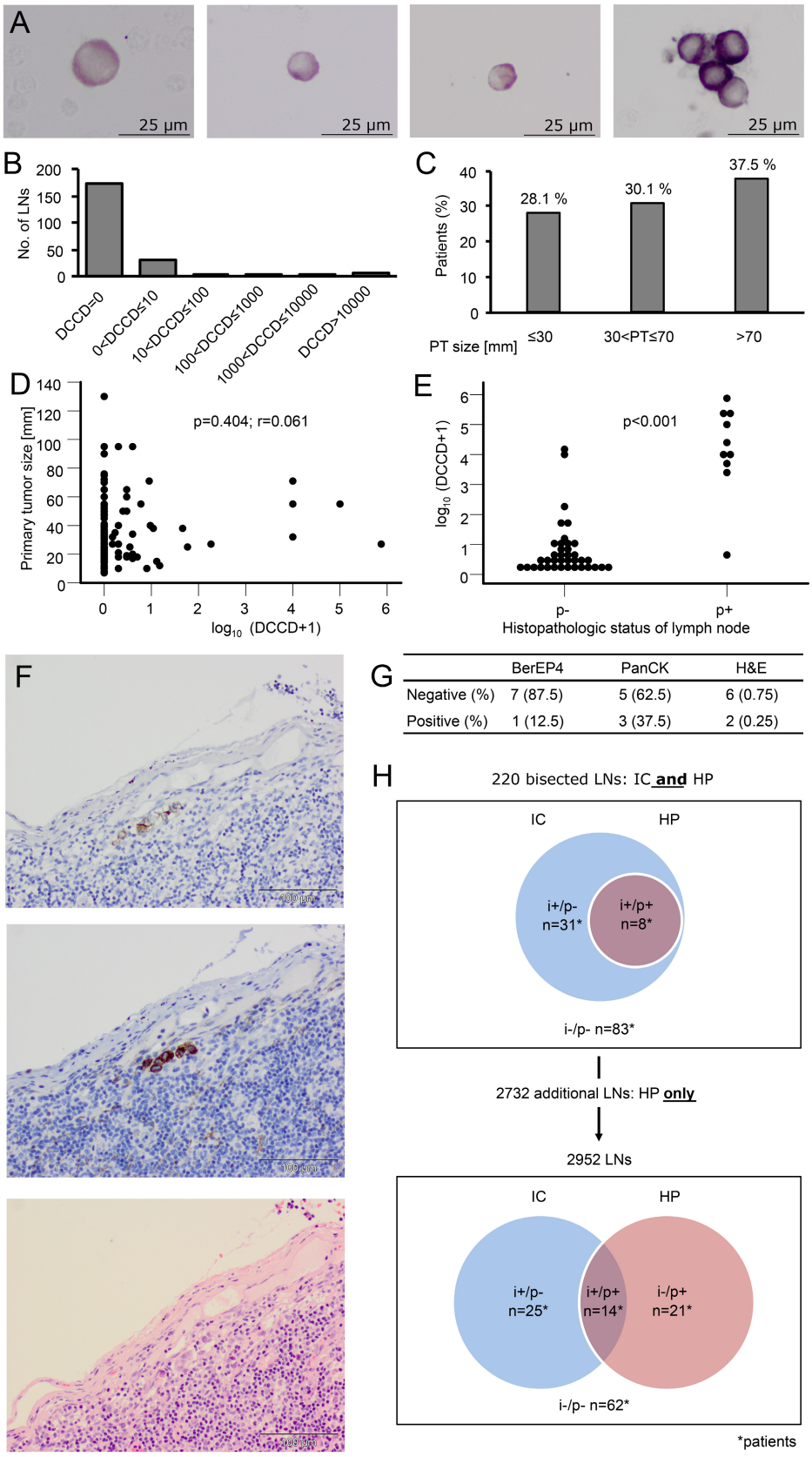

Figure 1 Legend on next page.

Table 2. Associations between LN-DCCs and clinicopathological variables for 122 patients (if values were missing, the correct number is given in parentheses).

Parameter	Patients		p value
	n	%	
Sex			
Male	86	70.5	0.289
Female	36	29.5	
Age at surgery (years)			
≤67	63	51.6	0.658
>67	59	48.4	
Histology ($n = 116$)*			
ADC	66	56.9	0.433
SCC	50	43.1	
T category† ($n = 104$)‡			
pT1/2	79	76.0	0.437
pT3/4	25	24.0	
Tumour size (mm) ($n = 104$)‡			
≤30	57	54.8	0.670
>30	47	45.2	
N category†			
pN0	87	71.3	0.228
pN1/2/3	35	28.7	
UICC stage ($n = 104$)‡			
I/II	85	81.7	0.852
III/IV	19	18.3	
Residual tumour			
R0	112	91.8	0.202
R1/no surgery	10	8.2	
Grading			
G1/2	51	41.8	0.608
G3	71	58.2	

*χ^2 was only tested for ADC versus SCC.
†According to TNM Classification of Malignant Tumours, 8th edn, 2016 [3].
‡Patients with neoadjuvant treatment and/or multiple tumour nodules in the same lobe of the lung are excluded.

immunohistochemically detected cells had an increased nuclear to cytoplasmic ratio; were larger than surrounding lymphocytes; and did not bear any morphologic resemblance to reticulocytes. Of note, the cells were located in the subcapsular LN sinus, which is known as the predilection site where metastatic cancer cells first arrive and lodge in the LN before outgrowth. Furthermore, cells with enlarged nuclei and prominent eosinophilic nucleoli could also be identified in corresponding locations on matched H&E sections in two cases (Figure 1F,G).

However, despite thorough re-examination using step sections and IHC, five of eight LNs (62.5%) remained negative in HP. Although we cannot exclude a sampling error when splitting the LN, the significantly higher DCCD values observed for immunohistochemical positive compared with negative LNs [mean 8.3 versus 2.5, respectively; $p = 0.037$ (Student's *t*-test)] suggest that IC is also superior to extended histopathological analysis by IHC even when only small numbers of lymphatically spread cancer cells are present below a DCCD of about 8. We did not observe that IC missed cancer cells, whereas HP clearly did.

We then tested the need for a complete histopathological analysis in addition to IC based on two LNs alone and included all resected LN samples and applied routine HP to the additional 2,732 LNs, for which no IC had been performed. This revealed LN metastases in 21 out of 83 patients categorized before as i−/p− and in 6 out of 31 patients classified before as i+/p− (Figure 1H). In summary, IC is much more sensitive than HP performed on a single H&E-stained slide with up to three cutting levels. However, a ten-fold increase in the number of histologically analysed LNs identifies few additional patients that are missed when only one or two LNs are analysed by IC.

Survival impact of LN-DCCs

To assess the clinical relevance of DCCs detected by IC in LNs, we performed survival analysis in the total cohort of NSCLC patients. We excluded patients with a history of neoadjuvant therapy without (complete) surgical resection of the PT or a follow-up time less than 6 months to assess the impact of systemic cancer spread to LNs. For the remaining 217 patients, altogether 407 LNs were analysed by IC. Of these, 101 patients and 127 LNs tested positive for DCCs, whereas 116 patients were classified as DCC-negative. At a median follow-up time of 64.5 months (range 8.5–125.1 months), 87 patients had died. We used log-rank testing for univariable analysis. Dissemination of single tumour cells to LNs indicated a significantly shorter overall survival (OS) for NSCLC patients and was slightly more prognostic than pN status ($p = 0.002$ versus $p = 0.003$) (Figure 2A,B). IC relevance could be

Figure 1. Immunocytochemistry for Ber-EP4 using the APAAP method. (A) Ber-EP4$^+$ single cells and a small aggregate of Ber-EP4$^+$ cells (rightmost). The size of Ber-EP4+ cells is markedly increased compared with surrounding lymphocytes. Nevertheless, Ber-EP4$^+$ cells show a certain variability in cell size. The single cells shown in the two leftmost pictures harboured copy number alterations in their aCGH profiles, whereas the third single cell was balanced. Apparently, it is not possible to identify cells with aberrant aCGH profiles by using morphological criteria. (B) Distribution of DCCD values assigned to different ranges for the 220 examined LNs. (C) Percentage of patients ($n = 104$) with LN-DCCs for PT sizes ≤30, 30 < PT size ≤ 70, and >70 mm. Patients with a history of neoadjuvant therapy, multiple PTs in the lung, or without surgical resection of the PT were excluded. (D) Graphical representation of the relationship between PT size and $\log_{10}$(DCCD + 1) for 190 LNs from the 104 patients. (E) Distribution of $\log_{10}$(DCCD + 1) in LNs without metastases (p−; $n = 38$ LNs) and LNs harbouring metastases (p+; $n = 10$ LNs). (F) Re-examination of i+/p− LNs by step sectioning and IC. Small DCCs aggregate in the subcapsular sinus of a lymph node (top). From the top to bottom: Ber-EP4, PanCK (AE1/AE3), and H&E staining. Aggregated cells show enlarged nuclei and prominent eosinophilic nucleoli and can be identified in corresponding locations. (G) Summary of the results for re-examination of i+/p− LNs ($n = 8$). (H) 'IC diagnostic gap' caused by the limited number of IC-examined LNs. On the basis of bisected LNs for which the corresponding halves were examined by both IC and HP, patients could be grouped into three different categories: i+/p−, i+/p+ and i−/p−. When the 2,732 additional LNs examined exclusively by HP were included, some patients were also categorized as i−/p+. Total number of patients $n = 122$.

Table 3. Immunocytology of disaggregated lymph node samples and routine histopathological work-up related to lymph node samples, patients when only bisected LNs were considered, and patients when all 2,952 histologically examined LNs were considered.

Lymph node samples			
Immunocytology	Histopathology		Total (%)
	No. of negative LNs (p−)	No. of positive LNs (p+)	
Negative (i−)	172	0	172 (78.2)
Positive (i+)	38	10	48 (21.8)
Total (%)	210 (95.5)	10 (4.5)	220 (100)
Patients (only bisected LNs considered)			
Immunocytology	Histopathology		Total (%)
	No. of negative patients (p−)	No. of positive patients (p+)	
Negative (i−)	83	0	83 (68.0)
Positive (i+)	31	8	39 (32.0)
Total (%)	114 (93.4)	8 (6.6)	122 (100)
Patients (all 2,952 histologically examined LNs considered)			
Immunocytology	Histopathology		Total (%)
	No. of negative patients (p−)	No. of positive patients (p+)	
Negative (i−)	62	21	83 (68.0)
Positive (i+)	25	14	39 (32.0)
Total (%)	87 (71.3)	35 (28.7)	122 (100)

further improved when we limited the analysis to patients who had two LNs analysed (one N1-LN and one N2-LN, respectively; $n = 190$; $p < 0.001$; Figure 2C). This analysis also revealed a significant quantitative effect of DCCD on OS when categorized to DCCD = 0, 0 < DCCD ≤ 100, and DCCD > 100 (Figure 2D). We then quantified risk factors in uni- and multi-variable analyses. Of the prognostically significant associated risk factors in univariable analysis (DCCD, pN status, pT status, age, and sex), only DCCD, pN status, and sex were included in the final model, with DCCD being the most significant predictor of outcome (Table 4).

The combined inclusion of DCCD and pN status in the Cox model suggested that pN status provided risk information in addition to DCCD (when limited to two LN halves), fully consistent with the observation that a ten-fold increase in the number of analysed LNs identifies additional patients at risk (Figure 1H). We therefore combined the results of IC and HP, i.e. a patient was counted as node-positive (N+) if at least one of the two methods detected tumour cells in at least one LN. This combined nodal status (N_{cHPIC}) resulted in the best overall stratification of subgroups N0 versus N1 ($p = 0.049$), N0 versus N2 ($p < 0.001$), and N1 versus N2 ($p = 0.079$; Figure 2E), and was superior to the survival curves generated separately by IC or HP alone (Figure 2F,G). When included in a multivariable Cox regression model in place of DCCD and pN status, the 'combined N status' (N_{cHPIC}) outperformed all other parameters as the most significant prognosticator for OS ($p < 0.001$, RR = 2.290, CI 1.407–3.728) (Table 4).

Genomic analysis of LN-DCCs

Apparently, early LN spread often escapes histopathological detection until the first colonies have been formed and tumour cells can faithfully be detected by routine HP analysis. To test for their malignant origin, molecular analysis of Ber-EP4 positive single cells was performed after micromanipulator-assisted isolation from LNs positive in HP (p+) and negative in HP (p−). Since HP performed on a few sections detects larger colonies more easily than single cells, we considered i+/p+ DCCs as representative for metastatic colonies as opposed to i+/p− DCCs isolated from LNs that were only positive in IC and more representative for DCCs before metastatic colony formation.

In a subset of 19 patients, 39 Ber-EP4-positive single cells (23 i+/p− DCCs and 16 i+/p+ DCCs) displaying high-quality genomic DNA were subjected to aCGH analysis. Cells displayed chromosomal alterations in 82% of cases and 18% had balanced profiles (i.e. no copy number variations detectable by aCGH; for details, see supplementary material, Table S4). Interestingly, we noted that all i+/p+ DCCs were aberrant without exception. Aberrant DCCs were found to display various quantities of copy number variations ranging from 2 to 41 per cell (supplementary material, Figure S4; median = 25.5; interquartile range = 12), with a median of 13.5 genomic gains (range 0–31) and 10.5 losses (range 0–22) per cell. For 14 patients, we could additionally extract DNA from the corresponding PT and compare the aCGH profiles of both DCCs and PT. Notably, DCCs from LNs showed both shared and divergent chromosomal alterations compared with the corresponding PTs (Figure 3A and supplementary material, Figure S5).

Category i+/p+ DCCs differed significantly from i+/p− DCCs with regard to quality and quantity of alterations [median number of alterations 27.5 versus 17.5, respectively; $p < 0.001$ (Mann–Whitney *U*-test)], with genomic gains, unlike chromosomal losses, being significantly more frequent in i+/p+ DCCs [17 versus 2.5; $p < 0.001$ (Mann–Whitney *U*-test); Figure 3B,C], suggesting continuing molecular evolution during colony formation within the LN. Of note, all significantly differing changes, gains or losses, were increased in i+/p+ DCCs and never more abundant in i+/p− DCCs (Table 5).

Targeted sequencing analysis of LN-DCCs

We then performed targeted Sanger sequencing of genes known to be frequently mutated in NSCLC. This included the mutational hot spots in *EGFR* exons 18–21, *HER2* exon 20, *KRAS* exon 2, *PIK3CA* exons 9 and 20, and *TP53* exons 5–8, which were tested in

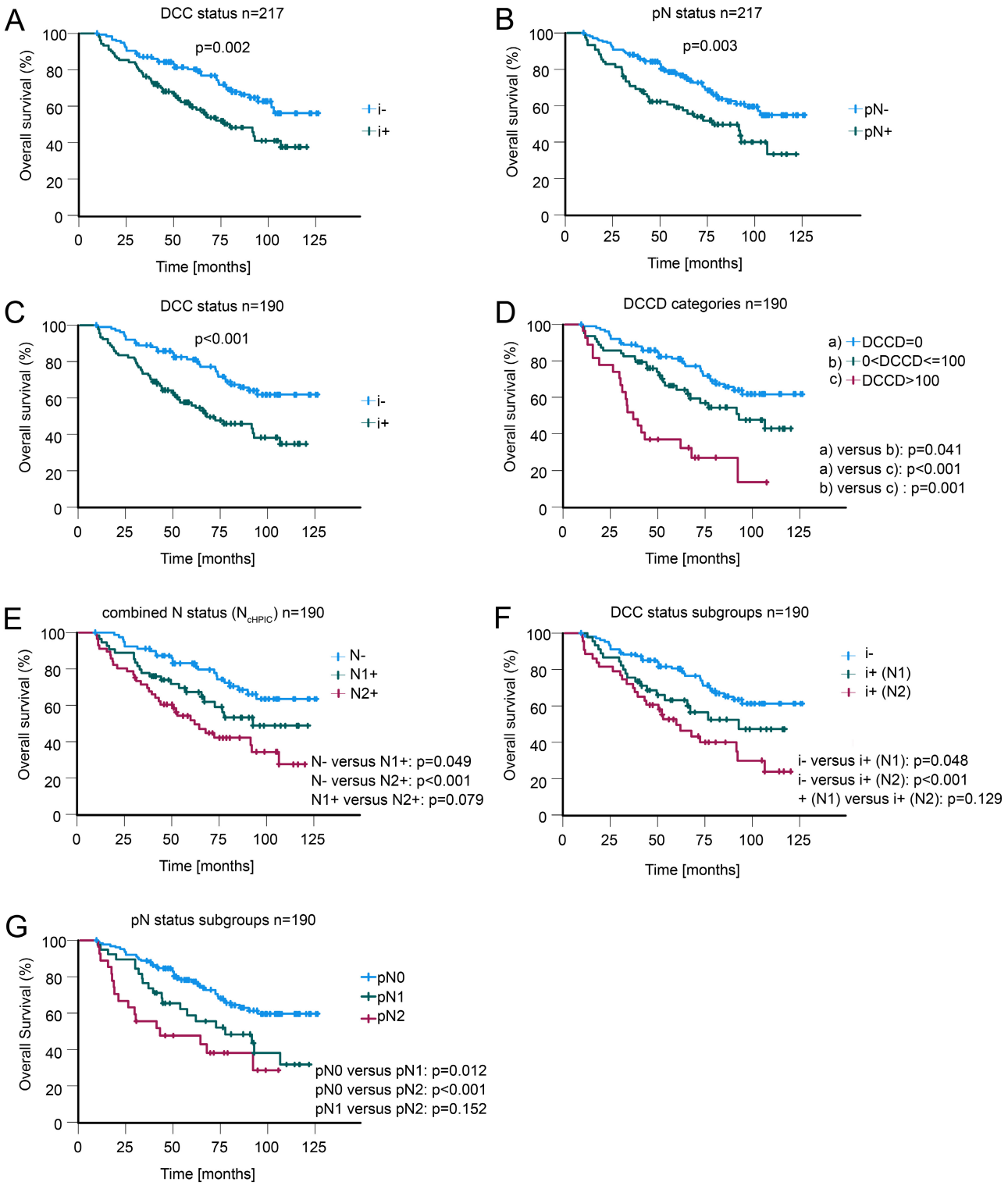

Figure 2. Survival impact of DCCs and pN status. All *P* values were calculated using the log-rank test. Kaplan–Meier curves based on (A) DCC status and (B) conventional pN status, respectively ($n = 217$). (C) The impact of DCC status on OS becomes more evident if only patients with two IC-examined LNs are considered ($n = 190$). (D) Survival impact of different quantitative DCCD ranges [(a) DCCD = 0, (b) 0 < DCCD ≤ 100, (c) DCCD > 100] for patients with two IC-examined LNs ($n = 190$). (E) Combined N status from IC and HP (N_{cHPIC}). (F) Prognostic information from N1 and N2 LNs as detected by IC (N− = negative; N1+: lymphatic spread in N1-LN; N2+: lymphatic spread in N2-LN). (G) Prognostic information from N1 and N2 LNs as detected by HP ($n = 190$, as in D, E, and F).

54 Ber-EP4-positive single cells (including those 39 cells examined by aCGH). Of the analysed cells, 48% harboured mutations, with 7% of these cells showing more than one mutation (Figure 4A). Cells with aberrant aCGH profiles harboured mutations in 71% of cases. None of the seven cells with a balanced aCGH profile showed any mutation in the tested genes. For 20 patients, we analysed DNA extracted from the matched PT for the same mutational hot spots as the Ber-EP4-positive single cells. The paired PT–DCC samples showed a marked

Table 4. Cox regression analysis for overall survival in patients with two LNs analysed by IC ($n = 190$).

Including DCC status and pN status	Univariable analysis*	Multivariable analysis†	
	P value	*P* value	Relative risk (CI)
DCC status (i− versus i+)	<0.001	0.006	1.952 (1.213–3.141)
pN status (pN− versus pN+)	<0.001	0.011	1.826 (1.148–2.904)
Sex (male versus female)	0.007	0.012	1.970 (1.161–3.344)
Age (≤66 years versus >66 years)	0.016		n.i.‡
pT status (pT1/2 versus pT3/4)	0.121		n.i.
Histology (ADC versus SCC)	0.814		n.i.
Grading (G1/2 versus G3/4)	0.662		n.i.

Including combined N status of HP and IC (N_{cHPIC})	Multivariable analysis†	
	P value	Relative risk (95% CI)
N_{cHPIC} (− versus +)	<0.001	2.290 (1.407–3.728)
Sex (male versus female)	0.007	2.072 (1.223–3.510)
Age (≤66 years versus >66 years)		n.i.
pT status (pT1/2 versus pT3/4)		n.i.
Histology (ADC versus SCC)		n.i.
Grading (G1/2 versus G3/4)		n.i.

*Univariable *P* values were calculated using the log-rank test.
†Risk factor selection for multivariable analysis was performed by applying the forward and the backward likelihood ratio (LR) algorithm in SPSS on all variables with $p \leq 0.05$ in univariable analysis, which both delivered exactly the same result.
‡Not included in the final Cox model.
n.i., not informative.

disparity in 70% of patients (Figure 4B). Only 10% of patients had identical mutational status of DCCs and matched PTs simultaneously for all mutations tested and 20% were simultaneously wild type in matched PTs and DCCs. Category i+/p+ DCCs displayed more genomic alterations in aCGH analysis and also harboured more mutations than i+/p− DCCs [94% versus 27%, respectively; $p < 0.001$ (χ^2 test); Figure 4C,D], indicating acquisition of mutations at the distant site during colony formation and growth. A detailed list of all the detected mutations is provided in supplementary material, Figure S6.

Acquisition of genetic changes outside the primary tumour

The striking difference between i+/p− DCCs and i+/p+ DCCs could be caused by either (1) different waves of early and late DCCs or (2) acquisition of genetic changes within the LN. In the first scenario, mutations generated late in the primary tumour would be linked to larger diameters of the primary tumour, whereas in the second scenario their detection in DCCs would be independent of the tumour diameter but associated with cell divisions within the LN. Thus, for each alteration, we tested whether its acquisition is linked to cell doublings within the PT as reflected by tumour diameter (PT size) or linked to cell divisions within the LN as reflected by DCCD. Genetic alterations can then be grouped into four categories: category 1, with exclusive dependence on the number of cell divisions before

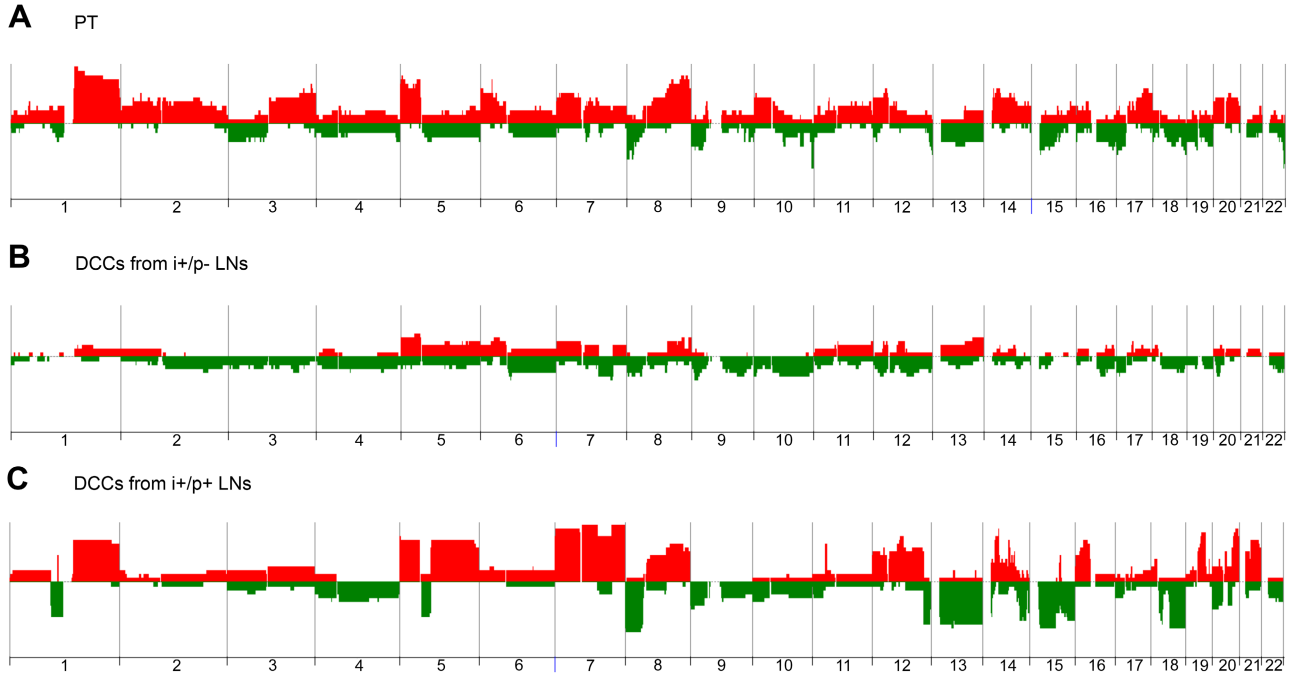

Figure 3. Array comparative genomic hybridization of isolated Ber-EP4$^+$ DCCs and the corresponding PTs. (A) Penetrance plot of PTs ($n = 14$) corresponding to DCCs. The histogram shows the relative proportion of cells with specific chromosomal alterations. Cumulated gains are depicted in red and losses in green. The chromosomal position is indicated by chromosome numbers along the *x*-axis. Gonosomes were excluded. (B) Penetrance plot of all aberrant i+/p− DCCs ($n = 16$). (C) Penetrance plot of all i+/p+ DCCs ($n = 16$).

Table 5. The 13 most divergent genomic regions (as quantified by FDR-adjusted *P* values) between i+/p− DCCs and i+/p+ DCCs in aCGH analysis. All samples were chemotherapy-naive.

Locus	19q		7q22.1–7q32.2		21q		20q13.13–20q13.33		7q11.22–7q21.3		7q32.3–7q36.3		14q11.1–14q24.3	
Adj *P* value	0.0002		0.0002		0.0002		0.0006		0.0017		0.0017		0.018	
DCC type	i+/p−	i+/p+	i+/p−	i+/p+	i+/p−	i+/p+	i+/p−	i+/p+	i+/p−	i+/p+	i+/p−	i+/p+	i+/p−	i+/p+
Gains (%)	0	81	0	75	12	69	12	88	25	94	25	94	19	69
Losses (%)	19	0	38	19	0	25	0	0	25	0	12	0	12	19
Locus	**11p13–11p11.11**		**16p**		**7p**		**1p21.3–1p11.1**		**18q**		**12q11–12q15**			
Adj *P* value	**0.023**		**0.024**		**0.025**		**0.025**		**0.025**		**0.034**			
DCC type	i+/p−	i+/p+	i+/p−	i+/p+	i+/p−	i+/p+	i+/p−	i+/p+	i+/p−	i+/p+	i+/p−	i+/p+		
Gains (%)	12	69	12	69	31	88	12	0	6	6	25	75		
Losses (%)	12	6	19	6	19	0	6	56	25	75	25	0		

dissemination; category 2, with dependence on the total number of divisions either in the primary or at the distant site; and categories 3 and 4, alterations that display a greater dependence on cell divisions in the LN and result in a real and observable dependence on the number of divisions within the LN, i.e. the DCCD. Using this approach, we previously demonstrated in melanoma that decisive genetic alterations are acquired within the SLN [19].

To expand the number of samples for this analysis, we included additional chromosomal CGH data from single cells of a previous study (supplementary material, Table S5), for which all information of the primary tumour and the DCCD were equally available. As in melanoma, most alterations of NSCLC-DCCs were acquired within the LN (Figure 5A–D and supplementary material, Table S6), whereas only a minority had obviously been acquired as the primary tumours grew larger. We further noted a remarkable overlap ($p = 0.033$, Fisher's exact test) between DCCD-associated alterations (Figure 5E) and those that had been previously found to differ between i+/p− DCCs and i+/p+ DCCs (Table 5). Although higher numbers

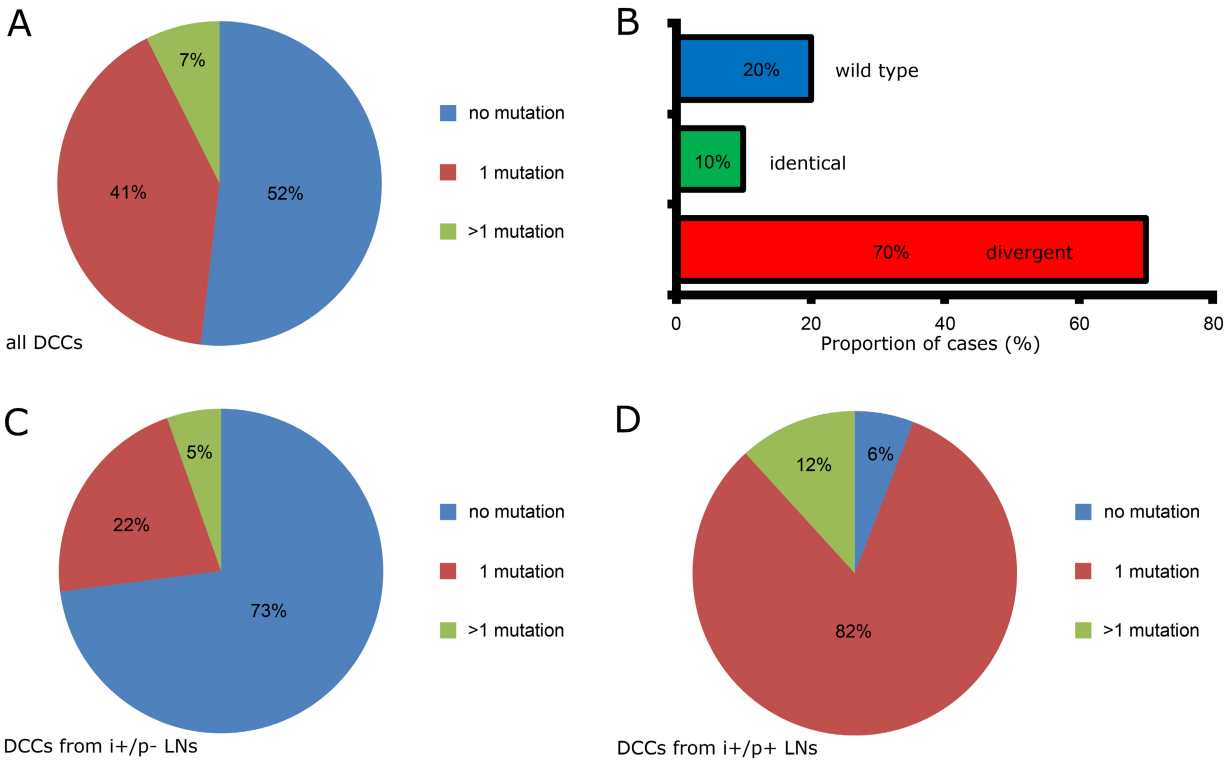

Figure 4. Mutation analysis of Ber-EP4$^+$ DCCs and the corresponding PTs. (A) Targeted Sanger sequencing of mutational hot spots in *EGFR* exons 18–21, *HER2* exon 20, *KRAS* exon 2, *PIK3CA* exons 9 and 20, and *TP53* exons 5–8 performed for $n = 54$ Ber-EP4$^+$ cells. The DCCs were isolated from 29 LNs of 25 different patients. (B) Mutational heterogeneity in paired PT–DCC samples. The mutational status of the corresponding PTs was obtained for 20 patients (49 Ber-EP4$^+$ cells from 24 LNs). The bars depict the cumulative mutational status per patient (wild type: no mutation in DCCs and the corresponding PT; identical: same mutation in DCC and matched PT; divergent: different mutational status, i.e. either different mutations in DCC and matched PTs or mutation in one compartment and wild type in the other). (C) Targeted sequencing of 37 single cells from i+/p− LNs (23 LNs from 22 patients). (D) Targeted sequencing of 17 single cells from i+/p+ LNs (six LNs from five patients). Note the markedly higher proportion of cells harbouring mutations in p+/i+ LNs. All samples were collected from chemotherapy-naive patients.

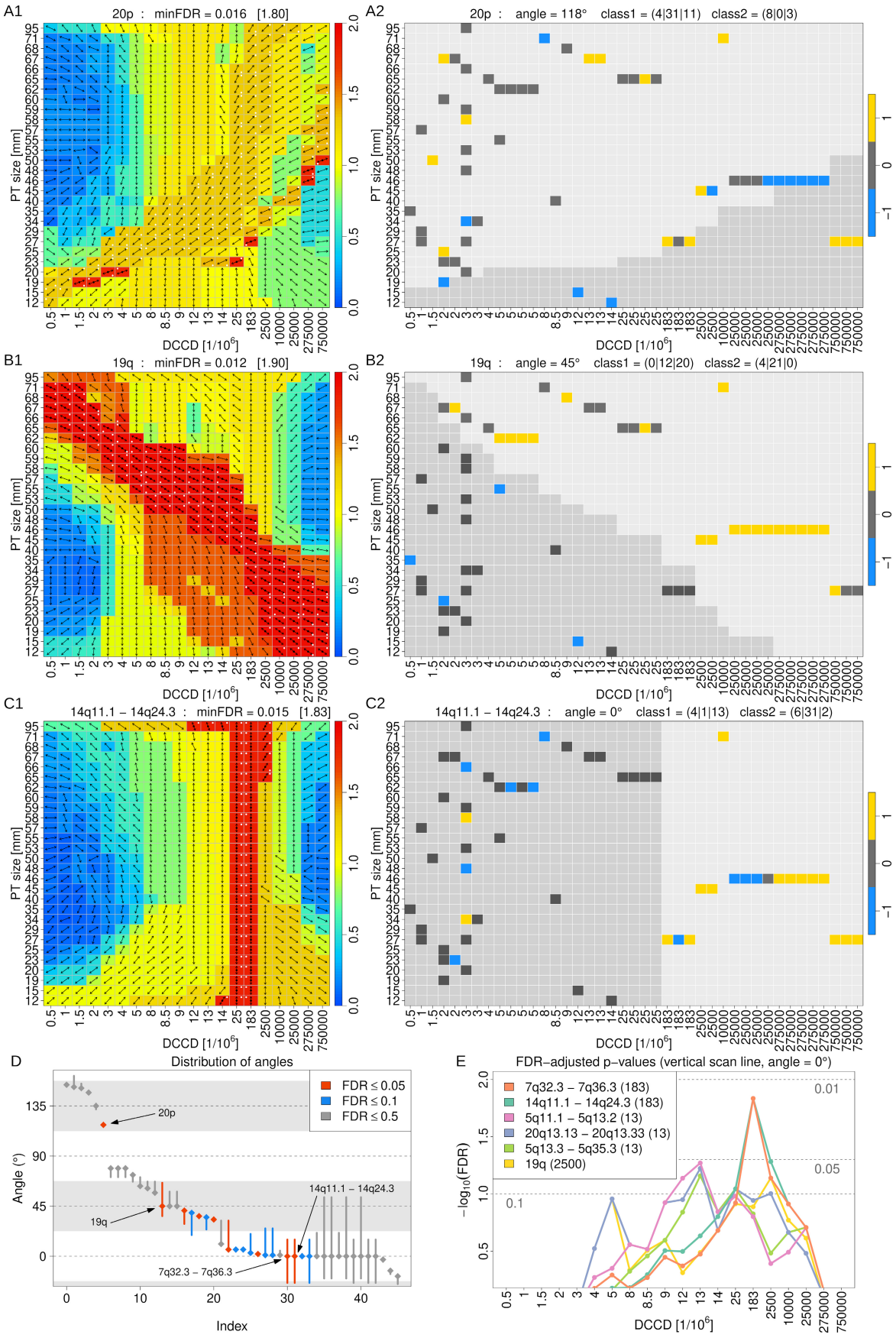

Figure 5 Legend on next page.

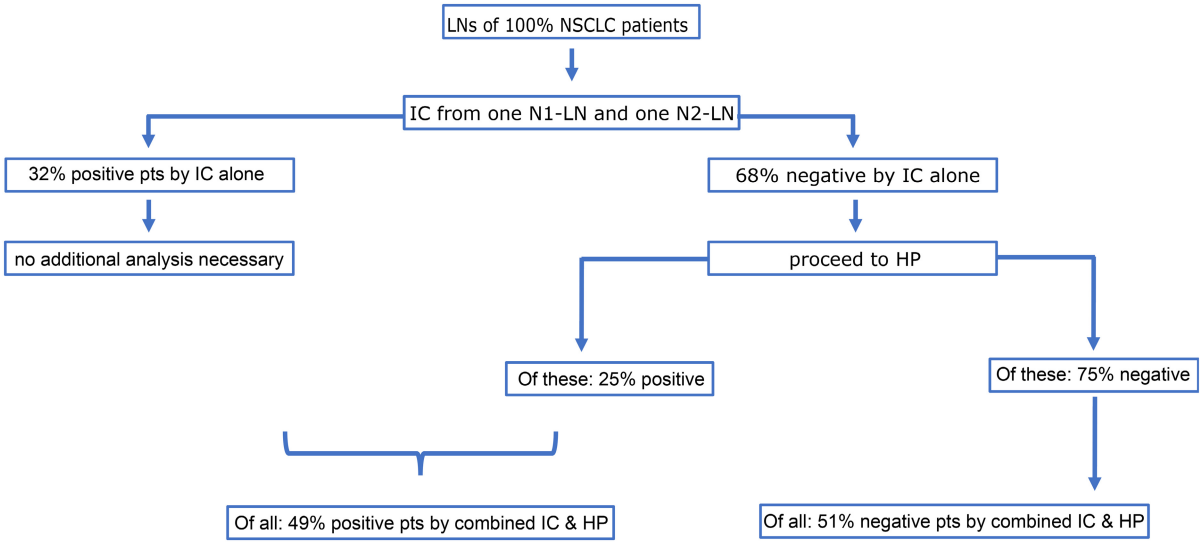

Figure 6. Suggested workflow for LN sample processing in diagnostic procedures. Sequential LN analysis by IC and HP, in which IC of two split LNs, one N1- and one N2-LN, identifies first about 1/3 of LN-positive patients. For patients with i-N2 station, all remaining LNs of the respective station are examined by HP to detect LN spread in 17% of patients that harbour DCCs in additional LNs and cHPIC determines the final N status.

of cells need to be analysed, it is safe to state that most alterations seem to be acquired during colony formation, i.e. when DCCs increase from DCCD > 10 to DCCD > 100 (Figure 5E).

Discussion

Here, we explored the usefulness of quantitative IC for the assessment of lymphatic cancer spread in NSCLC. For this, we addressed two aspects: (1) whether, and to what extent, LN-IC can improve risk assessment; and (2) how single-cell omics, made possible by LN-IC, promote our understanding of systemic cancer progression in NSCLC.

We first compared IC versus routine HP work-up for 220 LNs from 122 patients; IC showed an approximately five-fold higher detection rate for lymphatic cancer spread, indicating that HP procedures currently used in daily routine underestimate LN involvement in NSCLC patients. The IC results were confirmed by step sectioning and IHC. Although dividing LNs in half inevitably harbours the risk of generating sampling errors affecting both methods, the extended HP work-up confirmed the presence of cancer cells detected by IC, yet not in all cases. As the LN level containing DCCs has to be exactly hit but is easily missed in cases of single isolated cells or very small clusters, IC – after homogenization of LN tissue and enrichment of the mononuclear cell fraction – is advantageous. Due to the higher sensitivity, IC reduces false-negative LNs, enabling a better separation of positive and negative LNs for prognostic stratification. DCC detection by IC demonstrated clear clinical relevance, outperforming all other variables for predicting OS (including pN and pT status) in univariable and multivariable

Figure 5. DCCs acquire genetic alterations within and outside the PT. Genome alteration patterns have been explained by four different risk scenarios in Werner-Klein *et al* [19] depending on the number of cell divisions in the PT (n_T) and the LN (n_L): scenario n_T: exclusive risk of incurring alterations in the PT; $n_T + n_L$: equal risk in PT and LN; $n_T + 2n_L$: doubled risk in LN compared with PT; and n_L: exclusive risk from cell divisions in the LN. In the present study, significant evidence was found only for the first, third, and fourth scenarios (A1–C2). (A1–C1) Classification results for three prototypic CGH results (loci) using linear classifiers with 2D set points according to the indicated DCCD and PT size values and directions as illustrated by the small arrows within each rectangle of the colour matrix (see Supplementary materials and methods for details). Colour encodes the negative decadic logarithm of the FDR-adjusted *P* values (Fisher's exact test) of each classification. Basically, the red/orange areas separate regions differing in their distribution of genomic alterations. Their orientation parallels equal risk lines (see ref 19). DCCD and PT size values were chosen according to experiment, while classifier set points were slightly displaced relative to these values (white points included for the 5% most significant FDR values). Minimum FDR values are indicated by black points. $N = 57$ DCCs, 32 patients, 50 loci. (A2–C2) Corresponding measurement results as well as the class assignments of the best (lowest FDR) classifiers indicated by the light (class1) and dark (class2) grey areas. The class boundary does not necessarily appear linear as samples are listed according to rank, while linear classification was performed in $\log_{10}$(DCCD)−$\log_{10}$(PT size) space. The classifier angle and the distribution of deletions (−1), balances (0), and amplifications (+1) in each class are given in the title. DCC genetic alterations acquired in the SLN correspond to colonization. (D) Distribution of best classifier angles for different loci. These can be grouped according to the prototypic categories (scenarios) n_T, $n_T + n_L$, $n_T + 2n_L$, and n_L (see ref 19). Vertical bars correspond to the range of angles associated with the same minimal FDR value; diamonds indicate angles that are closest to the ideal prototypes (dashed lines). The 46 loci with FDR ≤ 0.5 are displayed. All of these have at least ten samples per class. (E) DCCD dependence of genomic changes (quantified by FDR-adjusted *P* values) derived using a vertical classifier (angle = 0°) positioned according to the experimentally measured DCCD values (set points left of these values). Only loci in categories $n_T + 2n_L$ and n_L whose FDR values fall below 0.075 are shown. In the legend, numbers in parentheses following the loci [e.g. 7q32.3–7q36.3 (183)] indicate peak DCCD. Note that any peak between DCCD 25 and 183 will appear at 183.

analyses. The findings are consistent with reported prognostic relevance of minimal lymphatic spread in NSCLC patients using Ber-EP4 or antibodies against cytokeratins (AE1/AE3, CK5/6, CK7) for DCC detection [5–14,29]. Nevertheless, some studies report a lack of prognostic significance of DCCs or micrometastases [30–32], which may be related to the size of the patient cohort, patient selection biases, the length of follow-up, treatment effects, and the heterogeneity of the protocols used for detection of DCCs. Indeed, our data indicate that the low sensitivity of limited HP may blur true and false-negative cases, precluding clear separation of patients with and without local spread. In contrast, using IC, we observed defined quantitative effects, as in melanoma [18,33], i.e. a higher risk for patients with DCCD values above and below 100.

Despite the higher sensitivity of IC, HP identified 17% additionally positive patients when all available LNs were analysed (in our study, about ten-fold higher LN numbers). This is consistent with the complex pulmonary lymphatic drainage [34] and the lack of SLNs for NSCLC [35]. In addition, metastatic LN skipping is a well-documented phenomenon [36,37] impeding rational LN selection. We therefore combined IC and HP (N_{cHPIC}), which resulted in the best overall stratification of prognostic subgroups. The N_{cHPIC} status outperformed all risk factors in multivariable analysis. This result may prompt testing N_{cHPIC} in a sequential approach (Figure 6) that could reduce the workload while improving diagnostic precision. A more comprehensive analysis could address if IC, based on more than two LN halves, further improves sensitivity and prognostic precision. Automation of the workflow and (AI-based) prescreening of disaggregated LNs could help to examine all resected LNs by IC. Possibly, a number of LNs can be determined that optimizes the balance between diagnostic precision and workload. Nevertheless, there is an imperative need to collect more information on the survival impact of DCCs, because several studies failed to demonstrate a significant effect of LN-DCC detection [30–32]. It needs to be determined whether the heterogeneity of detection methods or the use of different detection markers contributed to these conflicting results. Until full validation, IC should be combined with HP that preserves comprehensive morphologic information, such as the detection of extranodular extension of LN metastases [38,39] or of (lymph-) angioinvasion in the perinodal fatty tissue [40] or of rare benign gland inclusions [41,42] which are a potential diagnostic pitfall in IC because of Ber-EP4 positivity.

We also explored whether IC combined with single-cell omics generates insight into the systemic progression of NSCLC. Here, we made three key observations: (1) detection of DCCs is independent of PT size; (2) DCCs and PTs are molecularly disparate; and (3) genetic alterations are acquired as DCCs form metastatic colonies in LNs. These findings are consistent with an increasing number of observations that cancers disseminate early and progress in parallel to primary tumour formation [43–46]. Specifically for NSCLC, lymphatic spread even in small-sized tumours is well documented [47–49]. Consequently, we found no significant correlation between the number of DCCs in the LN and the PT size, and the detection rate of DCCs only marginally increased with increasing PT size, in line with studies that found no correlation between DCC detection and T stage [5,9,11,14,32].

Array CGH analysis revealed aberrant genomes in more than 80% of single DCCs and confirmed their malignant origin. Whether the remaining cells are non-cancer cells or cells whose aberrations are below our detection limit is currently unclear [28]. In support of the latter, we previously noted DCCs of breast cancer patients in bone marrow displaying balanced profiles, yet harbouring high numbers of LOH events [50], suggesting dissemination before acquisition of relevant genetic lesions at the primary site. As in oesophageal cancer and melanoma [19,51], aCGH and Sanger sequencing revealed molecular disparity between NSCLC-DCCs and the corresponding PTs. Particularly the disparity between PTs and DCCs for mutations such as KRAS and p53 is in stark contrast to the postulated origin of metastases from advanced primary tumours [52] and indicates that clonal dynamics are more complex than currently thought. For example, several changes were clearly dependent on the number of cell divisions in the LN and not or less on the number of cell divisions in the PT. One of them was a gain of chromosome 19q in i+/p+ DCCs, which has been associated with activation of mTOR, NF-κB, and STAT3 pathways resulting in cancer progression [53]. Based on the observed parallel progression of DCCs and PTs, we expect that analysis of DCCs for therapy relevant ('druggable') alterations will sharpen therapeutic precision by identifying molecular vulnerabilities on the target cells of systemic therapies directly, as opposed to drug target extrapolation from primary tumours.

Limitations of our study are its single-centre origin, the number of patients, and molecular analyses. Furthermore, we did not systematically explore lymph-vessel invasion for comparison with DCC status and prognostic impact. Future research should not only validate the observed prognostic importance of DCCs in NSCLC and the advantages of IC for diagnostics but also exploit the unique chances of identifying candidate metastasis founder cells. To this end, LN-IC opens multiple opportunities to investigate the existence of different DCC subpopulations, as we previously noted for BM-DCCs in NSCLC [54], and the molecular phenotypes whose precise description may uncover novel therapeutic approaches.

Acknowledgements

We thank Isabell Blochberger, Sandra Grunewald, Manfred Meyer, Irene Nebeja, Thomas Schamberger, Katharina Schneider, Sonja Fleischmann, and Doris Gaag for excellent technical assistance. We are grateful to Armin Pauer (Tumorzentrum Regensburg) for clinical follow-up information. We also thank Wolfgang Dietmaier for help with *EGFR* primers and Matthias Evert and Arndt Hartmann for general support. This work was supported by the European Research Council (ERC-2012-ADG-322602) and the Deutsche

Forschungsgemeinschaft (DFG; KL 1233/10-2 and KL 1233/11-1) to CAK. FE was supported by a Mildred Scheel doctoral fellowship from the German Cancer Aid (Processing number 70110399).

Author contributions statement

CAK and HSH designed the clinical study. HSH, TR, SS, and WS performed sample acquisition and clinical data management. FE, TM, and WS gathered follow-up information. FE, RFY, ZC, GF, BP, ST, PR, FW, and HW carried out experiments. FE, MH, BP, CAK, and RFY analysed data. FE and MH elaborated the figures. FE, MH, and CAK wrote the original manuscript draft. All the authors were involved in manuscript review and editing, and approved the submitted and published versions.

Data availability statement

The data sets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

1. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. *Nature* 2018; **553:** 446–454.
2. SEER*Explorer, National Cancer Institute. 2021; Accessed 25 April 2022] Available from: https://seer.cancer.gov/statistics-network/explorer/application.html?site=47&data_type=4&graph_type=5&compareBy=sex&chk_sex_1=1&chk_sex_3=3&chk_sex_2=2&series=9&race=1&age_range=1&stage=101&advopt_precision=1&advopt_show_ci=on
3. Goldstraw P, Chansky K, Crowley J, *et al*. The IASLC lung cancer staging project: proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer. *J Thorac Oncol* 2016; **11:** 39–51.
4. Lambert AW, Pattabiraman DR, Weinberg RA. Emerging biological principles of metastasis. *Cell* 2017; **168:** 670–691.
5. Passlick B, Izbicki JR, Kubuschok B, *et al*. Immunohistochemical assessment of individual tumor cells in lymph nodes of patients with non-small-cell lung cancer. *J Clin Oncol* 1994; **12:** 1827–1832.
6. Izbicki JR, Passlick B, Hosch SB, *et al*. Mode of spread in the early phase of lymphatic metastasis in non-small-cell lung cancer: significance of nodal micrometastasis. *J Thorac Cardiovasc Surg* 1996; **112:** 623–630.
7. Osaki T, Oyama T, Inoue M, *et al*. Molecular biological markers and micrometastasis in resected non-small-cell lung cancer. Prognostic implications. *Jpn J Thorac Cardiovasc Surg* 2001; **49:** 545–551.
8. Wu J, Ohta Y, Minato H, *et al*. Nodal occult metastasis in patients with peripheral lung adenocarcinoma of 2.0 cm or less in diameter. *Ann Thorac Surg* 2001; **71:** 1772–1777; discussion 1777–1778.
9. Gu CD, Osaki T, Oyama T, *et al*. Detection of micrometastatic tumor cells in pN0 lymph nodes of patients with completely resected nonsmall cell lung cancer: impact on recurrence and survival. *Ann Surg* 2002; **235:** 133–139.
10. Tezel C, Ersev AA, Kiral H, *et al*. The impact of immunohistochemical detection of positive lymph nodes in early stage lung cancer. *Thorac Cardiovasc Surg* 2006; **54:** 124–128.
11. Yasumoto K, Osaki T, Watanabe Y, *et al*. Prognostic value of cytokeratin-positive cells in the bone marrow and lymph nodes of patients with resected nonsmall cell lung cancer: a multicenter prospective study. *Ann Thorac Surg* 2003; **76:** 194–201; discussion 202.
12. Herpel E, Muley T, Schneider T, *et al*. A pragmatic approach to the diagnosis of nodal micrometastases in early stage non-small cell lung cancer. *J Thorac Oncol* 2010; **5:** 1206–1212.
13. Rusch VW, Hawes D, Decker PA, *et al*. Occult metastases in lymph nodes predict survival in resectable non-small-cell lung cancer: report of the ACOSOG Z0040 trial. *J Clin Oncol* 2011; **29:** 4313–4319.
14. Martin LW, D'Cunha J, Wang X, *et al*. Detection of occult micrometastases in patients with clinical stage I non-small-cell lung cancer: a prospective analysis of mature results of CALGB 9761 (Alliance). *J Clin Oncol* 2016; **34:** 1484–1491.
15. Izbicki JR, Passlick B, Karg O, *et al*. Impact of radical systematic mediastinal lymphadenectomy on tumor staging in lung cancer. *Ann Thorac Surg* 1995; **59:** 209–214.
16. Wright G, Manser RL, Byrnes G, *et al*. Surgery for non-small cell lung cancer: systematic review and meta-analysis of randomised controlled trials. *Thorax* 2006; **61:** 597–603.
17. S3-Leitlinie Prävention, Diagnostik, Therapie und Nachsorge des Lungenkarzinoms. 2018; [Accessed 18 March 2022] Available from: https://www.awmf.org/uploads/tx_szleitlinien/020-007OL_l_S3_Lungenkarzinom_2018-03.pdf
18. Ulmer A, Dietz K, Hodak I, *et al*. Quantitative measurement of melanoma spread in sentinel lymph nodes and survival. *PLoS Med* 2014; **11:** e1001604.
19. Werner-Klein M, Scheitler S, Hoffmann M, *et al*. Genetic alterations driving metastatic colony formation are acquired outside of the primary tumour in melanoma. *Nat Commun* 2018; **9:** 595.
20. Klein CA. Parallel progression of primary tumours and metastases. *Nat Rev Cancer* 2009; **9:** 302–312.
21. Klein CA. Selection and adaptation during metastatic cancer progression. *Nature* 2013; **501:** 365–372.
22. Klein CA. Cancer progression and the invisible phase of metastatic colonization. *Nat Rev Cancer* 2020; **20:** 681–694.
23. Ulmer A, Fischer JR, Schanz S, *et al*. Detection of melanoma cells displaying multiple genomic changes in histopathologically negative sentinel lymph nodes. *Clin Cancer Res* 2005; **11:** 5425–5432.
24. Latza U, Niedobitek G, Schwarting R, *et al*. Ber-EP4: new monoclonal antibody which distinguishes epithelia from mesothelial. *J Clin Pathol* 1990; **43:** 213–219.
25. Fehm T, Braun S, Muller V, *et al*. A concept for the standardized detection of disseminated tumor cells in bone marrow from patients with primary breast cancer and its clinical implementation. *Cancer* 2006; **107:** 885–892.
26. Klein CA, Blankenstein TJ, Schmidt-Kittler O, *et al*. Genetic heterogeneity of single disseminated tumour cells in minimal residual cancer. *Lancet* 2002; **360:** 683–689.
27. Klein CA, Schmidt-Kittler O, Schardt JA, *et al*. Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. *Proc Natl Acad Sci U S A* 1999; **96:** 4494–4499.
28. Czyz ZT, Hoffmann M, Schlimok G, *et al*. Reliable single cell array CGH for clinical samples. *PLoS One* 2014; **9:** e85907.
29. Kubuschok B, Passlick B, Izbicki JR, *et al*. Disseminated tumor cells in lymph nodes as a determinant for survival in surgically resected non-small-cell lung cancer. *J Clin Oncol* 1999; **17:** 19–24.
30. Marchevsky AM, Gupta R, Kusuanco D, *et al*. The presence of isolated tumor cells and micrometastases in the intrathoracic lymph nodes of patients with lung cancer is not associated with decreased survival. *Hum Pathol* 2010; **41:** 1536–1543.
31. Rena O, Carsana L, Cristina S, *et al*. Lymph node isolated tumor cells and micrometastases in pathological stage I non-small cell lung cancer: prognostic significance. *Eur J Cardiothorac Surg* 2007; **32:** 863–867.
32. Rud AK, Boye K, Fodstad Ø, *et al*. Detection of disseminated tumor cells in lymph nodes from patients with early stage non-small cell lung cancer. *Diagn Pathol* 2016; **11:** 50.

33. Ulmer A, Dietz K, Werner-Klein M, *et al*. The sentinel lymph node spread determines quantitatively melanoma seeding to non-sentinel lymph nodes and survival. *Eur J Cancer* 2018; **91:** 1–10.
34. Riquet M, Hidden G, Debesse B. Direct lymphatic drainage of lung segments to the mediastinal nodes. An anatomic study on 260 adults. *J Thorac Cardiovasc Surg* 1989; **97:** 623–632.
35. Schirren J, Bergmann T, Beqiri S, *et al*. Lymphatic spread in resectable lung cancer: can we trust in a sentinel lymph node? *Thorac Cardiovasc Surg* 2006; **54:** 373–380.
36. Ilic N, Petricevic A, Arar D, *et al*. Skip mediastinal nodal metastases in the IIIa/N2 non-small cell lung cancer. *J Thorac Oncol* 2007; **2:** 1018–1021.
37. Baisi A, Raveglia F, De Simone M, *et al*. Micrometastasis and skip metastasis as predictive factors in non-small-cell lung cancer staging. *Eur J Cardiothorac Surg* 2013; **43:** 1075.
38. Lee YC, Wu CT, Kuo SW, *et al*. Significance of extranodal extension of regional lymph nodes in surgically resected non-small cell lung cancer. *Chest* 2007; **131:** 993–999.
39. Luchini C, Veronese N, Nottegar A, *et al*. Extranodal extension of nodal metastases is a poor prognostic moderator in non-small cell lung cancer: a meta-analysis. *Virchows Arch* 2018; **472:** 939–947.
40. Heldwein MB, Doerr F, Schlachtenberger G, *et al*. Lymphangiosis carcinomatosa independently affects long-term survival of non-small cell lung cancer patients. *Surg Oncol* 2021; **37:** 101611.
41. Lewis AL, Truong LD, Cagle P, *et al*. Benign salivary gland tissue inclusion in a pulmonary hilar lymph node from a patient with invasive well-differentiated adenocarcinoma of the lung: a potential misinterpretation for the staging of carcinoma. *Int J Surg Pathol* 2011; **19:** 382–385.
42. Fellegara G, Carcangiu ML, Rosai J. Benign epithelial inclusions in axillary lymph nodes: report of 18 cases and review of the literature. *Am J Surg Pathol* 2011; **35:** 1123–1133.
43. Gruber IV, Hartkopf AD, Hahn M, *et al*. Relationship between hematogenous tumor cell dissemination and cellular immunity in DCIS patients. *Anticancer Res* 2016; **36:** 2345–2351.
44. Hosseini H, Obradović MMS, Hoffmann M, *et al*. Early dissemination seeds metastasis in breast cancer. *Nature* 2016; **540:** 552–558.
45. Hüsemann Y, Geigl JB, Schubert F, *et al*. Systemic spread is an early step in breast cancer. *Cancer Cell* 2008; **13:** 58–68.
46. Sänger N, Effenberger KE, Riethdorf S, *et al*. Disseminated tumor cells in the bone marrow of patients with ductal carcinoma *in situ*. *Int J Cancer* 2011; **129:** 2522–2526.
47. Inoue M, Minami M, Shiono H, *et al*. Clinicopathologic study of resected, peripheral, small-sized, non-small cell lung cancer tumors of 2 cm or less in diameter: pleural invasion and increase of serum carcinoembryonic antigen level as predictors of nodal involvement. *J Thorac Cardiovasc Surg* 2006; **131:** 988–993.
48. Hamatake D, Yoshida Y, Miyahara S, *et al*. Surgical outcomes of lung cancer measuring less than 1 cm in diameter. *Interact Cardiovasc Thorac Surg* 2012; **15:** 854–858.
49. Zhang YK, Chai ZD, Tan LL, *et al*. Association of lymph node involvement with the prognosis of pathological T1 invasive non-small cell lung cancer. *World J Surg Oncol* 2017; **15:** 64.
50. Schardt JA, Meyer M, Hartmann CH, *et al*. Genomic analysis of single cytokeratin-positive cells from bone marrow reveals early mutational events in breast cancer. *Cancer Cell* 2005; **8:** 227–239.
51. Stoecklein NH, Hosch SB, Bezler M, *et al*. Direct genetic analysis of single disseminated cancer cells for prediction of outcome and therapy selection in esophageal cancer. *Cancer Cell* 2008; **13:** 441–453.
52. Jamal-Hanjani M, Wilson GA, McGranahan N, *et al*. Tracking the evolution of non-small-cell lung cancer. *N Engl J Med* 2017; **376:** 2109–2121.
53. Wang X, Zhang Y, Nilsson CL, *et al*. Association of chromosome 19 to lung cancer genotypes and phenotypes. *Cancer Metastasis Rev* 2015; **34:** 217–226.
54. Mederer T, Elsner F, Robold T, *et al*. EpCAM-positive disseminated cancer cells in bone marrow impact on survival of early-stage NSCLC patients. *Lung Cancer* 2022; **167:** 73–77.
55. Travis WD, Al Dayel FHB, Chung JH, *et al*. *WHO Classification of Tumours (5th edn), Volume 5: Thoracic Tumours*. IARC Press: Lyon, 2021.
56. Savic S, Tapia C, Grilli B, *et al*. Comprehensive epidermal growth factor receptor gene analysis from cytological specimens of non-small-cell lung cancers. *Br J Cancer* 2008; **98:** 154–160.
57. Detection of TP53 mutations by Sanger sequencing (IARC protocol, 2019 update); [Accessed 18 March 2022] Available from: https://tp53.isb-cgc.org/pdf/TP53_SangerSequencing_IARC
58. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2022. Available from: https://www.R-project.org
59. Bengtsson H. A Unifying Framework for Parallel and Distributed Processing in R using Futures. *R J* 2021; **13:** 208–227.
60. Microsoft, Weston S. foreach: Provides Foreach Looping Construct. R package version 1.5.1, 2021. Available from: https://CRAN.R-project.org/package=foreach

References 55–60 are cited only in the supplementary material.

SUPPLEMENTARY MATERIAL ONLINE

Supplementary materials and methods

Figure S1. Schematic overview of sample processing

Figure S2. Isolation of primary tumour DNA from FFPE tissue (referred to in Supplementary materials and methods)

Figure S3. Correlation of ADC subtypes and detection of DCCs in the LN ($n = 66$)

Figure S4. Copy number variations of aberrant DCCs ($n = 32$)

Figure S5. Hierarchical cluster analysis of isolated Ber-EP4$^+$ cells and matched primary tumours

Figure S6. Mutations detected in DCCs and matched PTs for 25 patients

Table S1. Primer sequences and PCR programs used for gene-specific amplification and subsequent Sanger sequencing (referred to in Supplementary materials and methods)

Table S2. Correlation between detection of DCCs in the LN and lymphatic vessel involvement ($n = 122$ patients)

Table S3. Results for re-examination of selected i+/p− LNs by step sectioning and IHC ($n = 8$)

Table S4. aCGH analysis of Ber-EP4$^+$ cells

Table S5. Characteristics of additional patients for chromosomal CGH analysis

Table S6. Characteristics of best classifiers for different loci