



A biobank of small cell lung cancer CDX models elucidates inter- and intratumoral phenotypic heterogeneity

Kathryn L. Simpson¹, Ruth Stoney^{1,7}, Kristopher K. Frese¹, Nicole Simms¹, William Rowe^{1,8}, Simon P. Pearce¹, Sam Humphrey¹, Laura Booth¹, Derrick Morgan¹, Marek Dynowski², Francesca Trapani¹, Alessia Catozzi¹, Mitchell Revill¹, Thomas Helps¹, Melanie Galvin¹, Luc Girard³, Daisuke Nonaka⁴, Louise Carter^{4,5}, Matthew G. Krebs^{1,6}, Natalie Cook^{4,5}, Mathew Carter¹, Lynsey Priest¹, Alastair Kerr¹, Adi F. Gazdar³, Fiona Blackhall^{4,5,6} and Caroline Dive^{1,6}✉

Although small cell lung cancer (SCLC) is treated as a homogeneous disease, biopsies and preclinical models reveal heterogeneity in transcriptomes and morphology. SCLC subtypes were recently defined by neuroendocrine transcription factor (NETF) expression. Circulating-tumor-cell-derived explant models (CDX) recapitulate donor patients' tumor morphology, diagnostic NE marker expression and chemotherapy responses. We describe a biobank of 38 CDX models, including six CDX pairs generated pretreatment and at disease progression revealing complex intra- and intertumoral heterogeneity. Transcriptomic analysis confirmed three of four previously described subtypes based on *ASCL1*, *NEUROD1* and *POU2F3* expression and identified a previously unreported subtype based on another NETF, *ATOH1*. We document evolution during disease progression exemplified by altered *MYC* and *NOTCH* gene expression, increased 'variant' cell morphology, and metastasis without strong evidence of epithelial to mesenchymal transition. This CDX biobank provides a research resource to facilitate SCLC personalized medicine.

SCLC is an aggressive neuroendocrine (NE) cancer with dismal prognosis, characterized by rapid growth, prevalent circulating tumor cells (CTCs) and early metastasis^{1–3}. SCLC is diagnosed by expression of NE markers including neural cell adhesion molecule (NCAM/CD56), chromogranin A (CHGA), synaptophysin (SYP) and of thyroid transcription factor-1 (NKX2-1)⁴. The genomic landscape of SCLC is characterized by almost universal and frequent loss of tumor suppressors *TP53* and *RBI*, respectively^{2,5,6}; amplified *MYC* family oncogenes (*MYC*, *MYCL* and *MYCN*) in ~20% cases^{2,5,7,8} and frequent mutations in *NOTCH* family genes^{2,5,6}. Abnormal epigenetic regulation is also a recurrent feature^{9,10,2,5}.

The pulmonary NE cell is the cell of origin for most SCLC^{11,12} where the transcription factor achaete-scute complex homolog-like 1 (*ASCL1*) induces neuronal and NE differentiation^{13,14}. *ASCL1* is expressed in 75% of human SCLC. Cell lines derived from *ASCL1*^{high} tumors grow with 'classic' morphology, as loosely adherent clusters, often with *MYCL* amplification¹⁵. *Ascl1* is required for tumorigenesis in genetically engineered mouse models (GEMMs) generated by triple knockout of *Trp53*, *Rb1* and *Rbl2* where *Mycl* is amplified^{16–18}. A second NETF, Neurogenic Differentiation Factor 1 (*NEUROD1*) is expressed in ~24% of SCLC tumors and in 19% of SCLC cell lines (with or without *ASCL1* coexpression)⁴. SCLC cells expressing *NEUROD1* display morphologies ranging from 'classic' to 'variant', growing as loosely adherent monolayers, often with *MYC*

amplification^{10,16,19,20}. *NEUROD1*^{high} tumors typically have lower diagnostic NE marker expression compared to *ASCL1*^{high} tumors. 'Variant' histopathology is recapitulated in a SCLC GEMM upon enforced *Myc* expression²¹, leading to elevated *Neurod1*, reduced NE markers and more aggressive tumors. Approximately 10% of human SCLCs express neither *ASCL1* nor *NEUROD1* (ref. ²). A Tuft Cell transcription factor POU class 2 homeobox 3 (*POU2F3*) defines a further subtype²² and a consensus on SCLC subtype classification was recently reported²³.

For the past 30 years, metastatic SCLC was treated with a platinum agent plus etoposide as standard of care (SoC)²⁴. Immune checkpoint inhibitors (ICIs) benefit some patients and were recently licensed for SoC in the United States^{25,26}. DNA damage repair inhibitors (DDRI) show efficacy in preclinical models and as DDR inhibition promotes STING-dependent activation of cytotoxic T cells there is rationale for DDRi and ICI combination therapy²⁷. Differential vulnerability to aurora kinase inhibitors was observed in preclinical models based on *MYC* family member expression^{21,28}. However, a comprehensive understanding of SCLC biology to support personalized therapy is hindered by difficulties obtaining sufficient biopsies adequate for research, especially at post-SoC disease progression²⁹.

SCLC research tools comprise established human cell lines and derived xenografts^{19,30}, GEMMs^{12,16,18,21,31}, Patient-derived xenografts

¹Cancer Research UK Manchester Institute Cancer Biomarker Centre, University of Manchester, Manchester, UK. ²Scientific Computing Core Facility, Cancer Research UK Manchester Institute, The University of Manchester, Manchester, UK. ³Hamon Center for Therapeutic Oncology Research, Department of Pharmacology, UT Southwestern Medical Center, Dallas, TX, USA. ⁴The Christie NHS Foundation Trust, Manchester, UK. ⁵Division of Cancer Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK. ⁶Cancer Research UK Lung Cancer Centre of Excellence, Manchester, UK. ⁷Present address: Manchester Institute of Biotechnology, The University of Manchester, Manchester, UK. ⁸Present address: Manchester Centre for Synthetic Biology of Fine and Speciality Chemicals (SYNBIOCHEM), Manchester Institute of Biotechnology, The University of Manchester, Manchester, UK. ✉e-mail: Caroline.Dive@manchester.ac.uk

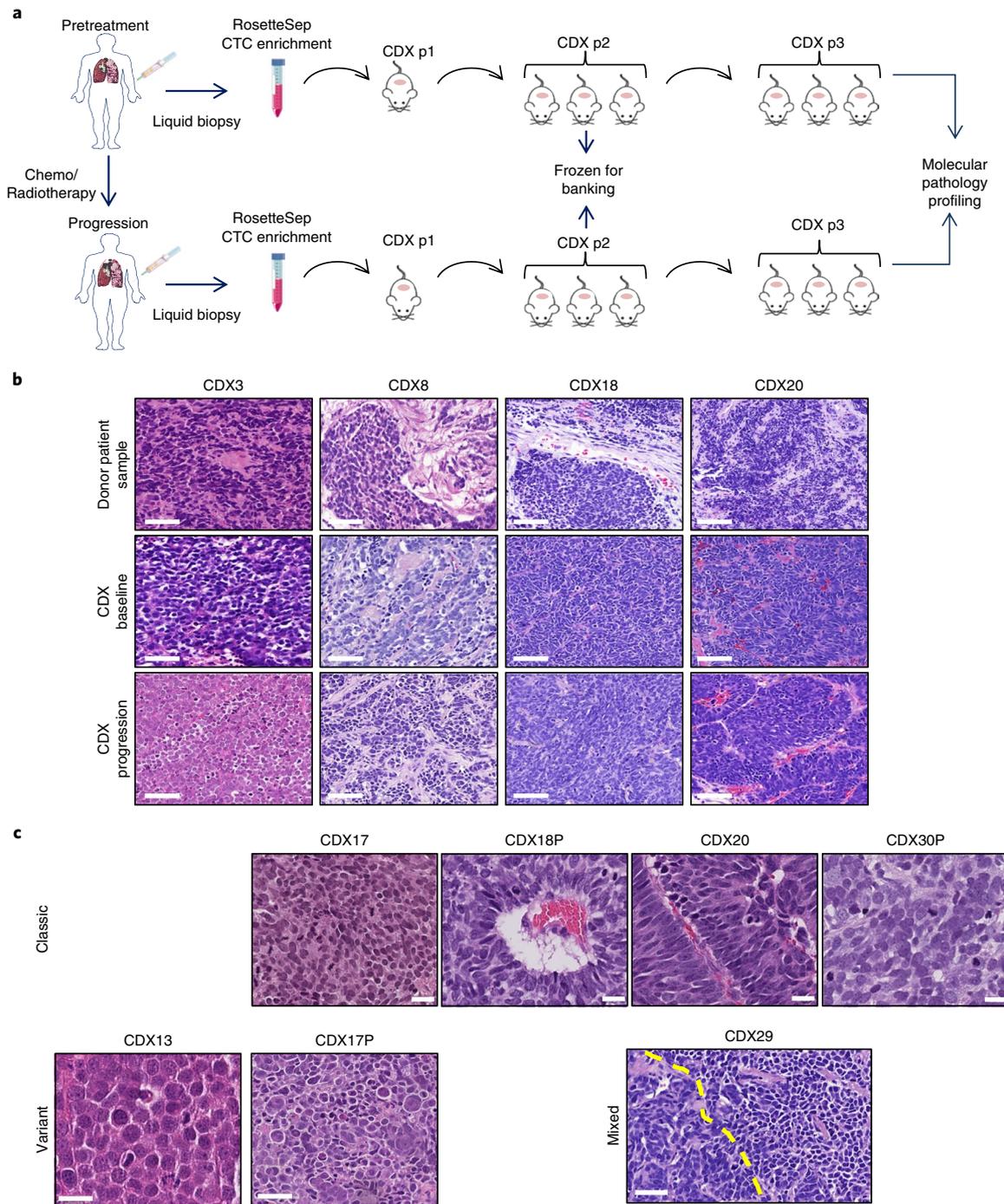


Fig. 1 | CDX generation and overview. **a**, CDX workflow. A single mouse per donor patient was engrafted with CTCs enriched from SCLC patient blood samples either at baseline (treatment naïve) or at post-treatment disease progression to generate the p1 CDX. Following successful engraftment p1 tumors were passaged into 3 × p2 animals. Resultant p2 tumors were either frozen or passaged into p3 animals for molecular characterization. **b**, Comparison of CDX and donor tumor specimen morphology. H&E staining of CDX donor diagnostic tissue and their CDX at passage p1. In each case, a whole-tumor section was analyzed and images shown are a representative area of these, shown for four patients where both a pretreatment baseline and disease progression CDX was generated. Scale bars, 50 μm. **c**, Classic and variant morphologies. H&E staining of selected CDX models. CDX17, CDX18P, CDX20 and CDX30P show typical classic morphology representative of most of the CDX panel and contain features present in patient samples such as pseudorosettes (for example, CDX18) and palisading (for example, CDX20). The rarer variant cellular morphology is depicted for CDX13 and CDX17P and intratumoral heterogeneity is shown for CDX29 (mixed), which contains both classic and variant cell morphologies (separated by dotted yellow line). CDX30P has large nuclei but classic morphology. Whole sections were analyzed per CDX tumor with three independent animals at p3 analyzed and a representative area is shown for each. Scale bars, 20 μm except variant examples; CDX13 and CDX17P and mixed CDX29, scale bars, 50 μm.

(PDXs) generated from patient tumor biopsies or pulmonary effusions^{9,10,32–36} and models generated from patients’ CTCs, so-called CDX^{29,33,37,38}. While GEMMs enable gene-disease causation stud-

ies, they do not recapitulate the extensive tobacco-induced mutagenic landscape of clinical SCLC present in human cell lines, PDX and CDX models^{9,29,33,37}. While cell lines are instructive and ame-

nable to experimentation, they suffer from ‘genomic drift’ during passage^{32,39} and are rarely obtained from the same patient before and after acquired chemoresistance. Cell lines and PDX can be forced to acquire chemoresistance in tissue culture or in mice^{9,36}, but whether those mechanisms commonly occur within patients is unclear. CDX models from serial blood samples before and after the patient develops progressive disease post-SoC^{33,37,38} now allow interrogation of the evolved SCLC biology present during clinical trials of new therapies. We described four CDX, which recapitulated the morphology and chemoresponse of their patient donor and genomic profile of donors’ CTCs³⁷, extended to ten CDX and ex vivo short term CDX cultures evaluating the combined efficacy of olaparib (a poly-ADP ribose polymerase inhibitor) and adavosertib (WEE1 inhibitor), which emphasized the importance of administering this combination before disease progression³⁸. Others have now adopted our CDX approach to augment their SCLC research³³.

Combined use of available SCLC models is enabling biomarker supported early clinical trials of targeted therapies. Deconvoluting heterogeneous SCLC phenotypes in large panels of patient relevant models will be essential for improved patient outcomes. We present a biobank of 38 CDX models from 31 SCLC patients who subsequently developed either chemosensitive or chemorefractory disease, including six paired baseline and disease progression models.

Results

Patient donors and generation of the CDX panel. Supplementary Table 1 summarizes clinical characteristics of the 31 patient donors, which successfully generated CDX (14 females, 17 males, average age 68 years), all former or current smokers (with the exception of CDX13, smoking history not recorded) who received SoC (patient treatment timelines shown in Supplementary Fig. 1). Most patients (28/31, 90%) presented with extensive stage (ES) disease with metastases to common sites; three patients presented with limited stage (LS) disease, two of those had no detectable metastases at diagnosis and one had local mediastinal lymph node metastasis.

We generated 38 CDX (Fig. 1a, Supplementary Table 1 and Supplementary Fig. 1). The number of implanted tumor cells for successful CDX generation is unknown although CDX generation rate was 17% (Extended Data Fig. 1). A diagnostic primary tumor biopsy was acquired in only 15 out of 31 patients (Supplementary Table 2). CDX were generated at a single time-point from 24 patients, 17 at pretreatment baseline and seven post-chemotherapy with disease progression (designated ‘P’). Six CDX were from longitudinally sampled patients at baseline and again after post-chemotherapy relapse, and for a seventh patient a paired model was generated from two post-chemotherapy relapse samples (15P/15PP) (Supplementary Table 1 and Supplementary Fig. 1). Of the three patients with LS SCLC, only CDX29 was derived at baseline and three CDX (15P, 15PP and 22P) were derived at disease progression. CellSearch (CS)-CTC counts (EpCAM and cytokeratin positive, CD45 negative) in a parallel blood sample for each CDX were variable (median 677, range 0–15,352 CTCs

per 7.5 ml blood) (Supplementary Table 2). Notably, the LS SCLC CDX29 donor had no CS-CTCs. Samples with a parallel CS-CTC number >49 per 7.5 ml blood resulted in a 49% CDX success rate compared to <1% for those with a parallel CS-CTC count of 0–49 (Extended Data Fig. 1). CS-CTC counts in longitudinal model donors did not always increase during progression with no obvious correlation between CS-CTC trajectories and overall survival (Supplementary Table 2).

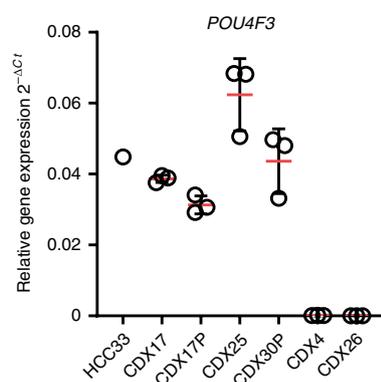
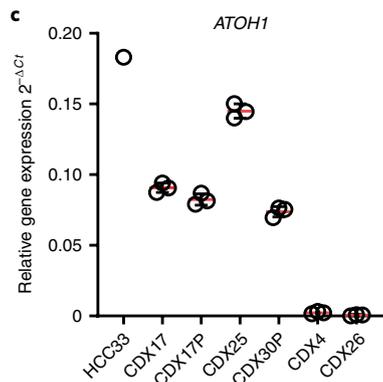
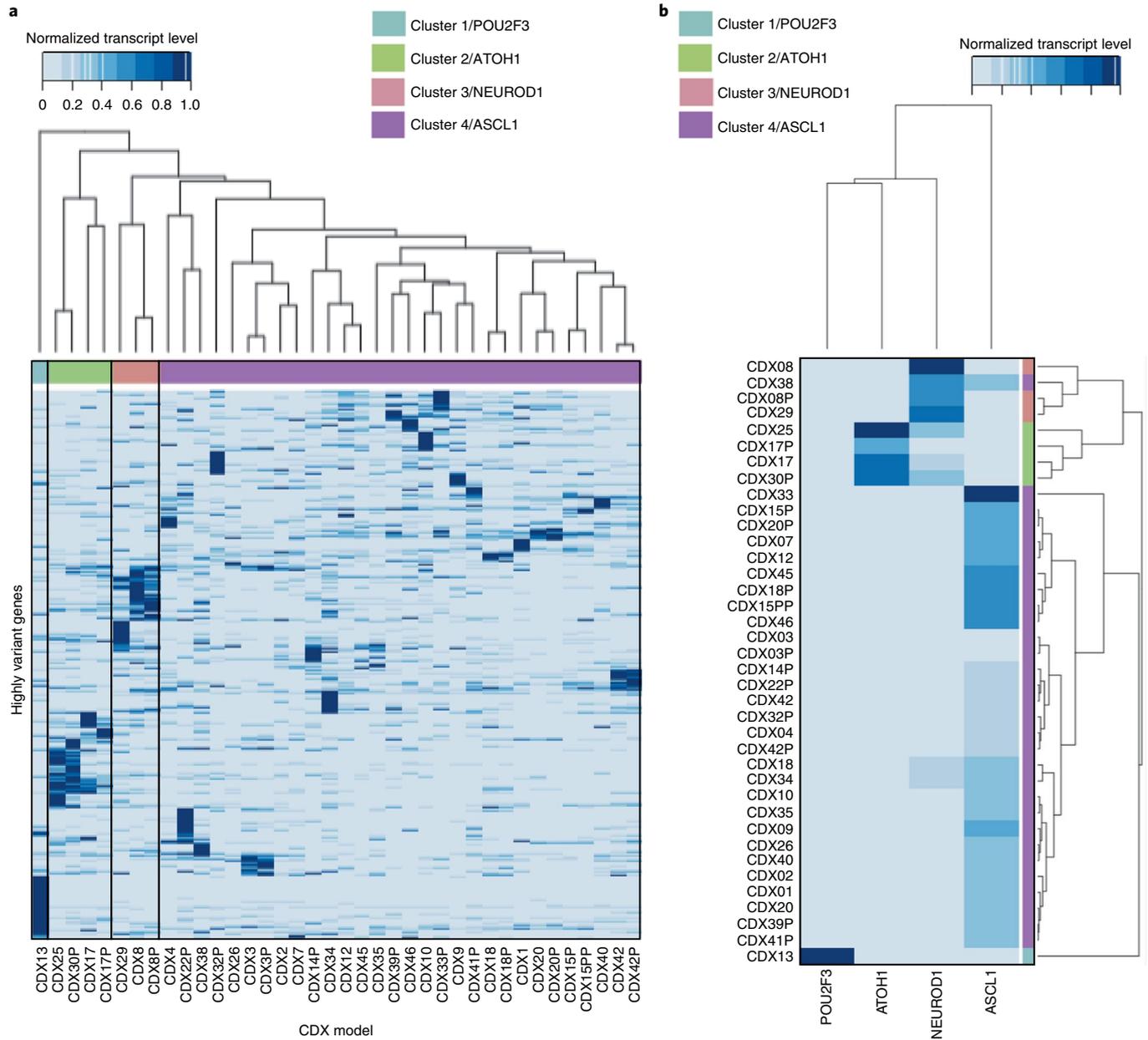
Morphology and NE markers in CDX and matched diagnostic patient specimens. Pathology evaluation (D.N. and A.F.G.) of donor patient biopsies confirmed SCLC (Fig. 1b) but there was insufficient specimen to assign classic and/or variant morphologies robustly. Previous studies using SCLC cell lines^{19,30} and GEMM models^{16,17,21} revealed frequent ‘classic’ and rarer ‘variant’ morphologies. CDX provides more tumor tissue for analysis than bronchoscopic biopsies²⁹ and morphologies were assessed by H&E staining at passage 1. Multiple morphological features of SCLC were observed, including ‘sheet-like’ cellular architecture (CDX3), pseudorosettes (CDX18), palisading and trabecular growth (CDX20; Fig. 1c). Most CDX (35 out of 38) contained neoplastic cells with small nuclei (20–40 μm), (Extended Data Fig. 2) consistent with ‘classic’ SCLC, as reported in 35 out of 50 (70%) human SCLC cell lines¹⁹. CDX13 and CDX17P had comparatively large nuclei (40–50 μm diameter, see Fig. 1c and Extended Data Fig. 2) consistent with ‘variant’ SCLC morphology^{16,19,30} but although CDX30P had large nuclei, pathology review (A.F.G.) reported classic morphology (Fig. 1c). CDX17 exhibited classic morphology suggesting that a switch to the variant morphology of CDX17P occurred during disease progression (Fig. 1c). CDX29 (limited stage donor) contained cells with classic and variant morphologies (Fig. 1c). Overall, this CDX biobank shows a range of classic to variant morphologies and that classic and variant cells can coexist within a tumor. Consistent with clinical data, all CDX except CDX13 expressed at least two out of the three diagnostic NE markers NCAM, CHGA and SYP by immunohistochemistry (IHC; Extended Data Fig. 3). Hierarchical clustering revealed overall that expression (loss or gain) of NE markers was unchanged between baseline and disease progression in paired models (Extended Data Fig. 3b).

Global transcriptomics analysis. The morphological differences and histopathological variations in NE markers (Fig. 1b,c and Extended Data Fig. 3) may reflect (in part), the recently described SCLC subtypes²³. We explored this possibility using unbiased hierarchical clustering of CDX transcriptomes that revealed four distinct clusters (Fig. 2a). Clusters 1, 3 and 4 comprised genes already associated with reported SCLC molecular subtypes, namely ASCL1 (cluster 4, 31/38), NEUROD1 (cluster 3, 3/38) and POU2F3 (cluster 1, the most divergent, 1/38)^{22,23} (Fig. 2b). Cluster 2 (4/38) represents an unrecognized subtype containing genes including the NETF *ATOH1* (the most significantly up-regulated cluster 2 gene relative to other clusters (adjusted *P* value = 4.58×10^{-44}) (Fig. 2b). *ATOH1* is both necessary and sufficient for

Fig. 2 | Global transcriptomic analysis of CDX models. **a**, Gene expression analysis of the 38 CDX models, demonstrating four clusters/subgroups derived from hierarchical clustering of the top 1,686 most variable genes across the CDX models with three independent CDX tumors analyzed per model. Cluster 1, light blue, POU2F3; Cluster 2, light green, ATOH1; Cluster 3, pink, NEUROD1 and Cluster 4, purple, ASCL1. **b**, Expression of four transcription factors: *ASCL1*, *NEUROD1*, *POU2F3* and *ATOH1* are significantly associated with four distinct subtypes of SCLC CDX. For the ASCL1 subtype, *ASCL1* is the most highly associated gene, for the NEUROD1 subtype *NEUROD1* is ranked 271st, for the POU2F3 subtype, *POU2F3* is ranked 56th, and for the ATOH1 subtype, *ATOH1* is the most highly associated gene. Cluster 1, light blue, POU2F3; Cluster 2, light green, ATOH1; Cluster 3, pink, NEUROD1 and Cluster 4, purple, ASCL1. **c**, RT-qPCR for *ATOH1* and *POU4F3* in CDX confirms high *ATOH1* and *POU4F3* expression in the ATOH1 subgroup compared to low expression in CDX4, CDX17 and CDX26 (ASCL1 subgroup) and with the cell line HCC33 as a positive control (high expression of *ATOH1* according to the CCLE). Relative fold change in gene expression is shown as $2^{-\Delta Ct}$, y axis, for each CDX model/cell line. *n* = 3 independent CDX tumors per model and *n* = 3 independent cell culture samples per cell line were analyzed, each in triplicate. Black bars represent mean \pm s.d.

hair cell differentiation in the mammalian auditory epithelium⁴⁰ and *Math1*, the mouse homolog of *ATOH1*, is essential for neuronal differentiation and mouse cerebellar development⁴¹. Reverse transcriptase–quantitative PCR (RT-qPCR) analysis validated high

expression of *ATOH1* and its target *POU4F3* (ref. ⁴⁰) in CDX17/17P, CDX25 and CDX30P and in the *ATOH1* expressing SCLC cell line HCC33 compared to CDX models with no *ATOH1* expression (CDX4 and CDX26) (Fig. 2c).



NE transcriptional regulators. The relative abundance of previously reported SCLC subtypes³⁰ was determined by IHC for ASCL1, NEUROD1, REST and POU2F3 (Fig. 3). The biobank represents a range of ASCL1/NEUROD1 expression profiles that segment into three groups (Fig. 3c). Twenty-two CDX (58%) were ASCL1⁺/NEUROD1⁻, eight CDX (21%) were ASCL1⁻/NEUROD1⁺ and eight CDX (21%) were ASCL1⁺/NEUROD1⁺. Where CDX were double positive by IHC, multiplex immunofluorescence revealed mutually exclusive intracellular expression of these NETFs (Fig. 3d). ASCL1/NEUROD1 status between baseline and progression was unaltered in five out of six models (Fig. 3a). At the time of writing there were no suitable antibodies to assess ATOH1 protein.

RE1 Silencing Transcription Factor (REST) is a master regulator of neurogenesis and a suppressor of NE differentiation in SCLC^{4,42,43} that directly targets ASCL1, CHGA and SYP⁴³. REST expression across the CDX biobank was either absent or low. Apart from CDX13 (86% REST-positive cells), the other six REST-expressing CDX had <10% REST-positive cells (Fig. 3a). CDX13 was negative for ASCL1, predicted by high REST expression (Figs. 2 and 3a), but low levels of REST across the CDX biobank precluded correlation of REST with known NE targets.

A scoring algorithm for NE classification aids characterization of SCLC⁴. Genes within this algorithm exhibited similar coexpression patterns in the CDX as seen in other preclinical models (Extended Data Fig. 4). Based on RNA-seq data with the exception of CDX13 (NE score, -0.14) all CDX had a positive NE score (Fig. 3e; range 0.28 (CDX22P) to 0.89 (CDX2, CDX18, CDX38) and 68% of CDX (26/38) had high NE scores of ≥ 0.80 (except CDX13 and CDX22P) in line with SCLC cell lines, GEMMs and resected tumors from patients with LS SCLC⁴. The low score for CDX22P (Fig. 3e, asterisk), which has classic morphology, expresses NE markers, with low REST, high ASCL1 and low NEUROD1 expression was unexpected. The negative score for CDX13 (Fig. 3c, asterisk) was expected as it has a variant morphology (Fig. 1c), high REST and lacks ASCL1 and NEUROD1 (Fig. 3a). The donor was diagnosed as having SCLC, with the pathology of CDX13 consistent with that diagnosis. Furthermore, this model is the only model that expresses POU2F3 (Fig. 3f) with the transcriptomic signature for the recently described Tuft Cell phenotype²², consistent with CDX13 belonging to the POU2F3 subtype (Extended Data Fig. 5). Further characterization of this model indicates that RB1 status is wild-type as shown by IHC (Fig. 3f) and confirmed by whole-exome sequencing, which also showed CDX13 was TP53 null (data not shown). While RB1 loss is common in SCLC, evidence that CDX13 represents a bona fide Tuft Cell variant (as opposed to a previously uncharacterized extreme nonNE SCLC variant) is supported by the SCLC Tuft Cell variant cell line NCI-H211 (one of four known Tuft Cell SCLC cell lines) that also expresses wild-type RB1 (ref. 44).

Expression of MYC gene family. To investigate MYC family gene expression in CDX, we assessed bulk tumor RNA expression for each gene (in absence of robust antibodies suitable for IHC for MYCN and MYCL⁴⁵). MYCL was most frequently expressed (Fig. 4a) with the highest expression in CDX17P. MYC and MYCN expression were relatively low; only two models express MYC (CDX13 and CDX17P) and three models express MYCN (CDX10, CD29 and CDX45; Fig. 4a). The relative distribution of MYC family genes is concordant with data from 81 resections from patients with SCLC² (Fig. 4b).

MYC protein was detectable in eight out of 38 (21%) CDX (Fig. 4c); in the six models where MYC transcripts were undetectable there were very low numbers of MYC positive cells (range 1–3%), and in CDX13 and CDX17P protein and RNA-seq data were concordant (Fig. 4a,b). Consistent with its variant, nonNE phenotype, in CDX13 63% of cells were MYC positive. MYC positive cells increased during progression from 1% in CDX17 to 10% in CDX17P where MYC was observed in ‘cell islands’ rather than dispersed throughout the tumors (Fig. 4d). MYC amplification is associated with more aggressive SCLC⁴⁶ and CDX17P is likely to represent clonal expansion of a small preexisting subpopulation present in CDX17.

MYC family gene amplification is widely reported as mutually exclusive in SCLC^{2,4,5}, although MYCL coexpression with either MYC or MYCN has been demonstrated in a few SCLC cell lines¹⁵. Assessment of MYC family gene coexpression in CDX tumors revealed a similar predominantly mutually exclusive pattern, with the exception of CDX17P that co-expressed MYCL and MYC, and CDX26 that coexpressed MYCN and MYCL (Fig. 4a), confirmed by western blot of CDX lysates (Fig. 4e).

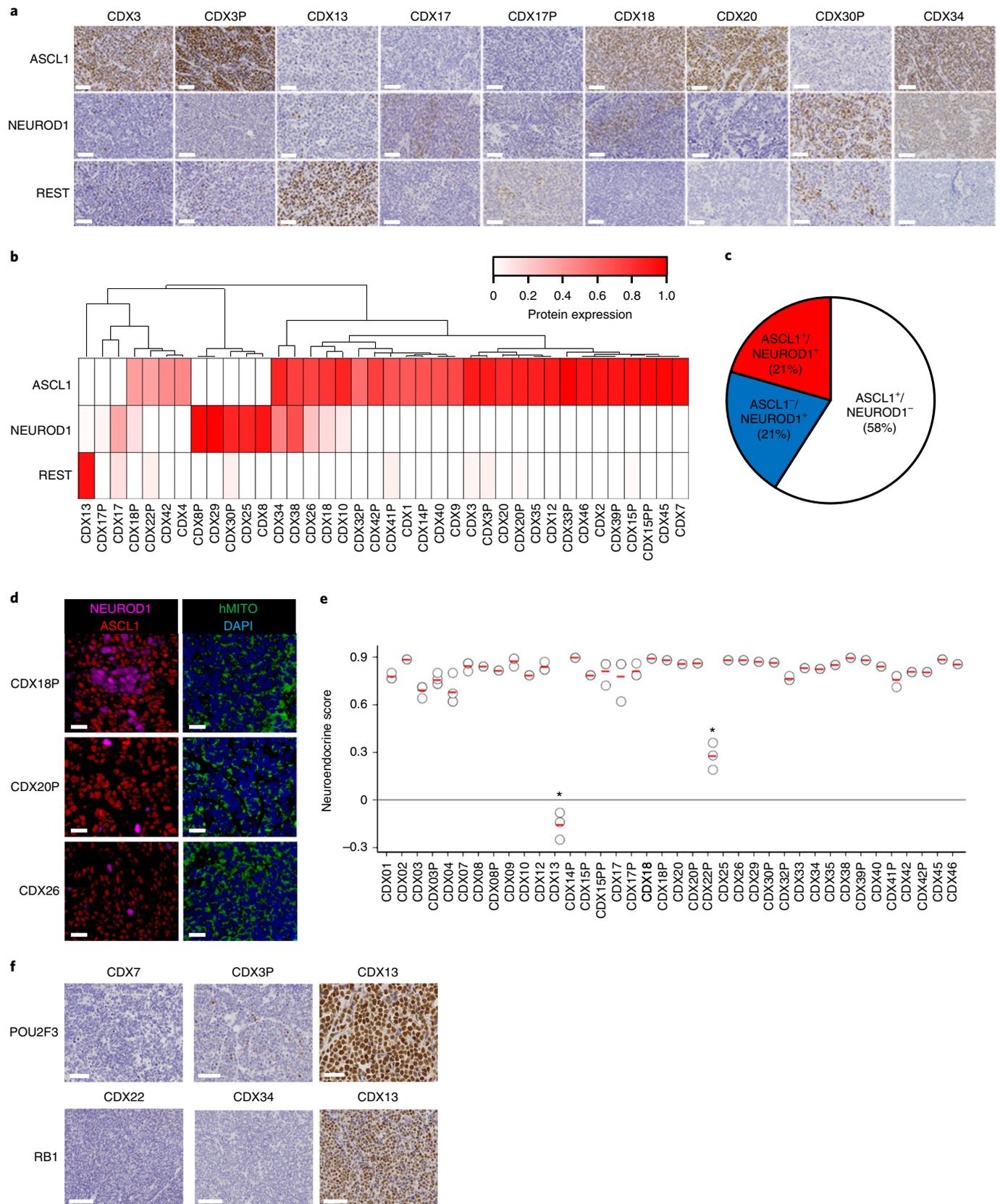
Classic SCLC is associated with expression of ASCL1 and MYCL¹⁵, whereas MYC associates with the variant phenotype in SCLC cell lines¹⁹. High MYC levels have also been associated with nonNE REST and POU2F3 and also with the reduced NE phenotype associated with NEUROD1 expression^{16,47}. In a GEMM, MYC drives an aggressive, variant form of SCLC associated with increased expression of NeuroD1 (ref. 21). Unsupervised clustering of CDX based on expression of these genes revealed an association between MYC, POU2F3 and REST (Extended Data Fig. 6). Highest MYCL expression was most closely associated with the ATOH1 subtype; although the ASCL1 subtype also expressed MYCL as in cell lines and GEMMs^{16,17}. NEUROD1 was more closely associated with MYCN rather than MYC, in contrast to the GEMM model accelerated by MYC²¹ (Extended Data Fig. 6).

Epithelial to mesenchymal transition (EMT). EMT is a feature of variant SCLC where loss of NE markers is associated with loss of epithelial cell markers including EpCAM and cytokeratins (CKs) and gain of the mesenchymal intermediate filament protein

Fig. 3 | Inter- and intratumoral heterogeneity of NE transcriptional regulators. **a**, Representative images of selected CDX models showing ASCL1, NEUROD1 and REST expression by IHC (brown stain). Scale bars, 50 μ m. In each case a whole-tumor section from three individual mice was analyzed. **b**, Heatmap showing hierarchical clustering and relative expression (0% positive cells, white, 100% positive cells, red) of ASCL1, NEUROD1 and REST in each CDX. For each IHC assay a whole-tumor section from three individual mice per CDX was analyzed, the mean value generated in HALO and shown on the heatmap. **c**, Pie chart showing the prevalence of SCLC subtypes regarding ASCL1 and/or NEUROD1 expression in CDX models. The pie chart was derived from data generated in **a**. **d**, Multiplex immunofluorescence assay showing mutually exclusive expression of ASCL1 (red) and NEUROD1 (purple) in CDX18P, CDX20P and CDX26. Human mitochondria antibody was also used (green) to show NEUROD1 and ASCL1 was present in cells of human origin. DAPI, blue. Scale bars, 20 μ m. Representative images are shown and each experiment constituted analysis of one whole-tumor section taken from three individual mice bearing CDX tumors. **e**, NE scores⁴ of each CDX derived from RNA-seq data (see Methods). All models have positive values confirming NE phenotypes with the exception of CDX13 (asterisk) with a negative score indicative that it is strongly nonNE. CDX22P (asterisk) is a noted outlier with a low but not negative NE score. Data were obtained from $n = 3$ independent tumors per CDX model (shown) and red lines represent the mean. **f**, CDX13 expresses POU2F3, the Tuft Cell marker and RB1 as determined by IHC (brown). Representative images are shown and for POU2F3 CDX7 and CDX3P are shown as examples of negative staining seen throughout the rest of the CDX panel. Scale bars, 50 μ m. For RB1, CDX22 and CDX34 are shown as examples of negative staining. Scale bars, 100 μ m. All analyses were performed on a whole-tumor section from two individual mouse tumors per CDX shown, with the exception of CDX13 RB1 staining, which was performed on a whole-tumor section from four individual mouse tumors.

vimentin^{4,48}. Variable expression of vimentin (using a human specific antibody) was present in 13 out of 38 CDX (34%, Fig. 5a,b) of which most contained <5% positive cells. CDX30P, CDX25 and CDX1 were 96, 58 and 9% positive for vimentin, respectively

(Fig. 5a,b). CDX13 that exhibits all known features of the variant, low NE subtype had undetectable vimentin and high expression of EpCAM and panCKs (Fig. 5a,b). With the exception of CDX42/CDX42P, hierarchical clustering of EPCAM, CK and VIM



expression showed that the paired models did not cluster together, indicating that changes had occurred during disease progression without consistent trends (Fig. 5b). CDX29 (matched CS-CTC count (EpCAM⁺/CK⁺) of zero) was the only model that did not express EpCAM/CKs or vimentin (Fig. 5a,b).

Principal component analysis of RNA-seq data shows an association between *ASCL1* (red circles) and *EPCAM*, while *NEUROD1* (blue circles) and *ATOH1* (orange circles) clusters associate with mesenchymal markers (Fig. 5b). The *ATOH1* and *VIM* association is concordant with protein expression and the two models with highest vimentin expression (CDX25 and CDX30P) are in the *ATOH1* cluster (Figs. 2 and 5).

Paired baseline and progression models. Longitudinal CDX facilitate assessment of SoC responses during disease progression (Fig. 6a). Chemotherapy response data were evaluated using a preclinical RECIST score^{49,50} and normalized time to progression (nTTP) for five baseline/paired progression models (CDX3/3P, CDX8/8P, CDX17/17P, CDX18/18P, CDX42/42P) and confirming previously published data for CDX8/8P^{37,38}. Baseline CDX tended to be more chemosensitive than their progression counterparts (Fig. 6). CDX20/CDX20P data are not yet available and relative sensitivity for CDX42/42P is inconclusive due to censoring of mice in the drug treatment arm of CDX42P (Fig. 6b).

To identify transcriptomic changes associated with progression and for future studies of therapy targets, differential expression analysis was conducted for the six paired CDX (Fig. 6c). Diverse changes in gene expression were observed during progression in each paired model, although few common changes across the six pairs were seen. Some trends toward differential gene expression during progression were observed across the six paired models (82 genes up-regulated and 270 genes down-regulated by paired statistical analysis); for example, up-regulation of NOTCH receptors in 4/6 pairs (NOTCH1 in three pairs: CDX8/8P, CDX20/20P and CDX42/42P and NOTCH2 in three pairs: CDX8/8P, CDX18/18P, CDX42/42P) (Fig. 6d). There were no consistent trends in *SLFN11* (ref. ⁹), the *MYC* regulon³³ or *WNT* family genes⁵¹, all previously reported as associated with drug resistance (data not shown).

CDX metastasis. Metastasis is a frequent and early event in SCLC⁵². Although not studied in all CDX models, mice bearing subcutaneous (s.c.) CDX14P, CDX17, CDX17P and CDX25 tumors showed evidence of metastatic dissemination (Fig. 7), with macro-metastases present in the brain and liver of mice bearing CDX14P (Fig. 7a). Dual immunofluorescence for NCAM and phospho-histone H3 (pHH3) demonstrated disseminated proliferating cells in the brain of a CDX14P-bearing mouse and liver metastases in a CDX17-bearing mouse (Fig. 7b). In all four models disseminated NCAM positive human cells were detectable within the mouse lungs (Fig. 7c).

Discussion

An increasing appreciation of SCLC phenotypic heterogeneity is emerging from studies of cell lines, GEMMs, patient-derived

preclinical models and patients' tumor specimens^{4,53}. The ability to subtype SCLC is an important first step to stratified treatment approaches, particularly now that new therapies are approaching and/or have entered early clinical trials where biomarkers are urgently required^{23,28}. Unbiased transcriptomic analysis of our CDX biobank revealed four molecular subtypes (Fig. 2): in agreement with previous studies the majority subgroup was dominated by *ASCL1*, with a smaller group expressing *NEUROD1* and a single model, CDX13, expressing *POU2F3*. Cumulative data regarding variant morphology (Fig. 1c), lack of NE marker expression (Fig. 3a and Extended Data Fig. 3), negative NE score (Fig. 3e), high *POU2F3* expression by IHC (Fig. 3f) and RNA-seq (Fig. 2b) and high expression of *SOX9*, *TRPM5* and *AVIL* compared to all other models (Extended Data Fig. 5) convincingly classify CDX13 as belonging to the SCLC Tuft Cell subtype²². Further characterization of CDX13 confirmed that it was null for *TP53* and wild-type for *RB1* with readily detectable RB1 protein expression (Fig. 3f). The short overall survival of the CDX13 donor is of interest given the proposed link between poor prognosis and wild-type *RB1* in SCLC⁴⁴.

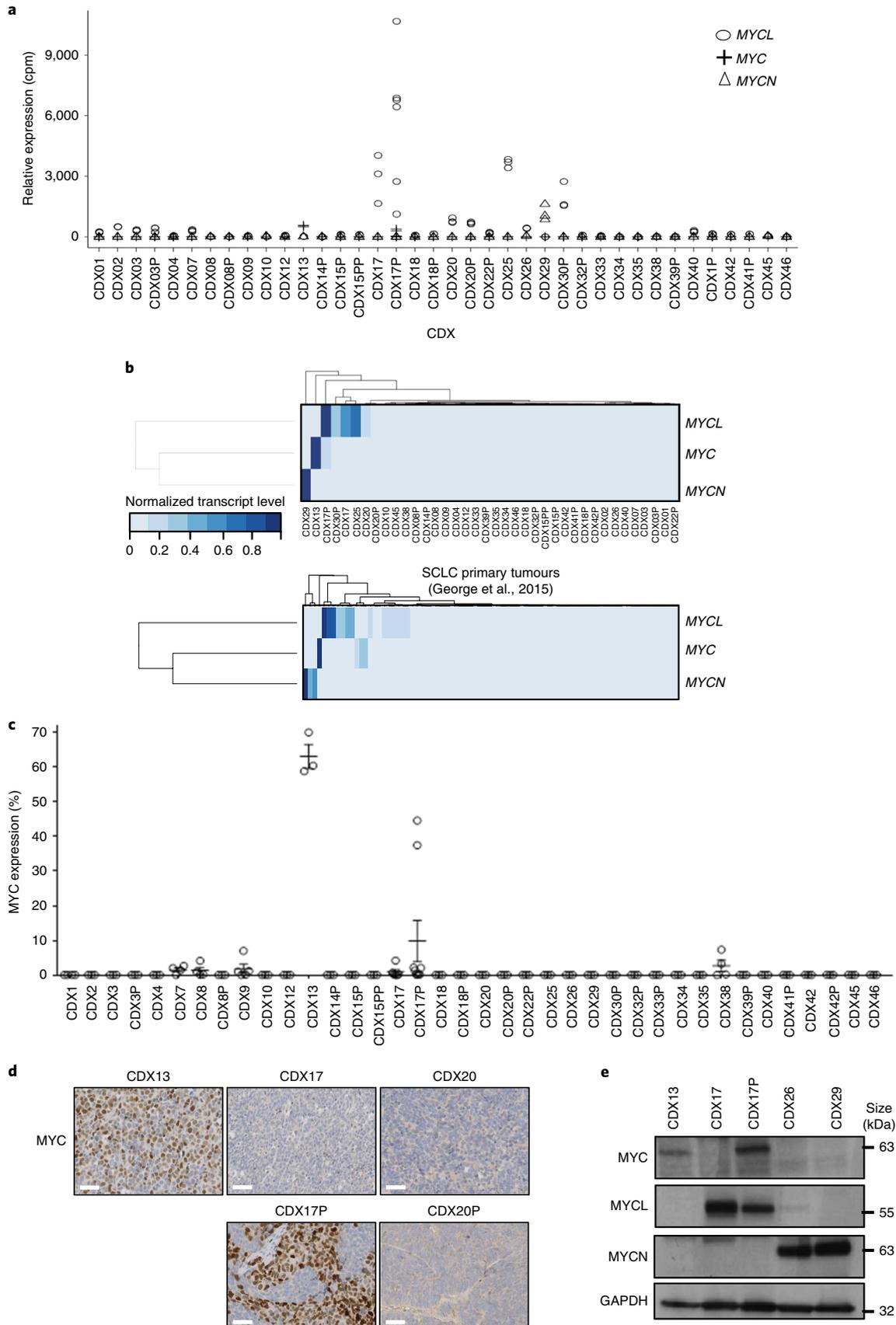
We also describe a previously unrecognized subgroup of SCLC based on high *ATOH1* expression and its reported downstream target *POU4F3* (Fig. 2). *ATOH1* is under-represented in other datasets^{2,6}, and present in only ~1% (1/81) surgically resected SCLC samples². Although larger datasets are required to confirm these findings, the higher prevalence of *ATOH1* in CDX may reflect their CTC origin with a potential for clonal selection and expansion in vivo, and/or in combination with the fact that these CDX donors had extensive stage rather than resectable disease, reflecting a hitherto unrecognized phenotypic diversity. High *YAP1* expression was used to define a SCLC subtype^{23,54}. *YAP1* expression is either extremely low (CDX4, CDX22P and CDX42P) or absent in our CDX biobank and *YAP1* expressing CDX do not correspond with *ATOH1* models (Extended Data Fig. 7a). *ATOH1* expression does not correlate with *YAP1* in either the surgically resected sample from the patient with SCLC reported by George et al., or in SCLC cell lines from the Cancer Cell Line Encyclopedia (CCLE)⁵⁵ (Extended Data Fig. 7b,c). Although most studies of *ATOH1* relate to its spatiotemporal expression required to drive differentiation and long term viability of auditory hair cells⁴⁰, it is also expressed in the intestinal epithelium where it plays an essential role in development of secretory cell lineages and control of proliferation^{56,57}. *ATOH1* is regulated by NOTCH and WNT signaling in prosensory auditory cells^{58–60} that are pertinent pathways in SCLC biology. *ATOH1* has been implicated in development of various cancers including medulloblastoma and Merkel cell carcinoma (a NE tumor), where it appears to act as an oncogene, while in colorectal cancer it resembles a tumor suppressor gene⁶¹.

Intratatumoral heterogeneity was evident at the level of *ASCL1*/*NEUROD1* expression in multiple CDX. IHC identified three subtypes (*ASCL1*⁺/*NEUROD1*⁻, *ASCL1*⁺/*NEUROD1*⁺, *ASCL1*⁻/*NEUROD1*⁺) and consistent with previous studies, CDX models with *ASCL1* expression (with or without co-expressed *NEUROD1*) were most common (Fig. 3). Intratumoral heterogeneity was observed

Fig. 4 | Expression of MYC family members. **a**, Normalized expression levels (cpm) of *MYC*, *MYCL* and *MYCN* RNA in CDX models from RNA-seq analysis from $n=3$ independent tumors per model. **b**, Heatmap showing normalized expression of *MYC* family genes across the CDX models and within the data reported in ref. ². **c**, *MYC* protein levels across the CDX panel quantified using HALO Software (see Methods). $n=3$ whole-tumor sections (biological replicates, different animals) were analyzed per CDX, with the exception of CDX17P where $n=6$ independent CDX tumors were analyzed. Mean values are shown with error bars representing \pm s.e.m. **d**, Immunohistochemistry of *MYC* for selected CDX: CDX13 has the highest *MYC* expression in the panel; an example of increasing *MYC* expression over time, from CDX17 (low expression) to CDX17P (intermediate expression) where focal *MYC* positive cells are observed; and CDX20 and CDX20P, a second pair of longitudinal models with no up-regulation of *MYC*. Scale bars, 50 μ m. Images are representative of analysis of three whole-tumor sections, with the exception of CDX17P where six whole-tumor sections were analyzed. **e**, Western blots for *MYC*, *MYCL* and *MYCN* in whole-CDX-tumor lysates from those CDX where RNA expression (in **b**) implies more than one family member is expressed (CDX17P and CDX26), or with high expression of *MYC* (CDX13) or *MYCN* (CDX29). Blots are representative of lysates from three independent tumors per model.

within ASCL1⁺/NEUROD1⁺CDX, showing mutually exclusive regions of ASCL1 and NEUROD1 positive cells that have not previously been noted (Fig. 3d). Whether and how this intratumoral

heterogeneity in NE transcriptional profiles affects or is affected by chemotherapy or changes during disease progression is under continued investigation. Drapkin et al.³³ found no evidence for



differential responses to SoC chemotherapy in *ASCL1* versus *NEUROD1* expressing patient-derived models, although there is evidence of cell lines gaining *NEUROD1* and a more mesenchymal phenotype after treatment⁶². Several examples of altered *NETF* expression in paired CDX were observed; such changes should be considered if future therapies are targeted to specific and distinct *NETF* regulated pathways (Fig. 3a), as was recently demonstrated for the antitumor effects of *LSD1* inhibition due to *NOTCH*-induced suppression of *ASCL1* (ref. 63). NE scores across the CDX panel (Fig. 3e) more closely resembled established cell lines (with values close to 1) than previously reported for samples from patients with *SCLC*⁴ (maximum value 0.89). We speculate that particular *SCLC* phenotypes are ‘fit’ enough to propagate in vitro to become established cell lines and overlapping elements of that ‘fitness’ enables survival in the circulation and tumorigenicity in mice.

The effects of different *MYC* family genes on *SCLC* behavior are beginning to be explored. With the exception of the *Rb*^{-/-}/*p53*^{-/-}/*Myc*^{T58A} (RPM) GEMM²¹, *SCLC* GEMMs express high levels of *Myc*^{18,64,65} that is required for tumor progression^{17,65}. *MYCL* was the most prevalent *MYC* family member expressed across the CDX biobank (Fig. 4). *MYC* family gene amplification in *SCLC* is thought to be mutually exclusive^{2,5,6}, however, *CDX17P* expressed both RNA and protein for *MYC* and *MYCL* (Fig. 4). *MYC* is associated with a variant morphology, a nonNE phenotype and closer linkage to *NEUROD1* rather than *ASCL1* (ref. 21). As predicted, unsupervised clustering of RNA-seq data revealed that the nonNE *CDX13* expressed high levels of *MYC*, while high *MYCL* was more closely associated with *ATOH1* than *ASCL1* (refs. 16,17), although *MYCL* was also expressed in *ASCL1* models. *MYCN* was only associated with *NEUROD1* models (Extended Data Fig. 6) although the small number of *MYCN* expressing models preclude robust conclusions. We hypothesize that although functional redundancy has been reported for murine *Myc* and *Mycn*⁶⁶, our data support a model in which *MYCL* and *MYCN* may mediate related but distinct NE programs, while *MYC* regulates a NonNE program in human *SCLC*. Studies are ongoing to determine whether *CDX13* responds to Aurora Kinase A/B inhibition, as predicted by its elevated *MYC* expression^{21,28}.

EMT provokes lack of cell-cell adhesion and substrate adherence and enhanced migratory and invasive capacity, collectively forming a hallmark of cancer⁶⁷ and may be particularly relevant in *SCLC* where metastasis is early and frequent⁵². In the *p53*^{-/-}/*Rb1*^{-/-} GEMM, a mechanism of cooperation between NE and mesenchymal-like nonNE cells is required for distant metastasis and EMT has been associated with the nonNE, variant subtype⁴⁸. Both *NEUROD1* and *ATOH1* are associated with mesenchymal markers in CDX models (Fig. 5c); *NEUROD1* is linked with migration and invasion in *SCLC*²⁰ and *ATOH1* associated with *VIM* RNA and protein expression (Fig. 5). Further studies will explore the propensity for EMT in *ATOH1*-high CDX. However, lack of vimentin expression in *CDX13*, *CDX17P* and *CDX29* suggests that a mesenchymal phenotype and variant morphology are not obligatorily linked. Multiple CDX underwent metastasis in vivo forming tumors at distant sites, notably those commonly seen clinically

including brain, liver and lungs and provides a patient-derived model of *SCLC* that metastasizes to the brain (Fig. 7) that was previously unavailable. Adopting the simplified definition of EMT (loss of CKs/*EpCAM* and gain of vimentin), the lack of evidence of EMT in *CDX14P* (*ASCL1*⁺/*NEUROD1*⁻, low/no *MYC* family gene expression) that routinely metastasizes to the brain shows that for this model at least, EMT may not be required for metastasis. The hypothesis that EMT associates with chemotherapy resistance⁶⁸ is also not clearly borne out by data generated in longitudinal CDX that acquire chemoresistance (Fig. 6a,b) and of the 15 post-chemotherapy progression models, only *CDX30P* had high vimentin and low CKs, consistent with chemotherapy-resistant EMT. *CDX4* (*ASCL1*⁺/*NEUROD1*⁻) with low *MYC*, classic morphology and no evidence of EMT is one of the most inherently chemorefractory CDX within the biobank³⁸.

The comparative tumorigenicity of *EpCAM* positive and *EpCAM* negative human *SCLC* CTCs in immune-compromised mice is unknown as we do not select for either when CTCs are enriched for implantation. We had surmised that there might be a correlation between the *EpCAM* positive CS-CTC count made in the matched blood sample and our ability to generate a CDX model and this was the case overall (Extended Data Figs. 1 and 3). However, the donor of *CDX29* had zero CS-CTCs suggesting additional factors beyond *EpCAM* expression contribute to tumorigenicity that warrant further study. Overall, our data are consistent with the notion that EMT (defined by vimentin high, *EpCAM*/*CK* low) is not required for CDX generation, is not always observed in nonNE cells, is not required for metastasis and is not the sole mechanism underpinning chemoresistance. The caveat that EMT is dynamic leaves the possibility that EMT occurs before our observations and is followed by a reversing mesenchymal to epithelial transition.

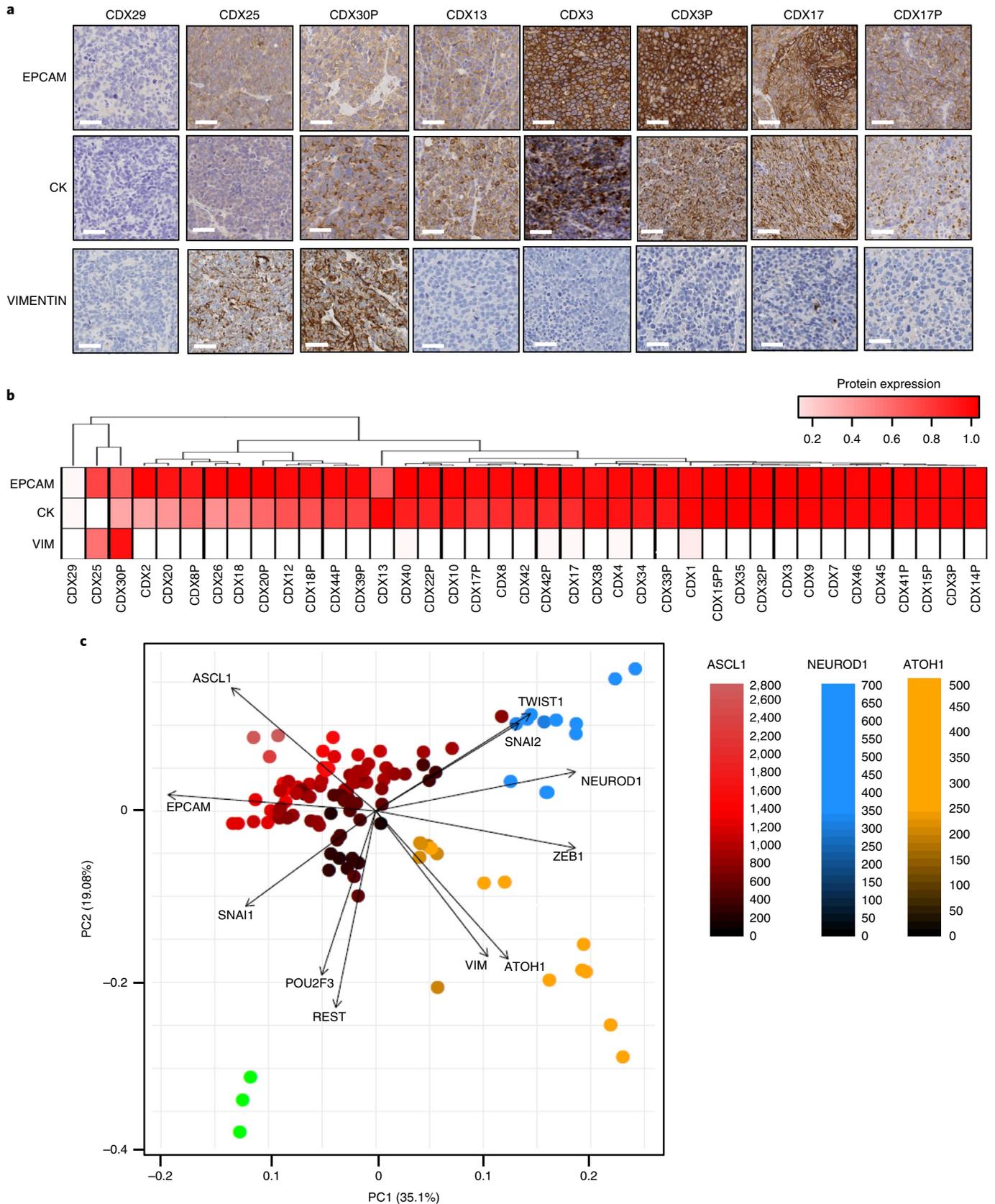
The important advantage of CDX models is the ability to make longitudinal models. In vivo studies treating CDX bearing mice with combination cisplatin/etoposide designed to mimic clinical SoC treatment show good concordance between preclinical response and patient overall survival (Fig. 6a,b). Furthermore, in paired models, the relapse model was generally more resistant to therapy than the baseline model, reflecting the advanced status of the patients’ disease (Fig. 6a) and supporting their use for pre-clinical modeling in the absence of patient biopsies. Transcriptomic analysis of the six longitudinal CDX pairs identified *NOTCH1* and *NOTCH2* up-regulation at progression (Fig. 6d) warranting further study. Although inactivating *NOTCH* mutations have been identified in up to 25% of *SCLC*², endogenous *NOTCH* activity activates a NE to nonNE transition in the RPP GEMM associated with the emergence of chemoresistance⁴⁵.

Overall, our CDX biobank recapitulates many previously reported features of *SCLC* with caveats that they are s.c., lacking the influence of the lung microenvironment and within immune-compromised hosts. Nevertheless, they represent a broad range of phenotypes encompassing established classic and variant morphologies, a range of *NETF* expression and the Tuft Cell lineage. This biobank highlights the complexity of CDX models compared to *SCLC* cell lines and GEMMs^{53,69,70} and confirms CTCs as

Fig. 5 | EMT gene expression. **a**, IHC (brown stain) for selected CDX of the epithelial markers *EpCAM* and panCK and the mesenchymal marker vimentin. Scale bars, 50 μm. Representative images are shown where a whole-tumor section from each of three independent CDX tumors per model was analyzed. **b**, Heatmap showing hierarchical clustering and relative expression (0% positive cells, white, 100% positive cells, red) of *EpCAM*, CK and *VIM* for each CDX. For each assay a whole-tumor section from three individual mice per CDX were analyzed and the mean value was generated in HALO and shown on the heatmap. **c**, Principal component analysis of RNA-seq expression in the CDX models, on 11 selected genes. Points represent individual CDX, colored according to normalized *ASCL1* (red), *NEUROD1* (blue), *POU2F3* (green) or *ATOH1* (orange) expression. Arrows correspond to PCA loadings for the four transcriptional factors, the epithelial marker *EPCAM*, mesenchymal markers *VIM*, *ZEB1*, *TWIST1* and *SNAI2*, and the nonNE marker *REST*.

a valuable alternative source of primary patient material. We have identified a previously unrecognized SCLC subgroup, designated *ATOH1*, and show that coexpression of MYC family proteins can

occur within a CDX. We did not demonstrate a convincing role for EMT in CDX although *ATOH1* expressing CDX were associated with high vimentin expression and metastases were present in two



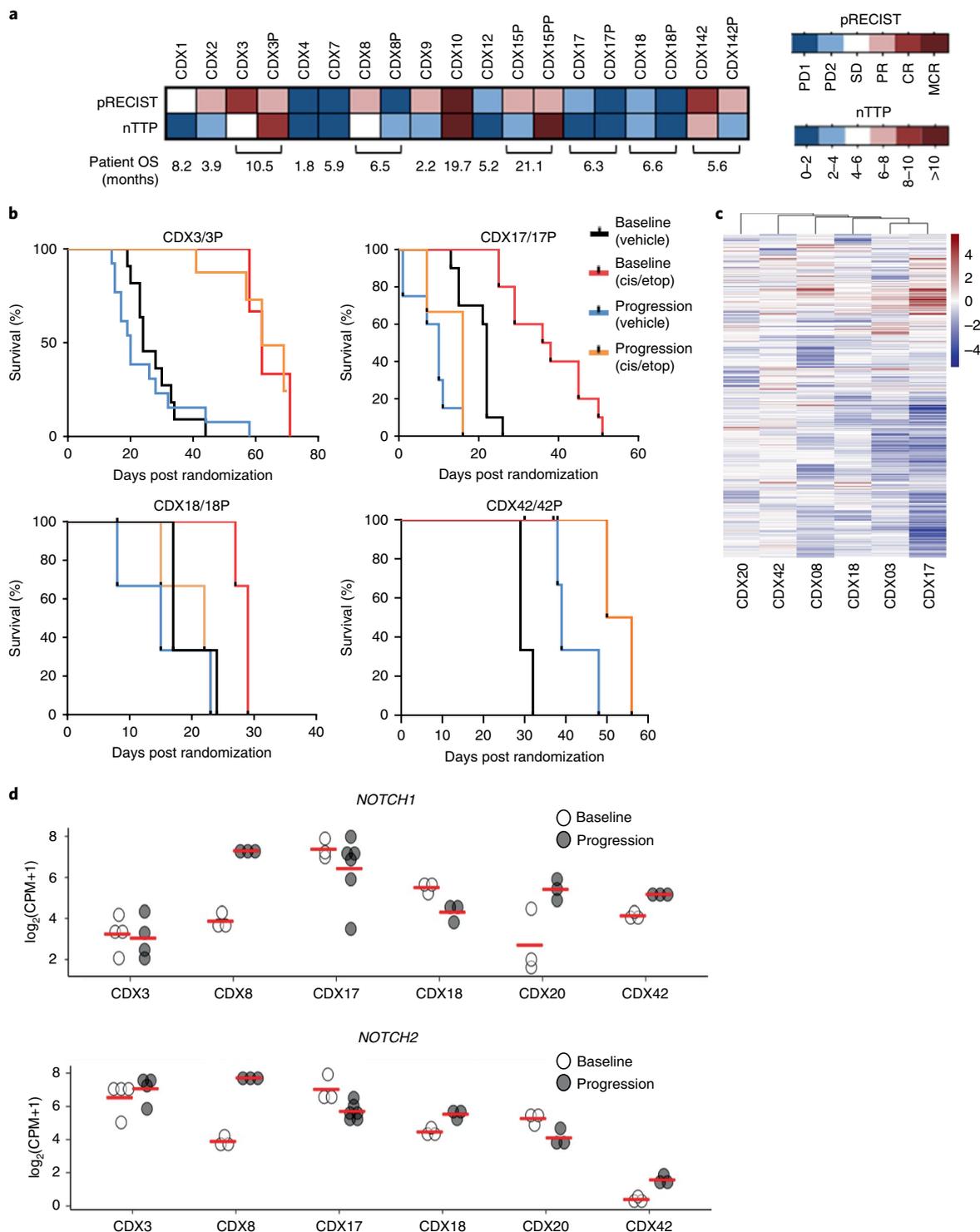


Fig. 6 | Analysis of paired CDX models. a, CDX Response to cisplatin/etoposide in vivo. Chemosensitivity of a range of CDX models according to pRECIST criteria and nTTP (months) reveal good concordance with patient overall survival (OS). Key: dark blue, progressive disease 1 (PD1) and nTTP, 0–2 months; light blue, progressive disease 2 (PD2) and nTTP, 2–4 months; white, stable disease (SD) and nTTP, 4–6 months; pink, partial response (PR) and nTTP, 6–8 months; light red, complete response (CR) and nTTP, 8–10 months and dark red, maintained complete response (MCR) and nTTP, >10 months. Number of mice per treatment group per CDX model were as follows: CDX12, CDX15P, CDX15PP, CDX18, CDX18P, CDX42, CDX42P, $n=3$; CDX7, $n=4$; CDX9; CDX17P, $n=6$; CDX17, $n=10$; CDX1, CDX3, CDX3P, CDX8P, CDX10, $n=11$; CDX2, $n=12$; CDX8, $n=13$ and CDX4, $n=14$. **b**, Kaplan-Meier survival curves for CDX17/CDX17P, CDX18/CDX18P and CDX42/CDX42P shows relative chemosensitivity of paired baseline and relapse models. Number of mice per vehicle and drug treated groups, CDX17, CDX17P, $n=10$ per group; CDX18, CDX18P, CDX42, CDX42P, $n=10$ per group. cis, cisplatin; etop, etoposide. **c**, Heatmap displaying the fold change in transcript level of differentially expressed genes between baseline and progression for each of the six paired CDX models. Patient variation was controlled, and the variance stabilizing transform of the mean counts was used for visualization. **d**, Increased *NOTCH1* and *NOTCH2* expression in progression models compared to their matched baseline. Circles represent independent CDX tumors; consisting of $n=3$ independent CDX tumors per model, with the exception of CDX3 and CDX3P, which had $n=4$ independent replicates. Red bar, mean value.

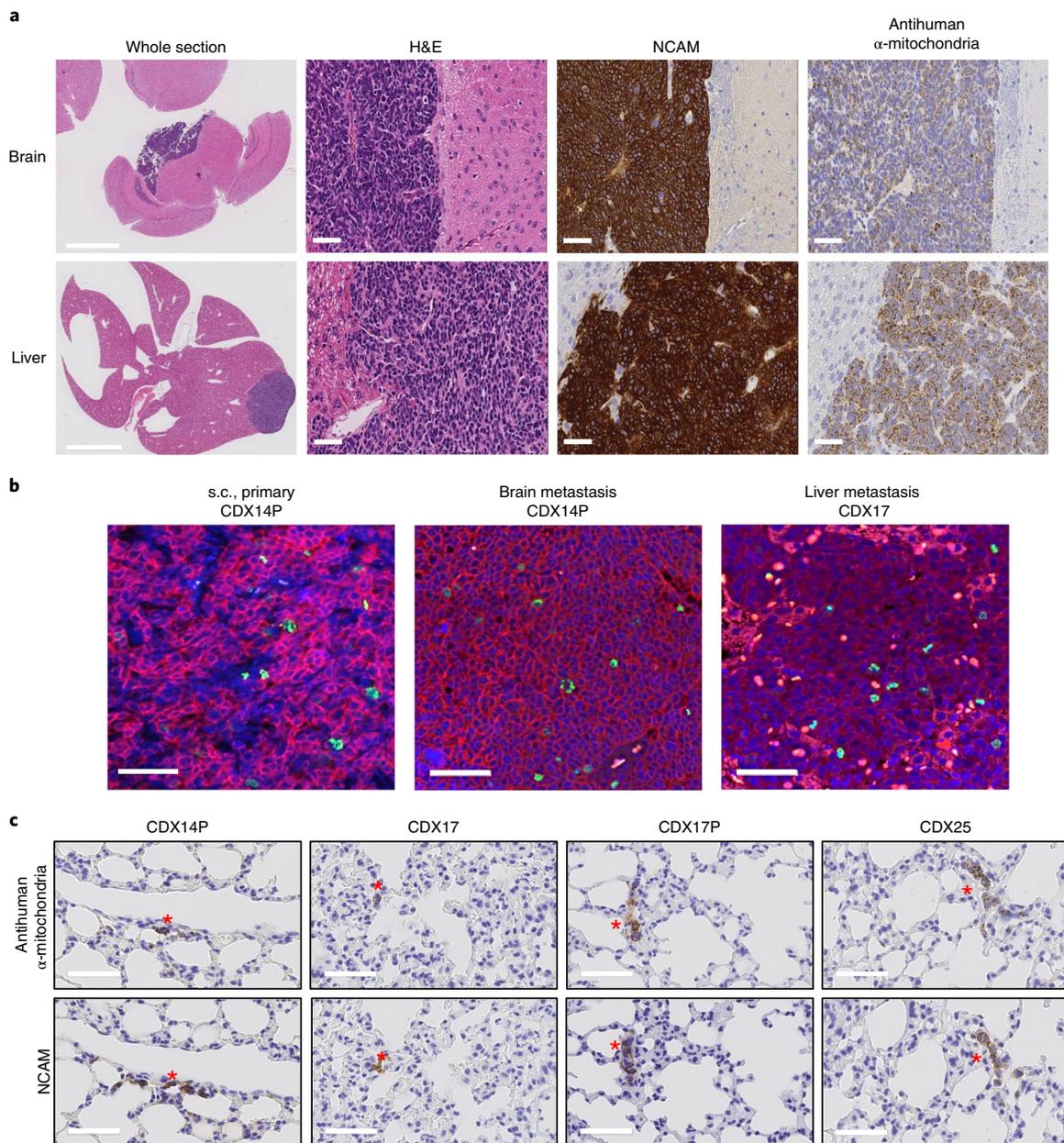


Fig. 7 | Metastasis of s.c. CDX cells to brain, liver and lung. a, Whole-organ sections and selected areas of H&E stained mouse brain and liver after resection of CDX14P grown s.c. on the flank to 1,200 mm². Tumor cells are indicated by the presence of the NE marker NCAM and antihuman α -mitochondrial staining (IHC, brown stain). Scale bars for whole-tumor sections of the brain, 2 mm and of the liver, 5 mm; all other panels, scale bars, 50 μ m. Three animals were resected when the s.c. tumor reached 1,200 mm³ and were culled after an additional 43, 115 and 130 d due to welfare/s.c., tumor regrowth to 1,200 mm³ and representative images are shown. **b**, Immunofluorescence analysis of the s.c. CDX14P tumor and brain metastases arising after its resection, and of a CDX17 liver metastasis detected on necropsy when the s.c. tumor had reached maximum volume (1,200 mm³). Cell nuclei are DAPI stained blue, NCAM positive NE cells are stained pink and proliferating cells are indicated by the presence of pHH3 foci stained green. Scale bars, 50 μ m, bottom left panel. A whole-tumor section from three individual tumors per CDX were analyzed and representative images are shown. **c**, Clusters of disseminated human tumor cells (marked by red asterisks) detected by immunohistochemistry of NCAM and human α -mitochondria (brown stain) present in lungs of mice bearing s.c. CDX14P, CDX17, CDX17P and CDX25. Scale bars, 50 μ m, bottom left panel. A whole-tumor section from three individual tumors per CDX were analyzed and representative images are shown.

of these models (CDX25 and CDX17P), which may implicate an aggressive phenotype linked to mesenchymal gene expression for further interrogation.

Only 10 ml of a patient's blood is taken to make CDX. This approach is being reproduced in many research centers worldwide and, with this biobank, forms a valuable addition to the preclinical models available for SCLC research. Detailed histopathology with

dovetailed gene expression studies in longitudinal models will support studies addressing important clinical questions: mechanisms of metastasis, development of chemotherapy resistance and the effects of intratumoral heterogeneity on therapy responses. They will also support an evaluation of the potential to stratify SCLC for personalized medicine with parallel development of predictive and response biomarkers.

Methods

Patient samples. The 31 patients described in this study had samples and (resultant data) obtained between February 2012 and August 2016 following informed consent and according to ethically approved protocols: European Union CHEMORIS FP6 Contract number LSHC-CT-2007-037665 (NHS Northwest 9 Research Ethical Committee) and The TARGET (tumor characterization to guide experimental targeted therapy) study, approved by the North-West (Preston) National Research Ethics Service in February 2015, reference 15/NW/0078. Patient metadata can be found in Supplementary Table 1.

CDX generation. CDX models were generated as previously described³⁷ and the workflow is summarized in Fig. 1a. In brief, EDTA blood from each donor patient (10 ml) was enriched for CTCs by negative selection of blood cells using RosetteSep (negative selection with antibodies against CD2, CD16, CD19, CD36, CD38, CD45, CD66b and glycophorin A, product no. 15167, Stem Cell Technologies) and s.c. injected into the flank of a 8–16 week old nonobese diabetic (NOD) severe combined immunodeficient (SCID) interleukin-2 receptor γ -deficient (NSG) mouse. When passage one (p1) CDX tumors reached 600 mm³ they were harvested and dissected into 3 × 3 mm³ fragments and each fragment surgically reimplanted s.c. into five NSG mice, designated p2 tumors. To generate sufficient material for histological and molecular characterization and enable future studies p2 tumors were first dissociated by gentleMACS dissociation (Miltenyi Biotec Ltd) and then depleted of red cells, dead cells and mouse cells according to manufacturer's instructions. In brief, tumors were cut into 2–4 mm pieces and transferred to a gentleMACS C tube containing 200 μ l of enzyme H, 100 μ l of enzyme R and 25 μ l enzyme A from the MACS tumor dissociation kit (Miltenyi Biotec Ltd) in 4.7 ml serum free RPMI medium. Tumors were dissociated using the Program h_tumor_01 on the gentleMACS (Miltenyi Biotec catalog no. 130-093-235). The sample was then incubated for 30 min at 37 °C under continuous rotation followed by further dissociation by the program h_tumor_02 gentleMACS program and a second incubation for 30 min at 37 °C under continuous rotation. A final dissociation was performed using the gentleMACS h_tumour_03 program run twice. Samples were resuspended in RPMI medium then passed through a 70 μ m cell strainer (BD catalog no. 352350) before washing twice in serum free RPMI. Cells were pelleted by centrifugation at 500g for 5 min and the remaining red blood cells were lysed in 1 × red blood cell lysis buffer (G-Biosciences) at room temperature for 5 min followed by stoppage by the addition of 30 ml PBS. Cells were then mixed 50:50 with trypan blue and a live/dead cell count obtained. Depletion of dead cells and mouse cells was performed using the MACS dead cell removal kit (Miltenyi Biotec Ltd) and the antimouse IgG2a + b (Miltenyi Biotec Ltd) kit, respectively. After dissociation and counting, samples were pelleted at 500g 5 min and resuspended in 200 μ l of 1 × dead cell binding buffer containing 100 μ l of prelabeled mouse cell beads and 100 μ l of dead cell beads per 10⁷ cells. Samples were incubated at room temperature for 15 min then eluted using a MACS LS column. Following counting, cells were then reimplanted into five NSG mice, designated p3.

Metastatic spread was evaluated in all organs following full necropsy after euthanasia of animals when the s.c. tumors reached maximum volume (1,200 mm³). CDX14P was an exception; when s.c. tumors reached 1,200 mm³ they were surgically removed and animals kept on study, these animals were culled either due to evident tumors reaching maximum volume or for welfare reasons. Metastases were then assessed after full necropsy.

All procedures were carried out in accordance with Home Office Regulations (UK), the UK Coordinating Committee on Cancer Research guidelines and by approved protocols (Home Office Project license 40-3306/70-8252 and Cancer Research UK Manchester Institute Animal Welfare and Ethical Review Advisory Board). In vivo studies were conducted to comply with The Animal Research: Reporting of In Vivo Experiments Guidelines. A parallel blood sample was also taken for CTC enumeration using the CellSearch CTC platform (CS, Menarini Silicon Biosystems) that captures EpCAM expressing cells and classifies CTCs as cells with an intact DAPI stained nucleus staining positively with a panCytokeratin (panCK) antibody and negatively for the white blood cell marker CD45.

CDX samples are available on reasonable request via discussion with the corresponding author (CDive).

In vivo drug studies. Therapeutic studies were carried out essentially as previously described³⁸. A total of 100,000 viable CDX cells in 100 ml 1:1 RPMI:matrigel were injected s.c. into the right flank of 8–10-week-old (20–25 g) female NSG mice. Mice were randomized when tumors reached 150–250 mm³, to be treated each morning with vehicle or cisplatin and etoposide as described in³⁷, with the following modifications: Three or more mice per cohort received from 1–3 cycles of cisplatin/etoposide dependent on tolerability. In all cases, 1 ml saline was administered on cycle 1, day 1 to improve platinum tolerance⁹. Mice were observed for a period of time after dose to ensure no adverse effects were seen. Tumors were measured twice a week by caliper until they reached four times initial tumor volume ($4 \times ITV$), as determined by the formula $V = (L \times W^2)/2$, where V is volume, L is length and W is width, or until animal health deteriorated.

Histology, immunohistochemistry and digital pathology. Histology and immunohistochemistry was performed on formalin-fixed paraffin embedded

tissue from CDX, 4 μ m sections were stained using antibodies detailed in Supplementary Table 3. Antibodies to Vimentin and to REST were human specific, removing ambiguity about contribution from tumor infiltrating mouse stromal and other mesenchymal murine cells. Chromogenic IHC assays were performed on a Leica Bond Max Platform with Bond Polymer Refine Detection kit (DS9800) or on a Roche Ventana Ultra with UltraMap DAB IHC Detection kit (760-151). Multiplex immunofluorescence staining was performed on a Leica Bond RX using PerkinElmer Opal 4-Color Automation IHC Kit (NEL800001KT). Isotype controls were performed on representative tissue sections to ensure antibody specificity. Whole sections were scanned and digitized using a Leica SCN400 for IHC, a Leica Versa was used for immunofluorescence. A minimum of three biological replicates per chromogenic assay were performed on all samples with the exceptions of one biological replicate for TTF1 (CDX4 and 33-45) and two replicates for Cytokeratin (CDX12), VIMENTIN (CDX15P, 20, 20P and 32) and EPCAM (CDX18). POU2F3 expression was examined in three biological replicates of CDX13, where RNA-seq indicated that POU2F3 would be highly expressed, as well as in an additional nine randomly selected CDX models as experimental controls. Chromogenic staining was quantified using Definiens Developer XD v.2.7.0 Tissue Studio v.4.4.2 (Definiens AG). Regions of interest (ROI) within tissue sections were identified using Definiens Tissue Studio via machine learning across pathological samples and tissue controls to define the full range of contrast. Within these ROI, tumor cells were detected and classified as positive or negative based on IHC thresholds. Expression level was quantified as percentage positive cells within the ROI. For each CDX, sections were manually and independently scored by two experienced analysts. A NE score for each CDX was calculated as described previously⁴.

RNA-seq analysis. RNA was extracted from 3–6 independent replicate tumors per CDX and RNA-seq was performed as previously described³⁸. Data were filtered to remove reads aligning preferentially to the mouse genome using the 'bamcmp' algorithm, previously described⁷¹. Reads were aligned to the GRCh38 assembly from Ensembl and counts per million (cpm) were determined⁷². Clustering of CDX models was generated using the Pearson correlation to generate distances and the average clustering method, using the R stats package. Heatmaps were generated using mean cpm from each set of replicates, transformed into normalized counts. Data were plotted using 'heatmap.2' from the R package 'gplot' with default parameters, unless otherwise stated. A model of mean expression versus coefficient of variation was fitted for all CDX models. The 1,912 genes that showed a significant variance, of which 1,686 were protein coding, from this fitted model ($P \leq 0.001$) were used as input for the clustering algorithm. Differentially expressed genes in each cluster were ranked and evaluated to begin to explore their potential for driving the molecular profiles of each cluster. Each cluster was then labeled based on conformance with known drivers of NE gene expression and/or established roles in SCLC⁴.

Gene expression analysis. Differential gene expression between baseline and progression models was performed using DESeq2 (ref. ⁷³) with log fold changes shrunk using apeGLM⁷⁴, incorporating patient as a factor in the design matrix. Genes were filtered on having an absolute fold change of at least 1, and an adjusted P value of ≤ 0.05 .

Reverse transcriptase-qPCR. RNA was isolated from RNAlater-treated or snap frozen tissue via the Tissuelyser LT and RNeasy mini kits according to Qiagen recommendations. cDNA synthesis was performed with the High-Capacity cDNA Reverse Transcription Kit (ThermoFisher Scientific). RT-qPCR was performed using Taqman gene expression master mix and gene expression assays for ATOH1 (Hs00944192_s1), POU4F3 (Hs00231275_m1), GAPDH (Hs02786624_g1) and ACTB (Hs01060665_g1) according to the manufacturer's recommendations. Data were analyzed with the dCt method by normalizing to ACTB and GAPDH housekeeping genes.

Statistics and reproducibility. With the exception of primary patient material and p1 mice generated from CTCs where only one sample was available, all analyses were performed on minimum $n = 3$ independent CDX tumor replicates or in the case of cell lines on three independent cell cultures per cell lines. In vivo responses to SoC were carried out ranging from $n = 3$ to 14 animals per group depending on availability of animals bearing CDX tumors at the time of randomization, with group sizes based on data in ref. ⁷⁵. Details of exact n numbers and statistical analyses used, including P values can be found in the relevant figure legends. All RNA-seq data presented were cross-checked against their respective passage 1 CDX tumor, and the respective donor's germline where available using NGSCheckMate⁷⁶. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

RNA-seq data that support the findings of this study (Figs. 2, 3e, 4a,b, 5c and 6c,d and Extended Data Figs. 2, 4 and 5–7a) have been deposited in the EMBL-EBI

ArrayExpress database under accession code *E-MTAB-8465*, with the title 'RNA of Small Cell Lung Cancer Circulating Tumor Cells Derived Explants', and can be accessed at <http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-8465>. Source data for Fig. 4c and Extended Data Fig. 7b can be accessed from ref. ² and source data for Extended Data Fig. 7c are publicly available from The Broad Institute CCLE at <https://portals.broadinstitute.org/ccle>. Source data for Figs. 3–5 and Extended Data Figs. 2 and 3 are presented with the paper. All other data supporting the findings of this study are available from the corresponding author on reasonable request.

Code availability

No algorithms or software were developed in this study. Software that was used is free and open source and details on acquiring them can be found in the associated references. All code that was used to generate the figures can be found at <https://doi.org/10.5281/zenodo.3574846>.

Received: 14 August 2019; Accepted: 26 February 2020;

Published online: 13 April 2020

References

- Bunn, P. A. Jr. et al. Small cell lung cancer: can recent advances in biology and molecular biology be translated into improved outcomes? *J. Thorac. Oncol.* **11**, 453–474 (2016).
- George, J. et al. Comprehensive genomic profiles of small cell lung cancer. *Nature* **524**, 47–53 (2015).
- Alexandrov, L. B. et al. Mutational signatures associated with tobacco smoking in human cancer. *Science* **354**, 618–622 (2016).
- Zhang, W. et al. Small cell lung cancer tumors and preclinical models display heterogeneity of neuroendocrine phenotypes. *Transl. Lung Cancer Res.* **7**, 32–49 (2018).
- Peifer, M. et al. Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. *Nat. Genet.* **44**, 1104–1110 (2012).
- Rudin, C. M. et al. Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer. *Nat. Genet.* **44**, 1111–1116 (2012).
- Kim, Y. H. et al. Combined microarray analysis of small cell lung cancer reveals altered apoptotic balance and distinct expression signatures of MYC family gene amplification. *Oncogene* **25**, 130–138 (2006).
- Wistuba, I. I., Gazdar, A. F. & Minna, J. D. Molecular genetics of small cell lung carcinoma. *Semin. Oncol.* **28**, 3–13 (2001).
- Gardner, E. E. et al. Chemosensitive relapse in small cell lung cancer proceeds through an EZH2-SLFN11 Axis. *Cancer Cell* **31**, 286–299 (2017).
- Poirier, J. T. et al. DNA methylation in small cell lung cancer defines distinct disease subtypes and correlates with high expression of EZH2. *Oncogene* **34**, 5869–5878 (2015).
- Park, K. S. et al. Characterization of the cell of origin for small cell lung cancer. *Cell Cycle* **10**, 2806–2815 (2011).
- Sutherland, K. D. et al. Cell of origin of small cell lung cancer: inactivation of Trp53 and Rb1 in distinct cell types of adult mouse lung. *Cancer Cell* **19**, 754–764 (2011).
- Borges, M. et al. An achaete-scute homologue essential for neuroendocrine differentiation in the lung. *Nature* **386**, 852–855 (1997).
- Johnson, J. E., Birren, S. J. & Anderson, D. J. Two rat homologues of *Drosophila* achaete-scute specifically expressed in neuronal precursors. *Nature* **346**, 858–861 (1990).
- Kato, F. et al. MYCL is a target of a BET bromodomain inhibitor, JQ1, on growth suppression efficacy in small cell lung cancer cells. *Oncotarget* **7**, 77378–77388 (2016).
- Borromeo, M. D. et al. ASCL1 and NEUROD1 reveal heterogeneity in pulmonary neuroendocrine tumors and regulate distinct genetic programs. *Cell Rep.* **16**, 1259–1272 (2016).
- Jahchan, N. S. et al. Identification and targeting of long-term tumor-propagating cells in small cell lung cancer. *Cell Rep.* **16**, 644–656 (2016).
- Schaffer, B. E. et al. Loss of p130 accelerates tumor development in a mouse model for human small-cell lung carcinoma. *Cancer Res.* **70**, 3877–3883 (2010).
- Carney, D. N. et al. Establishment and identification of small cell lung cancer cell lines having classic and variant features. *Cancer Res.* **45**, 2913–2923 (1985).
- Osborne, J. K. et al. NeuroD1 regulates survival and migration of neuroendocrine lung carcinomas via signaling molecules TrkB and NCAM. *Proc. Natl Acad. Sci. USA* **110**, 6524–6529 (2013).
- Mollaoglu, G. et al. MYC drives progression of small cell lung cancer to a variant neuroendocrine subtype with vulnerability to aurora kinase inhibition. *Cancer Cell* **31**, 270–285 (2017).
- Huang, Y. H. et al. POU2F3 is a master regulator of a tuft cell-like variant of small cell lung cancer. *Genes Dev.* **32**, 915–928 (2018).
- Rudin, C. M. et al. Molecular subtypes of small cell lung cancer: a synthesis of human and mouse model data. *Nat. Rev. Cancer* **19**, 289–297 (2019).
- Farago, A. F. & Keane, F. K. Current standards for clinical management of small cell lung cancer. *Transl. Lung Cancer Res.* **7**, 69–79 (2018).
- Antonia, S. J. et al. Nivolumab alone and nivolumab plus ipilimumab in recurrent small-cell lung cancer (CheckMate 032): a multicentre, open-label, phase 1/2 trial. *Lancet Oncol.* **17**, 883–895 (2016).
- Horn, L. et al. First-line atezolizumab plus chemotherapy in extensive-stage small-cell lung cancer. *N. Engl. J. Med.* **379**, 2220–2229 (2018).
- Sen, T. et al. Targeting DNA damage response promotes antitumor immunity through STING-mediated T-cell activation in small cell lung cancer. *Cancer Discov.* **9**, 646–661 (2019).
- Cardnell, R. J. et al. Protein expression of TTF1 and cMYC define distinct molecular subgroups of small cell lung cancer with unique vulnerabilities to aurora kinase inhibition, DLL3 targeting, and other targeted therapies. *Oncotarget* **8**, 73419–73432 (2017).
- Williamson, S. C. et al. Vasculogenic mimicry in small cell lung cancer. *Nat. Commun.* **7**, 13322 (2016).
- Gazdar, A. F., Carney, D. N., Nau, M. M. & Minna, J. D. Characterization of variant subclasses of cell lines derived from small cell lung cancer having distinctive biochemical, morphological, and growth properties. *Cancer Res.* **45**, 2924–2930 (1985).
- Meuwissen, R. et al. Induction of small cell lung cancer by somatic inactivation of both Trp53 and Rb1 in a conditional mouse model. *Cancer Cell* **4**, 181–189 (2003).
- Daniel, V. C. et al. A primary xenograft model of small-cell lung cancer reveals irreversible changes in gene expression imposed by culture in vitro. *Cancer Res.* **69**, 3364–3373 (2009).
- Drapkin, B. J. et al. Genomic and functional fidelity of small cell lung cancer patient-derived xenografts. *Cancer Discov.* **8**, 600–615 (2018).
- Faber, A. C. et al. Assessment of ABT-263 activity across a cancer cell line collection leads to a potent combination therapy for small-cell lung cancer. *Proc. Natl Acad. Sci. USA* **112**, E1288–E1296 (2015).
- Gardner, E. E. et al. Rapamycin rescues ABT-737 efficacy in small cell lung cancer. *Cancer Res.* **74**, 2846–2856 (2014).
- Allison Stewart, C. et al. Dynamic variations in epithelial-to-mesenchymal transition (EMT), ATM, and SLFN11 govern response to PARP inhibitors and cisplatin in small cell lung cancer. *Oncotarget* **8**, 28575–28587 (2017).
- Hodgkinson, C. L. et al. Tumorigenicity and genetic profiling of circulating tumor cells in small-cell lung cancer. *Nat. Med.* **20**, 897–903 (2014).
- Lallo, A. et al. The combination of the PARP inhibitor olaparib and the Wee1 inhibitor AZD1775 as a new therapeutic option for small cell lung cancer. *Clinical Cancer Res.* **24**, 5153–5164 (2018).
- Ben-David, U. et al. Patient-derived xenografts undergo mouse-specific tumor evolution. *Nat. Genet.* **49**, 1567–1575 (2017).
- Zhong, C., Fu, Y., Pan, W., Yu, J. & Wang, J. Atoh1 and other related key regulators in the development of auditory sensory epithelium in the mammalian inner ear: function and interplay. *Dev. Biol.* **446**, 133–141 (2019).
- Ben-Arie, N. et al. Math1 is essential for genesis of cerebellar granule neurons. *Nature* **390**, 169–172 (1997).
- Gao, Z. et al. The master negative regulator REST/NRSF controls adult neurogenesis by restraining the neurogenic program in quiescent stem cells. *J. Neurosci.* **31**, 9772–9786 (2011).
- Lim, J. S. et al. Intratumoural heterogeneity generated by notch signalling promotes small-cell lung cancer. *Nature* **545**, 360–364 (2017).
- McColl, K. et al. Reciprocal expression of INSM1 and YAP1 defines subgroups in small cell lung cancer. *Oncotarget* **8**, 73745–73756 (2017).
- Bragellmann, J. et al. Family matters: how MYC family oncogenes impact small cell lung cancer. *Cell Cycle* **16**, 1489–1498 (2017).
- Alves Rde, C., Meurer, R. T. & Roehle, A. V. MYC amplification is associated with poor survival in small cell lung cancer: a chromogenic in situ hybridization study. *J. Cancer Res. Clin. Oncol.* **140**, 2021–2025 (2014).
- Pozo, K., Minna, J. D. & Johnson, J. E. Identifying a missing lineage driver in a subset of lung neuroendocrine tumors. *Gene Dev.* **32**, 865–867 (2018).
- Calbo, J. et al. A functional role for tumor cell heterogeneity in a mouse model of small cell lung cancer. *Cancer Cell* **19**, 244–256 (2011).
- Geier, B., Kurmashev, D., Kurmasheva, R. T. & Houghton, P. J. Preclinical childhood sarcoma models: drug efficacy biomarker identification and validation. *Front. Oncol.* **5**, 193 (2015).
- Houghton, P. J. et al. The pediatric preclinical testing program: description of models and early testing results. *Pediatr. Blood Cancer* **49**, 928–940 (2007).
- Wagner, A. H. et al. Recurrent WNT pathway alterations are frequent in relapsed small cell lung cancer. *Nat. Commun.* **9**, 3787 (2018).
- Matthews, M. J., Kanhouou, S., Pickren, J. & Robinette, D. Frequency of residual and metastatic tumor in patients undergoing curative surgical resection for lung cancer. *Cancer Chemother. Rep.* **3**, 4, 63–67 (1973).
- Gazdar, A. F., Bunn, P. A. & Minna, J. D. Small-cell lung cancer: what we know, what we need to know and the path forward. *Nat. Rev. Cancer* **17**, 725–737 (2017).

54. Horie, M., Saito, A., Ohshima, M., Suzuki, H. I. & Nagase, T. YAP and TAZ modulate cell phenotype in a subset of small cell lung cancer. *Cancer Sci.* **107**, 1755–1766 (2016).
55. Barretina, J. et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
56. Shroyer, N. F. et al. Intestine-specific ablation of mouse atonal homolog 1 (Math1) reveals a role in cellular homeostasis. *Gastroenterology* **132**, 2478–2488 (2007).
57. Yang, Q., Bermingham, N. A., Finegold, M. J. & Zoghbi, H. Y. Requirement of Math1 for secretory cell lineage commitment in the mouse intestine. *Science* **294**, 2155–2158 (2001).
58. Kiernan, A. E., Cordes, R., Kopan, R., Gossler, A. & Gridley, T. The notch ligands DLL1 and JAG2 act synergistically to regulate hair cell development in the mammalian inner ear. *Development* **132**, 4353–4362 (2005).
59. Li, W. et al. Notch inhibition induces mitotically generated hair cells in mammalian cochlea via activating the Wnt pathway. *Proc. Natl Acad. Sci. USA* **112**, 166–171 (2015).
60. Abdolazimi, Y., Stojanova, Z. & Segil, N. Selection of cell fate in the organ of Corti involves the integration of Hes/Hey signaling at the Atoh1 promoter. *Development* **143**, 841–850 (2016).
61. Huang, C., Chan, J. A. & Schuurmans, C. Proneural bHLH genes in development and disease. *Curr. Top Dev. Biol.* **110**, 75–127 (2014).
62. Udyavar, A. R. et al. Novel hybrid phenotype revealed in small cell lung cancer by a transcription factor network model that can explain tumor heterogeneity. *Cancer Res.* **77**, 1063–1074 (2017).
63. Augert, A. et al. Targeting NOTCH activation in small cell lung cancer through LSD1 inhibition. *Sci. Signal* **12**, eaau2922 (2019).
64. Huijbers, I. J. et al. Rapid target gene validation in complex cancer mouse models using re-derived embryonic stem cells. *EMBO Mol. Med.* **6**, 212–225 (2014).
65. Semenova, E. A. et al. Transcription factor NFIB is a driver of small cell lung cancer progression in mice and marks metastatic disease in patients. *Cell Rep.* **16**, 631–643 (2016).
66. Malynn, B. A. et al. N-myc can functionally replace c-myc in murine development, cellular growth, and differentiation. *Gene Dev.* **14**, 1390–1399 (2000).
67. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
68. Zheng, H. C. The molecular mechanisms of chemoresistance in cancers. *Oncotarget* **8**, 59950–59964 (2017).
69. Gazdar, A. F., Hirsch, F. R. & Minna, J. D. From mice to men and back: an assessment of preclinical model systems for the study of lung cancers. *J. Thoracic Oncol.* **11**, 287–299 (2016).
70. McFadden, D. G. et al. Genetic and clonal dissection of murine small cell lung carcinoma progression by genome sequencing. *Cell* **156**, 1298–1311 (2014).
71. Khandelwal, G. et al. Next-generation sequencing analysis and algorithms for PDX and CDX models. *Mol. Cancer Res.* **15**, 1012–1016 (2017).
72. Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M. & Altman, D. G. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol.* **8**, e1000412 (2010).
73. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
74. Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* **35**, 2084–2092 (2019).
75. Murphy, B. et al. Evaluation of alternative in vivo drug screening methodology: a single mouse analysis. *Cancer Res.* **76**, 5798–5809 (2016).
76. Lee, S. et al. NGSCheckMate: software for validating sample identity in next-generation sequencing studies within and across data types. *Nucleic Acids Res.* **45**, e103 (2017).

Acknowledgements

We dedicate this manuscript to our collaborator A. Gazdar, a ground-breaking pathologist and pioneer in cell line and SCLC research. We thank J. Sage for his constructive review of the manuscript and the patients who donated their samples. This work was supported by The Christie Charitable Fund, and by Core Funding to CRUK Manchester Institute (grant no. A27412), Manchester CRUK Centre Award (grant no. A25254), the CRUK Manchester Experimental Cancer Medicines Centre (grant no. A25146) and the CRUK Lung Cancer Centre of Excellence (grant no. A20465). Patient recruitment was supported by the NIHR Manchester Biomedical Research Centre and NIHR Manchester Clinical research Facility at The Christie Hospital. Sample collection was undertaken via the ChemoRes Trial (molecular mechanisms underlying chemotherapy resistance, therapeutic escape, efficacy and toxicity—improving knowledge of treatment resistance in patients with lung cancer) and The TARGET Study. A.F.G. was supported by a grant from the National Cancer Institute, Bethesda, MD, USA: ‘Small Cell Lung Cancer Consortium Coordinating Center’ no. U24CA21327.

Author contributions

K.L.S. and C.D. supervised and devised the study, interpreted data and cowrote the manuscript. N.S., L.B., F.T., D.M., M.R. and T.H. carried out IHC and immunofluorescence including data analysis and interpretation. S.P.P., S.H., R.S., W.R. and A.K. carried out bioinformatics analyses and interpretation. M.D. carried out bioinformatics analysis. A.C. developed multiplex immunofluorescence protocols, analyzed data and generated resultant figures. K.K.F. has oversight of all CDX model generation and helped plan the study, analyze/interpret all data and edited the manuscript. M.G. is responsible for all in vivo work described. A.F.G. and L.G. generated the NE score and A.F.G. carried out an extensive pathology review of the models described as well as data interpretation and manuscript revision. D.N. carried out a pathology review of CDX models. L.C., M.G.K. and N.C. oversaw the acquisition of ethical permission and patient consent for patients on the TARGET study. M.C., L.F. and F.B. oversaw the acquisition of ethical permission and patient consent and the collection of blood samples for patients on the CHEMORES study. F.B. assisted with study design, data interpretation and manuscript revision and is the chief investigator of the CHEMORES study. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

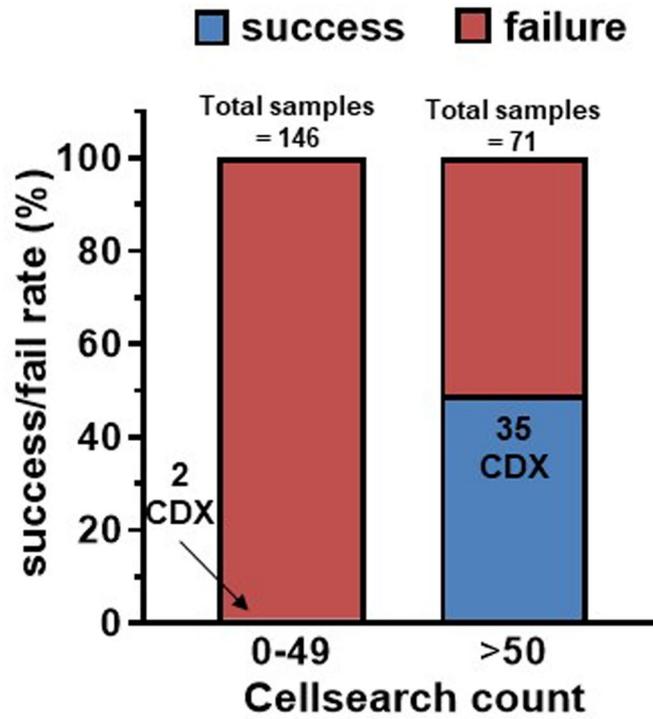
Supplementary information is available for this paper at <https://doi.org/10.1038/s43018-020-0046-2>.

Correspondence and requests for materials should be addressed to C.D.

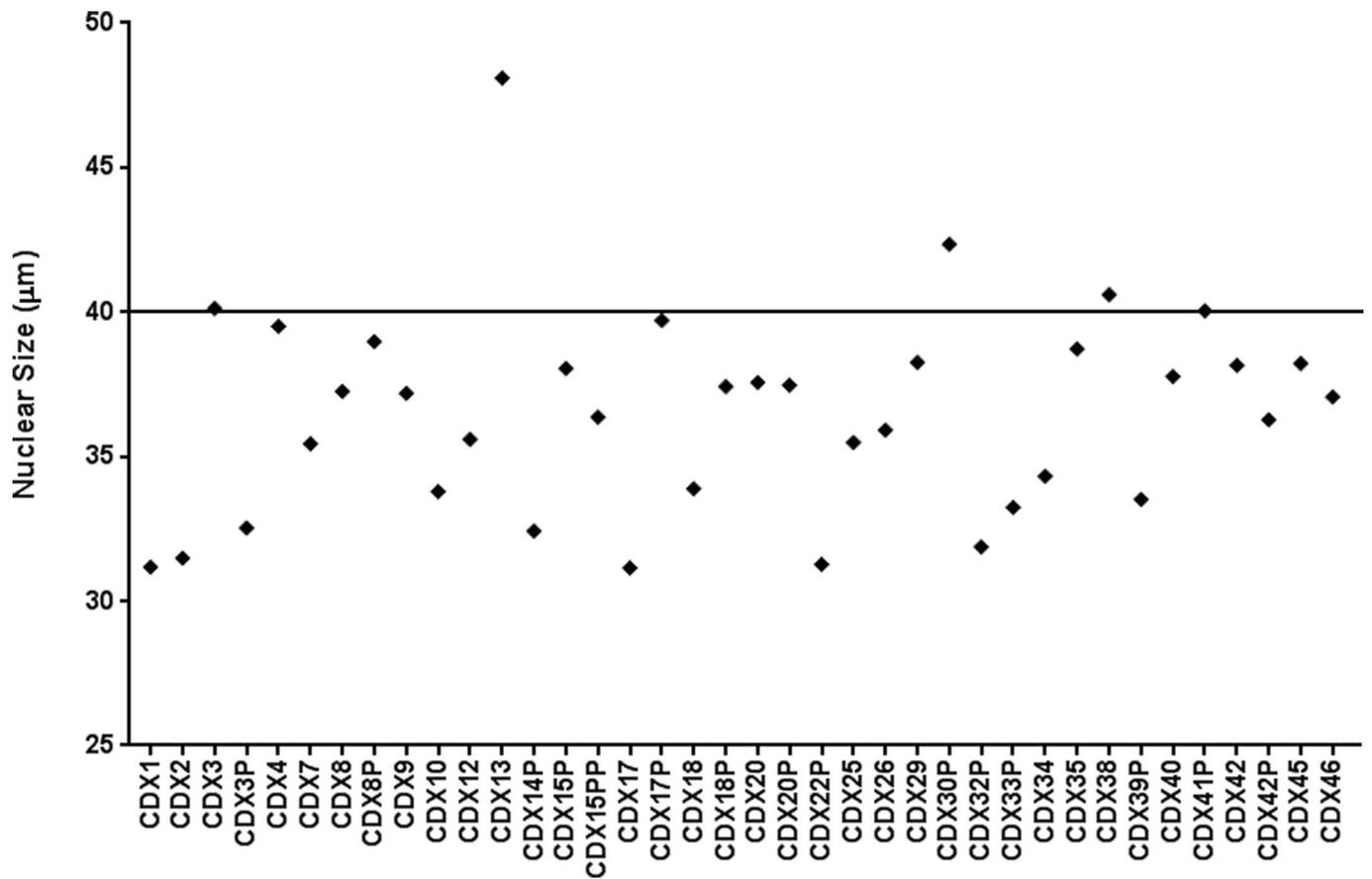
Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

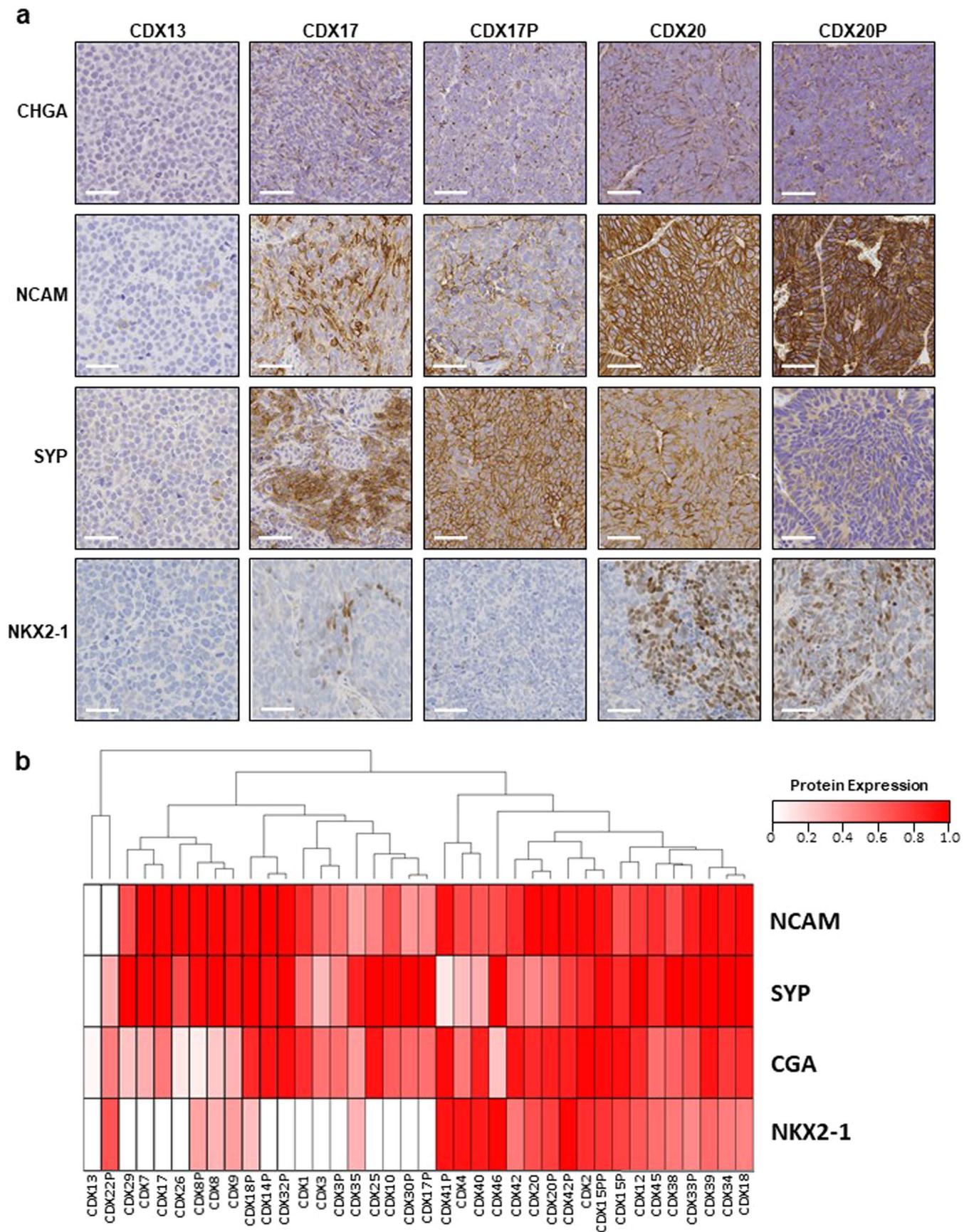
© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020



Extended Data Fig. 1 | Engraftment Rate For CDX Generation. For each CDX attempt, the CTC number was quantified by CellSearch in a parallel 7.5ml blood sample. Successful CDX generation was based on whether a measurable tumor grew within one year of implantation. Total samples implanted with matched CTC count <49 = 146, resulting in 2 successful CDX models, total samples implanted with matched CTC count >50 = 71, resulting in 35 successful CDX models.

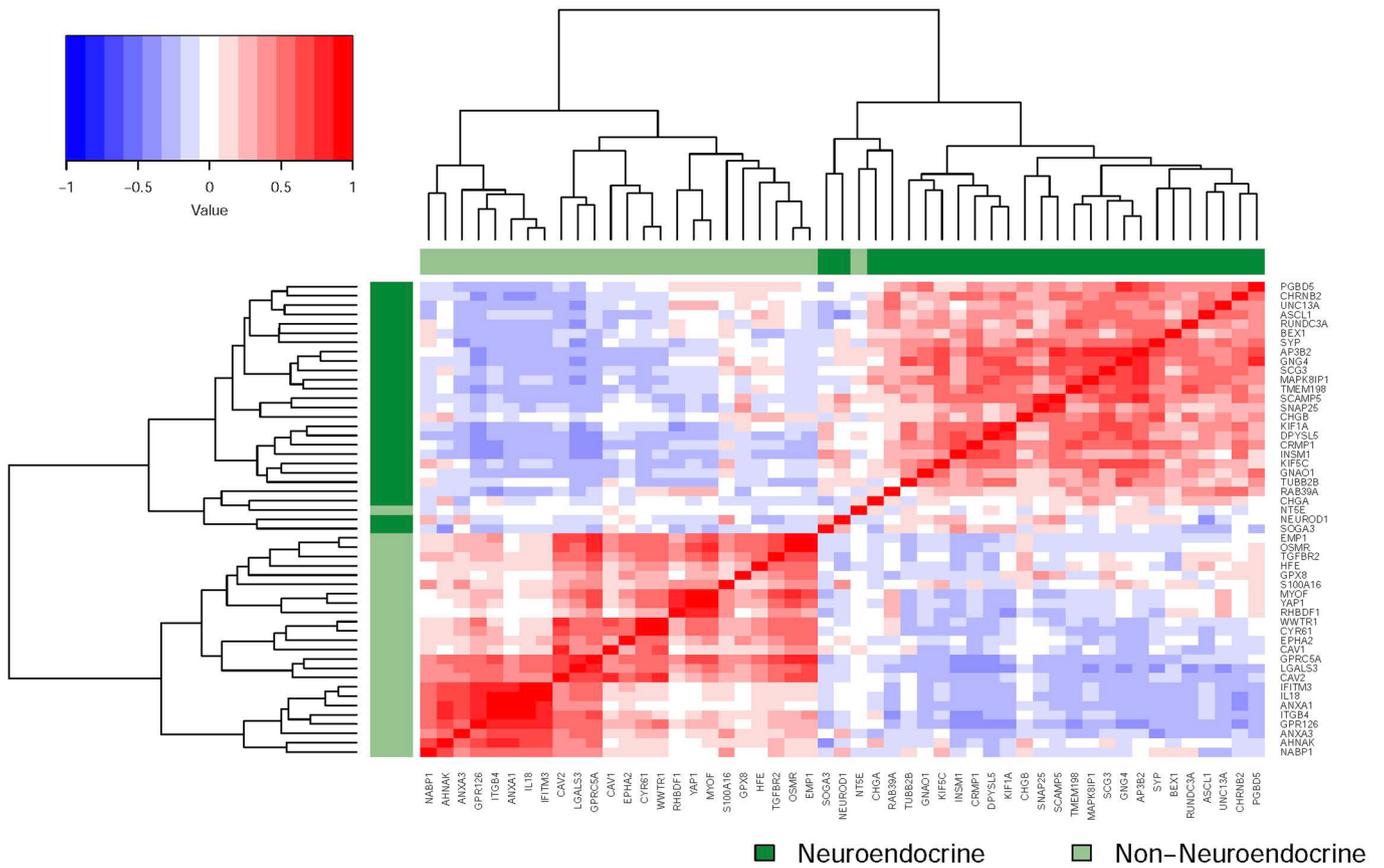


Extended Data Fig. 2 | CDX Cell Nuclear Size. Average Nuclear area (μm^2) was calculated for the CDX panel and used as a surrogate for total cell size and shows that the majority of CDX contain SCLC cells of a comparable size to human SCLC (approximately $40 \mu\text{m}^2$, ~size of 3–4 resting lymphocytes). Some CDX with relatively large cell nuclei were present (CDX3, CDX13, CDX17P, CDX30P and CDX38 and CDX41P $\geq 40 \mu\text{m}$, horizontal line). Mean value was calculated from is shown with error bars representing \pm SEM.

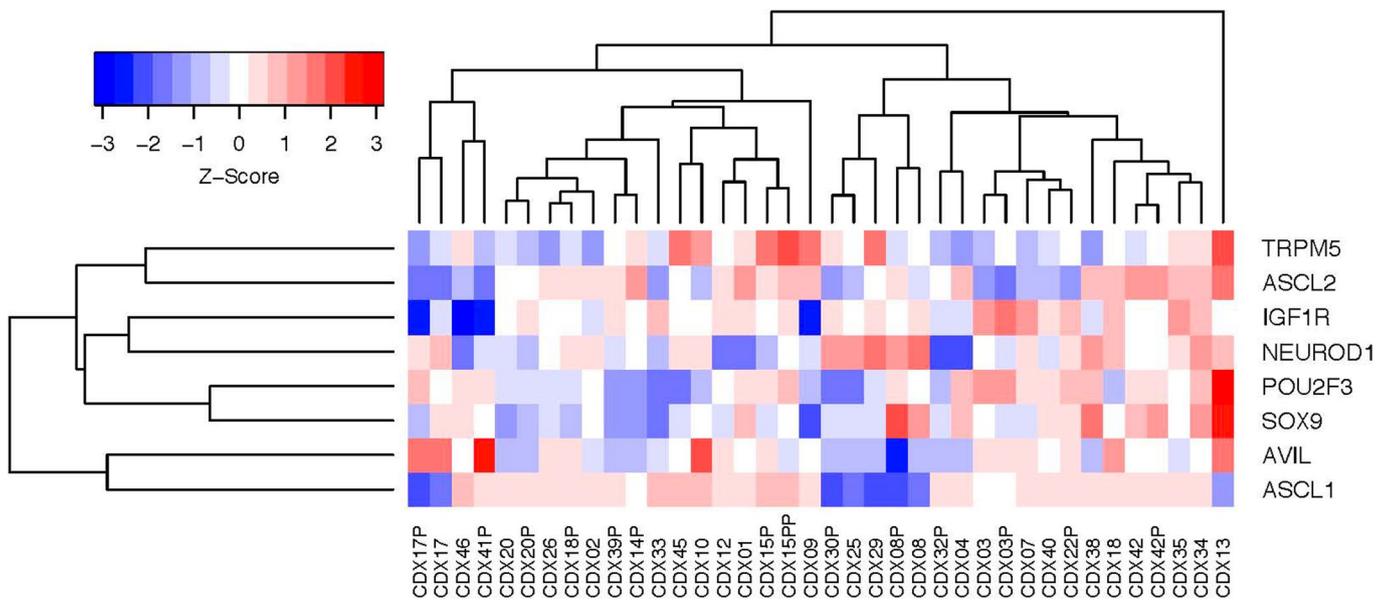


Extended Data Fig. 3 | see figure caption on next page.

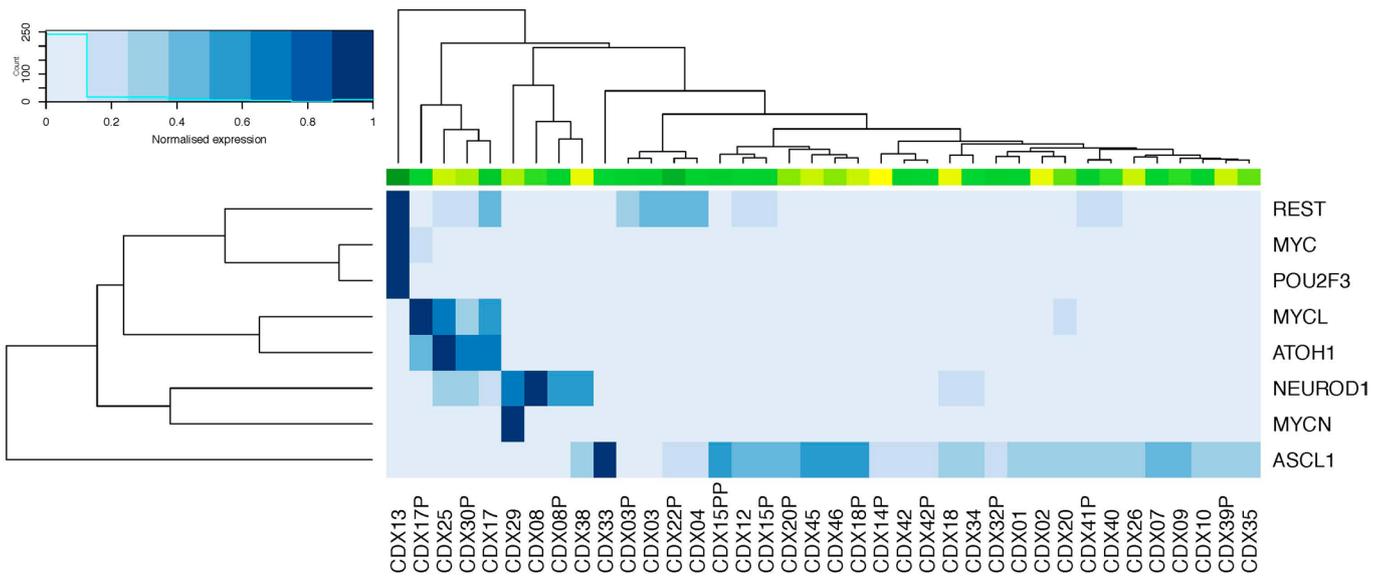
Extended Data Fig. 3 | Expression of SCLC Diagnostic Biomarkers. **a** Representative Immunohistochemistry (IHC, brown stain) for selected CDX models for NKX2-1 and the neuroendocrine markers CHGA, NCAM, and SYP. White scale bar, 50 μm . **b** Quantification of IHC data using Definiens software followed by hierarchical clustering (white, low, pink, intermediate, red, high expression, see methods). For a and b 3 whole tumour sections (biological replicates, different animals) were scanned and scored and the average value was used to generate the heatmap. Quantification was carried out according to the methods.



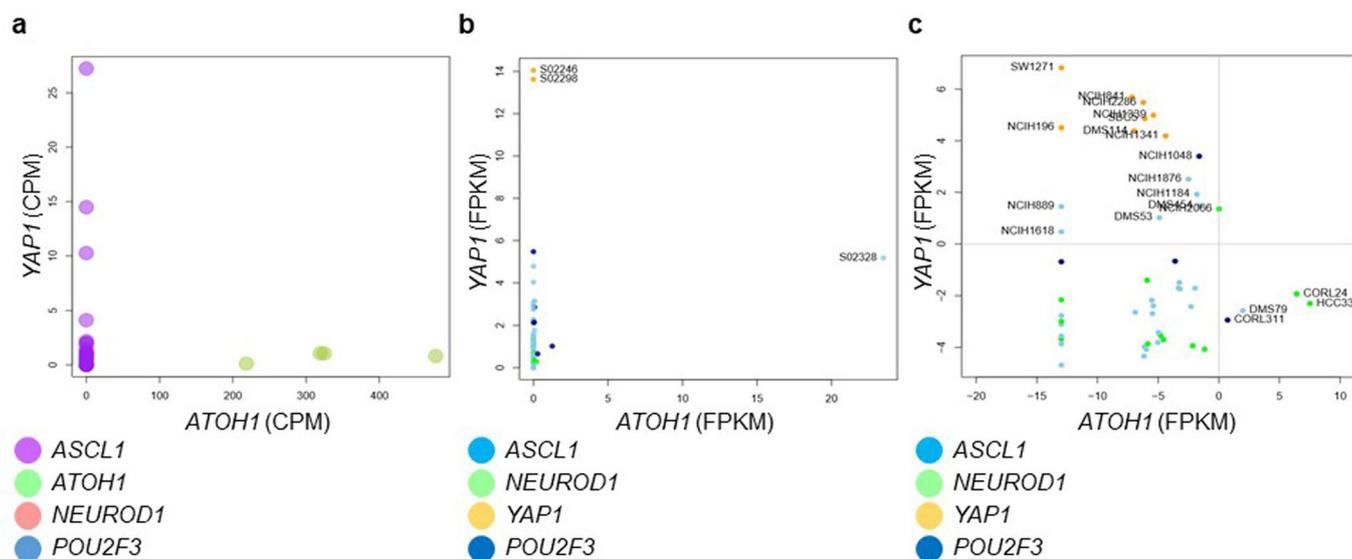
Extended Data Fig. 4 | Correlation of NE and non-NE genes in CDX. A previously derived 50 gene panel comprising NE and non-NE genes⁴ was mapped to CDX RNA sequencing data. Pearson correlation across the CDX dataset was calculated between all pairs of genes in the NE gene panel. Cells are coloured according to Pearson correlation between each pair of genes.



Extended Data Fig. 5 | Putative SCLC tuft cell markers in CDX Models. RNA-seq data of putative tuft cell marker expression in the CDX panel with cells coloured according to normalised z-scores. Data were pre-processed to remove reads of mouse origin, as described in ⁷⁰. Many of the key tuft markers are restricted to CDX13, which shows low expression of key transcription factors ASCL1 and NEUROD1.



Extended Data Fig. 6 | Comparison of MYC family expression with ASCL1, NEUROD1, POU2F3 and ATOH1 expressing CDX. RNA Expression analysis of ASCL1, NEUROD1, ATOH1, POU2F3 and REST with MYC family members. The bar at the top of the heatmap shows NE score. CDX13 has the only negative NE score and is unique in expressing POU2F3 and REST.



Extended Data Fig. 7 | ATOH1 and YAP1 Expression is Distinct in CDX, Patient Samples and Cell Lines. **a** Scatterplot RNA Expression (CPM) of ATOH1 and YAP1 in CDX. Samples are colour coded according to sub-groups defined by previous transcriptomics analysis: Purple, ASCL1, Blue, POU2F3, Pink, NEUROD1, Green, ATOH1; note that POU2F3 and NEUROD1 subgroups are obscured as the majority of samples represent the double negative population. **b** Scatterplot RNA Expression (RPKM) of ATOH1 and YAP1 in Surgically Resected SCLC Patient Samples. **c** Scatterplot RNA Expression (RPKM) of ATOH1 and YAP1 in SCLC Cell lines in the CCL. In both (b) and (c), samples are colour coded according to 4 sub-groups defined by Rudin et al²³; ASCL1, light blue, NEUROD1, green, POU2F3, dark blue, YAP1, orange.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data were collected as described in Khandelwal, G., et al., Next-Generation Sequencing Analysis and Algorithms for PDX and CDX Models. Mol Cancer Res, 2017. 15(8): p. 1012-1016 and Hodgkinson, C.L., et al., Tumorigenicity and genetic profiling of circulating tumor cells in small-cell lung cancer. Nat Med, 2014. 20(8): p. 897-903.

Data analysis

Data were filtered to remove reads aligning preferentially to the mouse genome using the 'bamcmp' algorithm, as described in Khandelwal et al., (2017) Mol Cancer Res: 15:1012. Reads were aligned to the GRCh38 assembly from Ensembl and Counts per million (CPM) values were determined. Heatmaps were generated using mean CPM values from each set of replicates, transformed into normalised counts. Data were plotted using 'heatmap.2' from the R package 'gplot' with default parameters, unless otherwise stated (Wasserman et al (1994) Cambridge: Cambridge University Press). Expression analysis was performed using a previously reported gene panel for NE classification (zhang et al., (2018) Transl. Lung cancer Res: 7:32). Different levels of variation are observed between CDX model replicates and between different CDX models, to account for this underlying structure within the data the 'duplicate correlation' function within limma was applied (Ritchie et al., Nucleic Acids Research, Volume 43, Issue 7, 20 April 2015, Page e47, <https://doi.org/10.1093/nar/gkv007>). Differential expression analysis was performed using Voom to transform the next generation sequencing data for subsequent processing with limma. Comparisons between baseline and progression models were performed using DESeq255. These comparisons were initially made individually between matching baseline and progression models, then in a combined analysis between model pairs. In all protocols, analysis was restricted to protein coding genes. Gene set enrichment analysis was performed using Generally Applicable Gene-set Enrichment (GAGE)(Luo W1 et al., BMC Bioinformatics. 2009; 10: 161). Enrichments in KEGG, GO biological process, GO cellular component and GO molecular function sets were analysed. When assessing gene set enrichment in each of the CDX clusters analysis was performed using the mean expression of the model

replicates. Comparisons between baseline and progression models and CDX clusters were performed using the “as.group” option in GAGE. Results were reduced to non-redundant gene sets using the “eset.grp” function.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw sequencing data for RNASeq can be accessed via array express with accession number E-MTAB-8465

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

The CDX Biobank is a collection of CDX models which are generated from CTCs from SCLC patients enrolled in the CHEMORES and TARGET Trials at the Christie Hospital, UK. CDX models reported here are from successful engraftment of CDX from SCLC patients enrolled on study between 2012 and 2017. No sample size calculations were carried out as this is a hypothesis-generating study which seeks to evaluate the presence of diagnostic and novel SCLC biomarkers for future validation. Samples are continually collected and mice are monitored for tumour growth. p1 models were compared to donor patient biopsy histopathology analyses and by two trained pathologists, who verified that the resultant CDX matched the donor patient's disease using the clinically recognised diagnostic criteria. Subsequent passage then resulted in p3 tumours which were also compared back to the donor/p1 tumours and verified to be consistent with the initial features of that patient's disease. Transcriptomics analysis and exploratory analyses (IHC/IF etc) were carried out on 3 independent CDX tumours per model and found to be concordant between replicates (see source data). For in vivo efficacy studies sample sizes per cohort varied from 3 per group to 14 per group depending on number of animals with tumours available at the time of randomisation. Sample size justification for in vivo studies was according to Murphy et al., Can Res 76(19) 5798-5809 (2016).

Data exclusions

No data were excluded from the study

Replication

Biological replicates of 2-9 independent tumours from each CDX model were generated and tested (stated in relevant methods). Histopathology was carried out on whole tumor sections and western blotting was carried out on lysates from 3 independent biological replicates per model. RNAseq analyses were conducted on 3-6 independent biological replicates per model as described in relevant methods and legends. Replicate data was not excluded from the study as this was primarily hypothesis generating, i.e in order to determine the feasibility of generating a SCLC CDX biobank as a surrogate for patient biopsy.

Randomization

There was no randomisation in this study. The study is an attempt to phenotypically characterise a number (n=37) of novel SCLC models for hypothesis generating and future validation of pre-clinical/clinical compounds/combinations and optimisation of biomarkers.

Blinding

During analysis all samples were given a unique identifier, but these were not blinded. This was not relevant as a) the study was for hypothesis generation and b) the clinical meta-data associated with each sample were not available/complete until the end of study. Furthermore all antibodies were fully validated prior to use in this study.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involvement
<input checked="" type="checkbox"/>	Involved in the study
<input checked="" type="checkbox"/>	Unique biological materials
<input checked="" type="checkbox"/>	Antibodies
<input checked="" type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	Palaeontology
<input checked="" type="checkbox"/>	Animals and other organisms
<input checked="" type="checkbox"/>	Human research participants

Methods

n/a	Involvement
<input checked="" type="checkbox"/>	Involved in the study
<input checked="" type="checkbox"/>	ChIP-seq
<input checked="" type="checkbox"/>	Flow cytometry
<input checked="" type="checkbox"/>	MRI-based neuroimaging

Unique biological materials

Policy information about [availability of materials](#)

Obtaining unique materials

The unique CDX models described in this manuscript are property of the Dive Laboratory. Prof. Dive is open to discuss collaborations which may involve sharing of CDX resources with the SCLC community (see contact information on manuscript).

Antibodies

Antibodies used

Antibody information is as follows (name, manufacturer, catalogue no., clone name, dilution: CD56: Leica Biosystems, PA0191, CD564, ready to use. Chromogranin A: ventana, 760 2519, LK2H10, ready to use. Cytokeratin: Dako, M3515, Y69, 1/100. c-Myc: ventana, 7904628, Y69, ready to use and Abcam, ab39688, Y69, 1:75. EpCAM: Cell Signalling, 2929, VU1D9, 1:100. MASH1: BD Pharmingen, 556604, 24B72D11.1, 1:250. mitochondria: Abcam, ab213725, 113-1, 1:500. NeuroD1: Abcam, ab213725, EPR20766, 1:100. PhosphoHistoneH3: Millipore, 06-570, Ser10, 1:250. POU2F3: sigma-Aldrich, HPA019652, polyclonal, 1:250. Retinoblastoma: Leica Biosystems, RB-358-L, 13A10, 1:50. REST: ThermoFisher Scientific, MA5-24606, CLO381, 1:250. Synaptophysin: Leica Biosystems, pA0299, 27G12, ready to use. TTF1: DAKO, M3575, 8G7G3/1, 1:200. Vimentin: Ventana, 7902917, V9, ready to use.

Validation

In the case of each primary antibody (CD56, Chromogranin A, Cytokeratin, cMYC, EpCAM, MASH1, Mitochondria, NeuroD1, PhosphoHistone H3, POU2F3, Retinoblastoma, REST, Synaptophysin, TTF1 and Vimentin) validation was carried out using positive and negative control human cancer cell lines with known +/- expression of the analyte in question as defined by the CCLE and/or the manufacturer/datasheet. Validation was also carried out in CDX tissue and xenografts from SCLC cell lines, as well as isotype control. Validation of antibodies was carried out according to methods described in: Smith and Womack, A matrix approach to guide IHC-based tissue biomarker development in oncology drug discovery. J Path. 2014 Jan;232(2):190-8. MYC IHC was carried out by the Christie Hospital Pathology Laboratory and is an accredited diagnostic test.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals

8-16 week old female NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ (NSG) mice. All procedures were carried out in accordance with Home Office Regulations (UK), the UK Coordinating Committee on Cancer Research guidelines and by approved protocols (Home Office Project license 40-3306/70-8252 and Cancer Research UK Manchester Institute Animal Welfare and Ethical Review Advisory Board). In vivo studies were conducted in order to comply with The Animal Research: Reporting of In Vivo Experiments (ARRIVE) Guidelines. Mice numbers per group were n=1 (in the case of initial CDX implantation) to 3 per group for molecular profiling studies to at least 3 per group for therapeutic testing studies.

Wild animals

study did not involve wild animals

Field-collected samples

study did not involve field-collected samples

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics

The patient data relating to this study can be found in Supplementary Figure 1 and Supplementary Table 2 which includes age, sex, diagnosis, sites of metastases, 1st line treatment, number of cycles, treatment response and survival (months). Broadly speaking there were 32 patients recruited following informed consent under CHEMORES or TARGET ethics (European Union CHEMORES FP6 Contract number LSHC-CT- 2007-037665 (NHS Northwest 9 Research Ethical Committee) and The TARGET (Tumour characterisation to Guide Experimental Targeted Therapy) study, approved by the North-West (Preston) National Research Ethics Service in Feb 2015, reference 15/NW/0078) and were between 55-79 years old (mean and median 68 years) with Extensive Stage SCLC, with metastasis to the lymph nodes, lung, liver, brain and bone. patients received 1-4 cycles of Carboplatin and Etoposide as Standard of Care for SCLC in the UK and survived between 1.3 - 28.9 months (mean= 7.9, median = 6.3) months post-diagnosis.

Recruitment

Participants were recruited by physician referral as described in the methods as follows: Patient samples and data were obtained between February 2012 and August 2016 according to ethically approved protocols: European Union CHEMORES FP6 Contract

number LSHC-CT- 2007-037665 (NHS Northwest 9 Research Ethical Committee) and The TARGET (Tumour chARacterisation to Guide Experimental Targeted Therapy) study, approved by the North-West (Preston) National Research Ethics Service in Feb 2015, reference 15/NW/0078. Patients were diagnosed with SCLC and receiving standard of care treatment, with their survival times in keeping with published averages, thus although the study could be said to be biased towards generation of models from patients with extensive stage disease, this is typical of SCLC in general, where the majority of patients will present with extensive stage at diagnosis.