This is a repository copy of *Molecular heterogeneity of non-small cell lung carcinoma patient-derived xenografts closely reflect their primary tumors*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/113009/

Version: Accepted Version

Article:
Wang, D., Pham, N.-A., Tong, J. et al. (30 more authors) (2017) Molecular heterogeneity of non-small cell lung carcinoma patient-derived xenografts closely reflect their primary tumors. International Journal of Cancer, 140 (3). pp. 662-673. ISSN 0020-7136

https://doi.org/10.1002/ijc.30472

This is the peer reviewed version of the following article: Wang, D., Pham, N.-A., Tong, J., Sakashita, S., Allo, G., Kim, L., Yanagawa, N., Raghavan, V., Wei, Y., To, C., Trinh, Q. M., Starmans, M. H.W., Chan-Seng-Yue, M. A., Chadwick, D., Li, L., Zhu, C.-Q., Liu, N., Li, M., Lee, S., Ignatchenko, V., Strumpf, D., Taylor, P., Moghal, N., Liu, G., Boutros, P. C., Kislinger, T., Pintilie, M., Jurisica, I., Shepherd, F. A., McPherson, J. D., Muthuswamy, L., Moran, M. F. and Tsao, M.-S. (2017), Molecular heterogeneity of non-small cell lung carcinoma patient-derived xenografts closely reflect their primary tumors. Int. J. Cancer, 140: 662–673. doi:10.1002/ijc.30472, which has been published in final form at https://doi.org/10.1002/ijc.30472. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Molecular heterogeneity of non-small cell lung carcinoma patient-derived xenografts closely reflect their primary tumors

Journal:	*International Journal of Cancer*
Manuscript ID	IJC-16-0893
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	15-May-2016
Complete List of Authors:	Tsao, Ming-Sound; Princess Margaret Cancer Centre Wang, Dennis; Princess Margaret Cancer Centre Pham, Nhu-An; Princess Margaret Cancer Centre Tong, Jiefei; Hospital for Sick Children, Molecular Structure and Function Sakashita, Shingo; Princess Margaret Cancer Centre Allo, Ghassan; Princess Margaret Cancer Centre Kim, Lucia; Princess Margaret Cancer Centre Yanagawa, Naoki ; Princess Margaret Cancer Centre Raghavan, Vibha; Princess Margaret Cancer Centre Wei, Yuhong ; Hospital for Sick Children, Molecular Structure and Function To, Christine ; Princess Margaret Cancer Centre Trinh, Quang ; Ontario Institute for Cancer Research Starmans, Maud; Ontario Institute for Cancer Research, Informatics & Biocomputing Chan-Seng-Yue, Michelle; Ontario Institute for Cancer Research, Informatics & Biocomputing Chadwick, Dianne; University Health Network, UHN Biospecimen Sciences Program Li, Lei ; Hospital for Sick Children, Molecular Structure and Function Zhu, Chang-Qi ; Princess Margaret Cancer Centre Liu, Ni ; Princess Margaret Cancer Centre Li, Ming ; Princess Margaret Cancer Centre Lee, Sharon ; Princess Margaret Cancer Centre Ignatchenko, Vladimir ; Hospital for Sick Children, Molecular Structure and Function Strumpf, Dan ; Princess Margaret Cancer Centre Taylor, Paul ; Hospital for Sick Children, Molecular Structure and Function Moghal, Nadeem ; Princess Margaret Cancer Centre Liu, Geoffrey ; Princess Margaret Cancer Centre Boutros, Paul; Ontario Institute for Cancer Research, Informatics & Biocomputing Kislinger, Thomas ; Princess Margaret Cancer Centre Pintilie, Melania ; Princess Margaret Cancer Centre Jurisica, Igor ; Princess Margaret Cancer Centre Shepherd, Frances; Princess Margaret Cancer Centre McPherson, John; Ontario Institute for Cancer Research

	Muthuswamy, Lakshmi ; Ontario Institute for Cancer Research Moran, Michael; Hospital for Sick Children, Molecular Structure and Function
Key Words:	xenograft, whole exome next generation sequencing, copy number aberration, transcriptome, phosphotyrosine-proteomics, mass spectrometry, DNA methylation, non-small cell lung cancer

SCHOLARONE™
Manuscripts

For Peer Review

Molecular heterogeneity of non-small cell lung carcinoma patient-derived xenografts closely reflect their primary tumors

RUNNING TITLE: Molecular landscape of NSCLC patient-derived xenografts

Dennis Wang,[1] Nhu-An Pham,[1] Jiefei Tong,[2] Shingo Sakashita,[1,3] Ghassan Allo,[1,3] Lucia Kim,[1,3] Naoki Yanagawa,[1,3] Vibha Raghavan,[1] Yuhong Wei,[2] Christine To,[1] Quang M. Trinh,[4] Maud H.W. Starmans,[4]-Michelle A. Chan-Seng-Yue,[4] Dianne Chadwick,[1] Lei Li,[2] Chang-Qi Zhu,[1] Ni Liu,[1] Ming Li,[1] Sharon Lee,[1] Vladimir Ignatchenko,[1] Dan Strumpf,[1] Paul Taylor,[2] Nadeem Moghal,[1,5] Geoffrey Liu,[1,5,6] Paul C. Boutros,[3,5,7] Thomas Kislinger,[1,5] Melania Pintilie,[1] Igor Jurisica,[1,5,8] Frances A. Shepherd,[1,6] John D. McPherson,[3,5] Lakshmi Muthuswamy,[1,3,5] Michael F. Moran,[2,9] and Ming-Sound Tsao[1,3,5]

[1]Princess Margaret Cancer Centre, University Health Network, Toronto; [2]Program in Molecular Structure and Function, Hospital for Sick Children, Toronto; [3]Department of Laboratory Medicine and Pathobiology, University of Toronto; [4]Ontario Institute of Cancer Research, Departments of [5]Medical Biophysics, [6]Medicine, [7]Pharmacology and Toxicology, [8]Computer Science, [9]Molecular Genetics, University of Toronto, Toronto, Ontario, Canada

CORRESPONDING AUTHOR

Dr. Ming-Sound Tsao, Room 14-401, Toronto Medical Discovery Tower, 101 College Street, M5G 1L7, Toronto, Ontario, Canada. Phone: 1-416-340-4737; Fax: 1-416-340-5517; email: Ming.Tsao@uhn.on.ca

For Peer Review

Disclosure of Potential Conflicts of Interest

The Authors have no conflicts to disclose.

Original Publication

This original manuscript is not under consideration by any other journals.

Keywords: xenograft, whole exome next generation sequencing, copy number aberration, transcriptome, phosphotyrosine-proteomics, mass spectrometry, DNA methylation, non-small cell lung cancer

Novelty and Impact

Non-small cell lung cancer is characterized by heterogeneous histology and genetics profile, driving the need for translational models that are clinically relevant. Many models would be required to represent the potential diverse mechanisms of lung cancer pathogenesis.
Our PDX resource currently has the largest number of lung cancer models representing the major histological subtypes. Comprehensive genomic proteomic profiles showed fidelity to the molecular pathology of their original human tumors in a subset of these models.

For Peer Review

ABSTRACT

Availability of lung cancer models that closely mimic human tumors remains a significant gap in cancer research, as tumor cell lines and mouse models may not recapitulate the spectrum of lung cancer heterogeneity seen in patients. We aimed to establish a patient-derived tumor xenograft (PDX) resource from surgically resected non-small cell lung cancer (NSCLC). Fresh tumor tissue from surgical resection was implanted and grown in the subcutaneous pocket of non-obese severe combined immune deficient (NOD SCID) gamma mice. Subsequent passages were in NOD SCID mice. A subset of matched patient and PDX tumors and non-neoplastic lung tissues were profiled by whole exome sequencing, single nucleotide polymorphism (SNP) and methylation arrays, and phosphotyrosine (pY)-proteome by mass spectrometry. The data were compared to published NSCLC datasets of NSCLC primary and cell lines. 127 stable PDXs were established from 441 lung carcinomas representing all major histological subtypes: 52 adenocarcinomas, 62 squamous cell carcinomas, 1 adeno-squamous carcinoma, 5 sarcomatoid carcinomas, 5 large cell neuroendocrine carcinomas, and 2 small cell lung cancers. Somatic mutations, gene copy number and expression profiles, and pY-proteome landscape of 36 PDXs showed greater similarity with patient tumors than with established cell lines. Novel somatic mutations on cancer associated genes were identified but only in PDXs, likely due to selective clonal growth in the PDXs that allows detection of these low allelic frequency mutations. The results provide the strongest evidence yet that PDXs established from lung cancers closely mimic the characteristics of patient primary tumors.

Introduction

For Peer Review

Lung cancers are grouped broadly into small cell lung cancer (SCLC, 15%) and non-small cell lung cancer (NSCLC, 85%). Except for very early stage tumors, they collectively lack curative treatments and account for the greatest number of cancer-related deaths worldwide [1]. Recent comprehensive genomic profiling of primary patient tumors belonging to the two most common NSCLC histological classes [2, 3], adenocarcinomas (AdC) and squamous cell carcinoma (SqCC) has provided new insights into the molecular complexity of these malignancies. Patient tumors are very heterogeneous histologically and genomically, with common cancer driver mutations mostly being restricted to <10% of patients, except for *TP53, KRAS* and *EGFR* [4]. This finding suggests that a large number of independent tumor models will be necessary to model the potential diverse mechanisms of lung cancer pathogenesis accurately.

Over decades, multiple tumor models have become available for the study of NSCLC, including cell lines established from patient tumors [5-7], genetically engineered mouse models (GEMMs) [8, 9], and patient-derived tumor xenografts (PDXs) [10, 11] . However, it remains unclear how well these preclinical tumor models recapitulate patient tumor biology. Some reports suggest that when the molecular profiles of patient tumors are compared to established cell lines, there is substantial genetic divergence between primary lung cancers and cell lines [12-14]. This observation raises the concern that cellular and biochemical mechanisms unique to culture-adapted cell lines growing on 2-dimensional conditions may not adequately reflect the full spectrum of key processes in patient tumors, and hence, may not accurately model patient responsiveness to cancer therapies [15-17]. This may contribute to and explain the high failure rates of early phase trials in cancer drug development [18], as most preclinical drug studies are conducted using established cell lines. Spontaneously formed tumors in GEMMs driven by single or limited number of genes also may be limited in addressing therapeutic responses

For Peer Review

relevant to patient tumors [19, 20], as they lack the substantial complexity of genomic alterations associated with patient tumors. Thus, PDX models ultimately may be better preclinical tumor models for oncology drug studies, as they may better mimic the genetic heterogeneity and biological complexity of clinical cancers [21, 22].

We and others have reported the establishment of transplantable NSCLC tumors directly from patients to immune-deficient mice. These PDXs bear close histopathological resemblance to their corresponding primary cancers and they recapitulate patient responses to a number of therapeutic agents [10, 19, 23, 24]. However, it is still unclear to what extent they fully recapitulate the genetic, epigenetic, and signaling mechanisms that collectively drive clinical disease, especially for their matched patient tumors. At a broad level, PDXs harbor similar NSCLC profiles of gene copy number [25], mRNA[10] and protein,[26] and retain specific "driver" mutations of the matching primary tumors [10, 11, 25-27]. Here, we now report the high concordance of genome-wide exomic single nucleotide variants (SNVs) and gene copy number variants (CNVs) between 25 PDXs and their matched patient tumors, which comprehensively establishes the ability of PDXs to faithfully recapitulate the genetic mechanisms driving pathobiology of their specific corresponding patient tumors. Furthermore, from a total of 36 PDXs, we use SNV, CNV and gene expression, as well as previously unexamined DNA methylation and phosphotyrosine (pY) proteome profiles to compare similarities between PDXs and publically described NSCLC primary tumors and established cell lines.

Materials and Methods

Patient Tumors and PDX Establishment

For Peer Review

The protocol for collection of tumor samples from surgically resected early-stage NSCLC to establish PDXs was approved by The University Health Network Human Research Ethics and Animal Care Committees. Patient tumor samples were obtained by informed consent. Tumor specimens were divided into small pieces (2 mm), mixed with 10% Matrigel at 4°C (Becton Dickinson Biosciences, San Jose, CA), and implanted into the subcutaneous flank tissue of non-obese diabetic severe combined immune deficiency (NOD SCID) gamma mice. All mice were bred in our facility, housed under sterile conditions, and given autoclaved food and water ad libitum. Once tumors reached 1-1.5 cm in diameter, as measured with calipers, mice were euthanized and tumors were implanted/passaged serially into new NOD SCID mice for at least three times to establish model stability. Harvested tumors were divided equally for cryopreservation, quick freeze in liquid nitrogen and preparation of formalin-fixed paraffin-embedded tissue blocks. Comprehensive profiling on PDXs was conducted from early passages 2-4.

Mutation Detection with Exome and Targeted Sequencing

Genomic DNA samples were isolated from 36 PDX tumors, 25 available matched snap-frozen banked primary tumors and 29 matched normal lung tissues. The whole exome was captured using a SureSelect Human 50Mbp kit (Agilent Technologies, Santa Clara, CA) according to the manufacturer's instructions. For targeted re-sequencing, we prioritized regions with somatic SNVs found by exome sequencing of both PDX and patient tumors and captured those regions using a SureSelect Custom 2Mbp kit. Captured DNA was sequenced using the HiSeq platform (Illumnina, San Diego, CA) and paired-end sequence reads were generated for each sample. Xenome [28] was used to first identify and remove contaminating reads from the

For Peer Review

mouse stroma for PDX samples. Basic alignment and sequence quality control was done using Novoalign v3 and Picard v1.78. Mapped exomes were processed by the standard Genome Analysis Toolkit (GATK) for quality control, variant calling, and mutational significance analysis. Somatic mutation calling with matched normal samples was also performed using Strelka v1.0 [29]. Mutations called by both GATK and Strelka were included in the targeted capture. High confidence variants in PDXs included only mutations called by both exome and targeted sequence analysis, or called in only exome sequence analysis but also found in the Catalogue of Somatic mutations in Cancer (COSMIC) database v70.

Copy Number Alterations

Genomic copy number was measured using the HumanOmni 2.5 BeadChip single nucleotide polymorphism (SNP) array platform (Illumina). The total signal intensity (LogR) and B-allele frequency (BAF) values were reported at each genomic locus that was profiled by the SNP array. The relative copy number gains or losses of genomic regions were identified using the allele-specific copy number analysis of tumors (ASCAT) segmentation algorithm [30]. Each gene was mapped to a genomic region if more than 50% of the gene overlapped with the region. We reported the average log likelihood ratio (LRR) for each gene, which quantifies the copy number gain in the tumor relative to normal tissue. For instance, a LRR=0.5 was considered copy number gain, and LRR=-0.5 was considered copy number loss. The gene copy number data from the 36 PDXs have been deposited in the Gene Expression Omnibus with GSE68929

Gene Expression Microarray

For Peer Review

RNA from 36 PDX samples was labeled and amplified using the whole genome gene expression DASL HT Assay (Illumina) for partially degraded, FFPE derived RNA. The labelled samples were hybridized onto Human HT-12 V4 BeadChips and processed as per the manufacturer's protocol, and scanned on the iScan (Illumina). The data files were quantified in GenomeStudio Version 2011.1. All the samples passed Illumina sample dependent and independent QC Metrics. Raw intensity data was log2 transformed and quantile normalized by the lumi R package 2.11.1. Batch effects from microarray assays conducted on different dates were adjusted using ComBat (sva R package) with default settings. A list of differentially expressed genes between SqCC and AdC PDXs was identified by analysis of variance (ANOVA) at adjusted $P<0.01$. We used Gene Set Enrichment Analysis (GSEA) to test whether the 131 upregulated genes in AdC and 289 genes upregulated in SqCC PDXs were also differentially expressed in patient tumors from the JBR.10 study, GSE14814 [31].

The gene expression data from the 36 PDXs and corresponding patient tumors have been deposited in the Gene Expression Omnibus with GSE68929.

DNA Methylation Chip

The bisulfite-converted DNA samples of tumors and matched patient normal samples were hybridized to the Illumina Infinitium HumanMethylation450 BeadChip and processed as per the manufacturer's protocol. The values M and Beta were background corrected and quantile normalized using the lumi R package 2.11.1. The log ratio between M values between tumor and normal samples defined the somatic methylation score. The methylation score for each gene was an average over all probes +1500bp and -500bp of the transcription start site.

Phosphotyrosine (pY) Proteome

For Peer Review

Total protein was extracted in urea lysis buffer, reduced and alkylated, and subjected to overnight tryptic digestion as described previously [26]. Peptides were purified by passing through a C-18 column, eluted with 50% acetonitrile and vacuum dried. The pY peptides were enriched using a PTMScan® Phosphotyrosine Mouse mAb (P-Tyr-100) kit as instructed by the manufacturer's protocol (Cell Signaling Technology, Danvers, MA). The enriched pY peptides were separated on an Easy-Spray column using an instrument with a combination of liquid chromatography and mass spectrometry systems (LC-MS/MS, Thermo Scientific, Odense Denmark) [32]. Acquired raw files were analyzed and quantified using MaxQuant software (version 1.3.0.5) [33]. The Andromeda probabilistic search engine [34] was used to search peak lists against the Swiss-Prot database (2013 version), with default search parameters as previously described [35]. The proteomics data described in this paper have been deposited to the ProteomeXchange Consortium (http://www.proteomexchange.org) via the PRIDE partner repository with the dataset identifier PXD000809. A detailed method is described in Supplementary Method S1.

Public Datasets

The Cancer Genome Atlas (TCGA) has the most extensive catalogues of mutations, gene copy changes, DNA methylation and gene expression data for the lung cancer subtypes AdC [3, 36] and SqCC [2]; data for 401 patients were downloaded from cBioportal (July 2014 version). Copy number gains and losses were determined using a LRR cutoff of ±0.5. RNA expression data derived from RNAseq by Expectation-Maximization (RSEM) format. The transcriptome of NSCLC was downloaded from GSE14814 dataset [31] . The phospho-proteome for NSCLC of 150 patients and 41 cell lines were acquired from Rikova et al. [37]. Broad CCLE cell line mutation,

For Peer Review

copy number and expression data were downloaded from the October 2012 version of their data portal (http://www.broadinstitute.org/ccle/data/).

Mutational signatures and altered pathways

All somatic mutations and their flanking bases were compiled and compared to pre-defined signatures from COSMIC (http://cancer.sanger.ac.uk/cosmic/signatures). We determined the linear combination of pre-defined signatures that most accurately reconstructed the mutational profile of a single tumor sample. A multiple linear regression model was used with the caveat that any coefficient must be greater than 0, as negative contributions are not possible in a biological sense.

Gene sets corresponding to pathways found in KEGG and Reactome databases were extracted. Mutations (somatic, wildtype) and copy number alterations (gain, loss, diploid) for each pathway were used as input features for iClusterPlus [38]. The model that maximized Bayesian information criterion was selected for each pathway and the proportion of variation explained was reported. To identify the expression pattern associated with each pathway alteration profile, we used the Random Forests algorithm (R package random Forest) running 1000 trees on all gene expression profiles and identified the top 100 expression profiles according to their importance scores.

Results

Patient tumors transplanted to mice create stable PDX models

A total of 441 resected primary lung carcinomas were implanted into NOD SCID mice and 127 (29%) patient-derived tumor xenografts (PDX) generated stable growth, which could be

John Wiley & Sons, Inc.

For Peer Review

passaged serially (Fig. 1A). Among 264 NSCLC patients with at least 3-years follow-up, patients whose tumors formed stable PDXs (*versus* those who did not) showed significantly worse disease-free survival (DFS, HR=3.12, 95% CI =2.02-4.83, P<0.0001) and overall survival (OS, HR=4.08, 95% CI =2.16-7.73, P<0.0001), even after adjusting for sex, stage and differentiation (Fig. 1B-C). Close histological similarities were observed between PDXs and their corresponding patient tumors (Fig. 1D-K), as shown in four paired tumors representing adenocarcinoma (AdC, D-E), squamous cell carcinoma (SqCC, F-G), sarcomatoid carcinoma (SC, H-I) and large-cell neuroendocrine carcinoma (LCNEC, J-K). Thirty six (15 AdC, 18 SqCC, two LCNEC and 1 SC) PDX models were comprehensively profiled by five-omics methods: exome, gene copy number, DNA methylation, mRNA expression, and pY proteome, and the matched patient tumor and adjacent normal tissue for a subset of PDXs were also analyzed by some of the platforms (Table 1 and Supplementary Tables S1 and S2).

Genomic fidelity of NSCLC PDXs to their corresponding patient tumors

Given the large number of heterogeneous mutations in NSCLC across patients, PDXs can be a powerful resource to study diverse genetic mechanisms of pathogenesis, provided they faithfully retain the mutation spectrum of the specific patient tumors from which they are derived. To investigate this possibility, we conducted whole exome sequencing of 36 PDXs, and their corresponding 25 patient tumors and 29 normal adjacent tissues, with few patients having insufficient specimen or poor specimen quality for analyses. The exome sequence of PDX tumor cells was additionally enriched over stromal sequence by removing any sequence reads that aligned better to the mouse rather than human genome. Approximately 88-99% of SNVs found in the primary tumor were retained in the matched PDX tumor (109-3373 SNVs/model)

For Peer Review

(Supplementary Fig. S1). Of common SNVs, 70% (7,507) were detected at <50% variant allele frequency (VAF) in both paired PDX-patient tumors, and 16% (1,763) had >50% VAF in PDXs, but <50% VAF in primary tumors (Supplementary Fig. S2). In contrast, only 6% of common SNVs (664) had VAFs greater in primary tumors than in PDXs. This enrichment of VAF in PDXs was significant (Chi-squared test $P<2.2 \times 10^{-6}$). We performed targeted re-sequencing of 31 PDX tumors at loci where somatic SNVs were found by exome sequencing and verified 6,419 unique non-synonymous SNVs, for a 59.4% verification rate (Supplementary Fig. S3).

PDXs also recapitulated genome-wide CNVs that are characteristic of their matched patient tumors (Fig. 2A, B). For example, between PDXs and patient tumors, common copy number gains were observed on chromosome 7 and 5p of AdCs, while focal amplifications of *SOX2* and *FGFR1* were shared in several SqCCs. We also observed that the signal intensity from copy number profiling on PDXs was higher than from matched patient tumors. This led to a higher resolution view of CNVs and histology-specific copy number gains and losses, using the PDXs as compared to patient tumors. A greater signal for copy number estimates were obtained from PDXs as compared to their matched primary tumors, as represented in the models 148 AdC and 77 SqCC (Supplementary Fig. S4). In addition, we observed an increased signal for loss of heterozygosity from the B allele frequencies at loci with known copy alterations.

We also compared the spectrum of genetic alterations in our patient tumors and corresponding PDXs to that observed in NSCLC TCGA patient and Cancer Cell Line Encyclopedia (CCLE) data. Of these validated non-synonymous SNVs, only 27% (1,728) have been reported in the Catalogue of Somatic Mutations in Cancer (COSMIC). The amount of mutation burden in each PDX was comparable to that in primary lung tumors from The Cancer Genome Atlas (TCGA) (Supplementary Table S3), and was lower than reported in the 70

For Peer Review

NSCLC CCLE cell lines [39]. In addition to known mutations, we found several non-synonymous somatic mutations in oncogenes and tumor suppressor genes that have not been reported to date in the COSMIC [40] (Supplementary Table S4).

We found that many broad and focal copy amplifications and deletions were shared (Supplementary Fig. S5). Notably, PDX AdC copy number variants (CNVs) were more similar to patient tumors (r=0.69) than to AdC cell lines (r=0.35) based on the log-R ratio of CNVs. Overall, the number of genes somatically altered (mutation and CNV) in the PDXs was most similar to patient tumors and was lower than in cell lines (Supplementary Fig. S6). These data support our PDXs being appropriate models to study the same type of genetic diversity observed in a larger cohort of NSCLC patients.

The phosphotyrosine (pY)-proteome of PDXs is more similar to patient tumors than to cell lines

Alterations in phosphotyrosine (pY) signaling occur widely in cancer, including NSCLC, and contribute to a number of oncogenic mechanisms [37, 41]. Although PDXs recapitulate the genetic alterations associated with primary patient tumors, it has not yet been reported whether they also retain global P-Tyr patterns. Due to limitations on patient material, we focused on using a liquid chromatography-mass spectrometry (LC-MS) method to detect pY sites on 351 proteins in 31 PDXs, and then compared these pY profiles to published pY datasets of NSCLC cell lines, patient tumors, and corresponding normal lung tissues [37]. Hierarchical and principal components clustering across tyrosine kinases and other subsets of genes in the pY-proteome showed greater similarity between patient and PDX tumors than to cell lines (Fig. 3A).

For Peer Review

Furthermore, the PDXs and patient tumors tended to have more variable and lower levels of tyrosine phosphorylation than cell lines (Fig. 3 B, C), suggesting cell lines may have an overall greater dependency on high levels of phosphotyrosine signaling by specific kinases.

Transcription and methylation landscapes characterize histological subtypes of NSCLC

Having established the fidelity of PDXs to recapitulate the genetic characteristics and P-Tyr signaling patterns of patient tumors, we then performed hierarchical clustering of the 200 most variable DNA methylation sites at gene promoters across 36 PDXs differentiated the tumors according to their histological subtypes (Fig. 4A). Similar to methylation, hierarchical clustering based on a profile of the 200 most variably expressed genes separated the PDXs according to their paired patient tumor histological subtypes (Fig. 4B). The PDX DNA methylation or gene expression patterns that correlated with histological subtypes also predicted correctly the subtypes of independent NSCLC patient tumors. Based on the identification of hyper- and hypo- methylated regions along the genome by probing genes and cytosine-phosphate-guanine (CpG) islands, a principal component clustering analysis showed that the methylation profiles of PDXs were more variable than patient tumors from the TCGA dataset (Fig. 4C); however, PDX and patient tumors of the same histological subtype clustered closely. Using 131 genes that were differentially up-regulated in AdC PDXs and 289 genes up-regulated in SqCC PDXs (absolute fold change >2, $P<0.001$) and applying them to 133 stage I and II NSCLC patient samples (GSE GSE14814), these sets of genes also were upregulated and enriched in the same subtypes of patient tumors (Fig. 4D, $p<0.001$).

Alterations in key cancer drivers and oncogenic pathways

For Peer Review

Known oncogenic mutations identified in PDXs included *TP53* in 31 (86%), *KRAS* in 9 (25%), *EGFR* in 2 (6%) and *PIK3CA* in 2 (6%) models (Fig. 5A). The co-occurrence of multiple somatic mutations and copy number aberrations were detected among the PDXs models. Gene sets were identified and associated with specific pathways that contained these alterations. Integrative clustering using iClusterPlus then was applied to the PDX genomic profiles to estimate the proportion of variation explained by alterations in each pathway (Supplementary Fig. S7A). Genes regulating cell cycle, including *TP53*, were altered in all PDX models except one (X207), and integrative clusters based on those alterations explained ~50% of the variation in the PDX genomic landscape (Fig. 5B). The occurrence of alterations in specific gene sets, such as AKT signaling, MAPK signaling and metabolic pathways was correlated with activity in cell signaling processes manifested in both the phospho-proteomic (Supplementary Fig. S7B) and mRNA levels (Fig. 5C). Genomic alterations in the CTLA4 signaling pathway were frequent across PDXs and co-occurred with alterations in cell cycle and AKT signaling pathways. For a subset of these PDXs, we also observed enrichment of genes related to the positive regulation of the immune system.

For Peer Review

Discussion

In this study, we have confirmed with twice the number of patients and longer follow-up our previous finding that the ability of NSCLC to grow as a xenograft in mice is a strong poor prognostic indication for patient survival [11]. Furthermore, we have utilized five comprehensive omics platforms including exome, gene copy number, DNA methylation, mRNA expression and pY-proteome, to characterize 36 of the PDX models. The separation of histological subtypes

AdC and SqCC was observed consistently at all levels of the omics profiles. The comparison of the PDX derived data to matched patient tumors, other cohorts of lung cancer patient tumors [2, 3, 36] and established cell lines [37, 39] showed that across gene mutations, copy number alterations and phospho-proteins, there were greater similarities between PDXs to patient tumors than to cell lines. This fidelity and diversity illustrates an advantage of PDX models as preclinical model over cell lines or other *in vivo* tumor models, including xenografts formed from established cell lines and genetically engineered mouse models (GEMMs). While GEMMs have an advantage in studying the biology of a single or combination of two genetic abnormalities, their utility is limited to the available models that represent few specific genetic abnormalities and histological types of lung cancers [42, 43].

The high recapitulation of the genomic landscape across NSCLC histological subtypes when PDXs were compared to lung primary tumors [2, 3, 36] as shown in this study, is consistent with omics studies comparing PDXs of breast and colon cancers to their corresponding patient tumors [44, 45]. This contrasts to cell lines, which in previous comparisons to patient tumors have shown low correlation between their point mutation frequencies ($\rho \cong 0.5$), and marked differences in their mRNA and copy number profiles [15, 39]. Recently, Gao *et al* [22] reported similar results of closer similarities for mutation rates and copy number landscape between PDXs models established across common adult solid cancers, including 42 NSCLC models, and data from primary patient tumors (TCGA) and cell lines (CCLE).

By measuring the variant allele frequency, we have shown that a significant proportion of SNVs were detected at higher frequencies in PDXs compared to their corresponding primary tumors. This could be due to higher tumor cellularity in PDXs compared to primary patient tumors, which often have greater stromal and inflammatory cell infiltrates, which are lost during

For Peer Review

engraftment in immune-deficient mice [46]. Higher VAFs could also result from selective enrichment of clones harboring mutations that favor higher growth rates and/or more aggressive cancer biology [47]. Genomic alterations in cell cycle pathways were the most frequently found changes in PDXs.

While the ability of these PDXs to model histology and the omic profiles is an advantage for their utility as preclinical models, their murine host may hinder specific experimental conditions. Our observation that lower levels of c-MET phosphorylation occurring in PDXs compared to cell lines may be a consequence of the inability of the murine hepatocyte growth factor to activate human c-MET tyrosine function [48]. The NOD SCID strain also lacks an immune system which would prevent the testing of therapeutic strategies that require activation of immune response. Possible alternate murine hosts could include nude or SCID strains which have a greater representation of the immune components, or a humanized mouse host by reconstitution of human hematopoietic stem cells [49]. Nevertheless, we observed a subgroup of PDX tumors with genomic alterations in CTLA4 pathways modulating immune checkpoints that resulted in up-regulation of immune system genes [50].

Preclinical models, including PDXs derived directly from patient tumors are emerging as the preferred platform for translational lung cancer research [39, 44]. Some evidence, including studies from our lab, shows that NSCLC PDX models respond similarly to pharmacological agents when compared to their matched patient tumors [19, 24], indicating that high throughput screening strategy using large number of PDX models may be used pre-clinically to assess population responses to novel therapeutics [22]. Genomics-characterized PDXs models may also enable systematic investigation on the underlying biology and mechanisms of drug response or

For Peer Review

treatment resistance, identify potential predictive biomarkers for novel drugs, and contribute to the design of early phase clinical trials.

Acknowledgements

Supported by the Ontario Research Fund-Research Excellence Grant (RE-03-020), Ontario Premier's Summit Award (FAS), Canadian Cancer Society grants #020527 (MST) and #701595 (MST), Canadian Institutes for Health Research (CIHR) grant MOP-102536 (MM), Princess Margaret Hospital Cancer Foundation Investment in Research Program (BW), Ontario Institute of Cancer Research (OICR) to PCB and JDM, Ontario Ministry of Health and Long Term Care, and Canada Research Chairs Program (TK, MFM, IJ). Dr. M.-S. Tsao is the M. Oasim Choksi Chair in Lung Cancer Translational Research. Dr. F.A. Shepherd is the Scott Taylor Chair in Lung Cancer Research. Dr. G. Liu is the Alan B. Brown Chair of Molecular Genomics and Cancer Care Ontario Chair in Experiment Therapeutics and Population Studies. Dr. Y. Wei was supported by CIHR Fellowship. Drs. S. Sakashita, G. Allo, L. Kim and N. Yanagawa were partially supported by the Terry Fox Foundation STIHR grant in Molecular Pathology of Cancer at CIHR (STP-53912) and OICR.

Authors' Contributions

Conception and design: D. Wang, N. Moghal, G. Liu, T. Kislinger, I. Jurisica, F.A. Shepherd, J.D. McPherson, M.F. Moran, M.-S. Tsao

For Peer Review

Administrative, material support (i.e., provision of tumor tissue): N.-A. Pham, D. Chadwick, M. Li, G. Liu, F.A. Shepherd, M.-S. Tsao

Data collection: D. Wang, N.-A. Pham, J. Tong , S. Sakashita, G. Allo, L. Kim, N. Yanagawa, Y. Wei, C. To, Q.M. Trinh, M. H.W. Starmans, D. Chadwick, L. Li, C.-Q. Zhu, N. Liu, M. Li, S. Lee, V. Ignatchenko, D. Strumpf, P. Taylor, P.C. Boutros, M. Pintilie, J.D. McPherson, L. Muthuswamy, M.F. Moran, M.-S. Tsao

Data analysis and interpretation: D. Wang, N.-A. Pham, J. Tong , S. Sakashita, G. Allo, L. Kim, N. Yanagawa, V. Raghavan, Y. Wei, C. To, Q.M. Trinh, M. H.W. Starmans, M.A. Chan-Seng-Yue, L. Li, C.-Q. Zhu, N. Liu, M. Li, S. Lee, V. Ignatchenko, D. Strumpf, P. Taylor, N. Moghal, G. Liu, P.C. Boutros, T. Kislinger, M. Pintilie, I. Jurisica, J.D. McPherson, L. Muthuswamy, M.F. Moran, M.-S. Tsao

Writing, review of manuscript: D. Wang, N.-A. Pham, J. Tong, L. Li, C.-Q. Zhu, N. Moghal, G. Liu, P.C. Boutros, T. Kislinger, M. Pintilie, I. Jurisica, F.A. Shepherd, J.D. McPherson, M.F. Moran, M.-S. Tsao

For Peer Review

Acknowledgments

We would like to thank Jenna Sykes, Christine Ng and Jonathan Chang for their technical support and helpful discussions.

Supporting Information

Supplementary Table S1: Sample information on the profiled samples and the gene-level molecular profiles of the PDXs.

Supplementary Method, Figures and Tables: Supplementary Method S1, Table S2-4, and Figure S1-7.

References

1. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer. *http://globocaniarcfr, accessed on 13/12/2013* 2013.

2. Cancer Genome Atlas Research N. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 2012;**489**: 519-25.

3. Cancer Genome Atlas Research N. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 2014;**511**: 543-50.

4. Sanders HR, Albitar M. Somatic mutations of signaling genes in non-small-cell lung cancer. *Cancer genetics and cytogenetics* 2010;**203**: 7-15.

5. Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, Greninger P, Thompson IR, Luo X, Soares J, Liu Q, Iorio F, et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* 2012;**483**: 570-5.

6. Gazdar AF, Gao B, Minna JD. Lung cancer cell lines: Useless artifacts or invaluable tools for medical science? *Lung cancer* 2010;**68**: 309-18.

7. Sharma SV, Haber DA, Settleman J. Cell line-based platforms to evaluate the therapeutic efficacy of candidate anticancer agents. *Nature reviews Cancer* 2010;**10**: 241-53.

8. Chen Z, Cheng K, Walton Z, Wang Y, Ebi H, Shimamura T, Liu Y, Tupper T, Ouyang J, Li J, Gao P, Woo MS, et al. A murine lung cancer co-clinical trial identifies genetic modifiers of therapeutic response. *Nature* 2012;**483**: 613-7.

9. Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong KK. Non-small-cell lung cancers: a heterogeneous set of diseases. *Nature reviews Cancer* 2014;**14**: 535-46.

10. Fichtner I, Rolff J, Soong R, Hoffmann J, Hammer S, Sommer A, Becker M, Merk J. Establishment of patient-derived non-small cell lung cancer xenografts as models for the identification of predictive biomarkers. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2008;**14**: 6456-68.

11. John T, Kohler D, Pintilie M, Yanagawa N, Pham NA, Li M, Panchal D, Hui F, Meng F, Shepherd FA, Tsao MS. The ability to form primary tumor xenografts is predictive of increased risk of disease recurrence in early-stage non-small cell lung cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2011;**17**: 134-41.

12. Daniel VC, Marchionni L, Hierman JS, Rhodes JT, Devereux WL, Rudin CM, Yung R, Parmigiani G, Dorsch M, Peacock CD, Watkins DN. A primary xenograft model of small-cell lung cancer reveals irreversible changes in gene expression imposed by culture in vitro. *Cancer research* 2009;**69**: 3364-73.

13. Herrmann D, Conway JR, Vennin C, Magenau A, Hughes WE, Morton JP, Timpson P. Three-dimensional cancer models mimic cell-matrix interactions in the tumour microenvironment. *Carcinogenesis* 2014;**35**: 1671-9.

14. Mishra DK, Creighton CJ, Zhang Y, Gibbons DL, Kurie JM, Kim MP. Gene expression profile of A549 cells from tissue of 4D model predicts poor prognosis in lung cancer patients. *International journal of cancer Journal international du cancer* 2014;**134**: 789-98.

15. Domcke S, Sinha R, Levine DA, Sander C, Schultz N. Evaluating cell lines as tumour models by comparison of genomic profiles. *Nature communications* 2013;**4**: 2126.

For Peer Review

16. Gillet JP, Varma S, Gottesman MM. The clinical relevance of cancer cell lines. *Journal of the National Cancer Institute* 2013;**105**: 452-8.

17. Johnson JI, Decker S, Zaharevitz D, Rubinstein LV, Venditti JM, Schepartz S, Kalyandrug S, Christian M, Arbuck S, Hollingshead M, Sausville EA. Relationships between drug activity in NCI preclinical in vitro and in vivo models and early clinical trials. *British journal of cancer* 2001;**84**: 1424-31.

18. El-Maraghi RH, Eisenhauer EA. Review of phase II trial designs used in studies of molecular targeted agents: outcomes and predictors of success in phase III. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2008;**26**: 1346-54.

19. Gandara DR, Mack PC, Bult C, Li T, Lara PN, Jr., Riess JW, Astrow SH, Gandour-Edwards R, Cooke DT, Yoneda KY, Moore EH, Pan CX, et al. Bridging tumor genomics to patient outcomes through an integrated patient-derived xenograft platform. *Clinical lung cancer* 2015;**16**: 165-72.

20. Singh M, Murriel CL, Johnson L. Genetically engineered mouse models: closing the gap between preclinical data and trial outcomes. *Cancer research* 2012;**72**: 2695-700.

21. Malaney P, Nicosia SV, Dave V. One mouse, one patient paradigm: New avatars of personalized cancer therapy. *Cancer letters* 2014;**344**: 1-12.

22. Gao H, Korn JM, Ferretti S, Monahan JE, Wang Y, Singh M, Zhang C, Schnell C, Yang G, Zhang Y, Balbin OA, Barbe S, et al. High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat Med* 2015;**21**: 1318-25.

23. Mattern J, Jager S, Sonka J, Wayss K, Volm M. Growth of human bronchial carcinomas in nude mice. *British journal of cancer* 1985;**51**: 195-200.

24. Stewart EL, Mascaux C, Pham NA, Sakashita S, Sykes J, Kim L, Yanagawa N, Allo G, Ishizawa K, Wang D, Zhu CQ, Li M, et al. Clinical Utility of Patient-Derived Xenografts to Determine Biomarkers of Prognosis and Map Resistance Pathways in EGFR-Mutant Lung Adenocarcinoma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2015;**33**: 2472-80.

25. Ilie M, Nunes M, Blot L, Hofman V, Long-Mira E, Butori C, Selva E, Merino-Trigo A, Venissac N, Mouroux J, Vrignaud P, Hofman P. Setting up a wide panel of patient-derived tumor xenografts of non-small cell lung cancer by improving the preanalytical steps. *Cancer medicine* 2015;**4**: 201-11.

26. Li L, Wei Y, To C, Zhu CQ, Tong J, Pham NA, Taylor P, Ignatchenko V, Ignatchenko A, Zhang W, Wang D, Yanagawa N, et al. Integrated omic analysis of lung cancer reveals metabolism proteome signatures with prognostic impact. *Nature communications* 2014;**5**: 5469.

27. Hao C, Wang L, Peng S, Cao M, Li H, Hu J, Huang X, Liu W, Zhang H, Wu S, Pataer A, Heymach JV, et al. Gene mutations in primary tumors and corresponding patient-derived xenografts derived from non-small cell lung cancer. *Cancer letters* 2015;**357**: 179-85.

28. Conway T, Wazny J, Bromage A, Tymms M, Sooraj D, Williams ED, Beresford-Smith B. Xenome--a tool for classifying reads from xenograft samples. *Bioinformatics* 2012;**28**: i172-8.

29. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* 2012;**28**: 1811-7.

30. Van Loo P, Nordgard SH, Lingjaerde OC, Russnes HG, Rye IH, Sun W, Weigman VJ, Marynen P, Zetterberg A, Naume B, Perou CM, Borresen-Dale AL, et al. Allele-specific

For Peer Review

copy number analysis of tumors. *Proceedings of the National Academy of Sciences of the United States of America* 2010;**107**: 16910-5.

31. Zhu CQ, Ding K, Strumpf D, Weir BA, Meyerson M, Pennell N, Thomas RK, Naoki K, Ladd-Acosta C, Liu N, Pintilie M, Der S, et al. Prognostic and predictive gene signature for adjuvant chemotherapy in resected non-small-cell lung cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2010;**28**: 4417-24.

32. Tong J, Taylor P, Moran MF. Proteomic analysis of the epidermal growth factor receptor (EGFR) interactome and post-translational modifications associated with receptor endocytosis in response to EGF and stress. *Molecular & cellular proteomics : MCP* 2014;**13**: 1644-58.

33. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* 2008;**26**: 1367-72.

34. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research* 2011;**10**: 1794-805.

35. Deeb SJ, D'Souza RC, Cox J, Schmidt-Supprian M, Mann M. Super-SILAC allows classification of diffuse large B-cell lymphoma subtypes by their protein expression profiles. *Molecular & cellular proteomics : MCP* 2012;**11**: 77-89.

36. Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, Cho J, Suh J, Capelletti M, Sivachenko A, Sougnez C, Auclair D, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* 2012;**150**: 1107-20.

37. Rikova K, Guo A, Zeng Q, Possemato A, Yu J, Haack H, Nardone J, Lee K, Reeves C, Li Y, Hu Y, Tan Z, et al. Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* 2007;**131**: 1190-203.

38. Mo Q, Wang S, Seshan VE, Olshen AB, Schultz N, Sander C, Powers RS, Ladanyi M, Shen R. Pattern discovery and cancer gene identification in integrated cancer genomic data. *Proceedings of the National Academy of Sciences of the United States of America* 2013;**110**: 4245-50.

39. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehar J, Kryukov GV, Sonkin D, Reddy A, Liu M, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012;**483**: 603-7.

40. Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, Ding M, Bamford S, Cole C, Ward S, Kok CY, Jia M, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic acids research* 2015;**43**: D805-11.

41. Balbin OA, Prensner JR, Sahu A, Yocum A, Shankar S, Malik R, Fermin D, Dhanasekaran SM, Chandler B, Thomas D, Beer DG, Cao X, et al. Reconstructing targetable pathways in lung cancer by integrating diverse omics data. *Nature communications* 2013;**4**: 2617.

42. Rodriguez E, Mannion L, D'Santos P, Griffiths M, Arends MJ, Brindle KM, Lyons SK. Versatile and enhanced tumour modelling in mice via somatic cell transduction. *The Journal of pathology* 2014;**232**: 449-57.

43. Trejo CL, Juan J, Vicent S, Sweet-Cordero A, McMahon M. MEK1/2 inhibition elicits regression of autochthonous lung tumors induced by KRASG12D or BRAFV600E. *Cancer research* 2012;**72**: 3048-59.

For Peer Review

44. Zhang X, Claerhout S, Prat A, Dobrolecki LE, Petrovic I, Lai Q, Landis MD, Wiechmann L, Schiff R, Giuliano M, Wong H, Fuqua SW, et al. A renewable tissue resource of phenotypically stable, biologically and ethnically diverse, patient-derived human breast cancer xenograft models. *Cancer research* 2013;**73**: 4885-97.

45. Lee WS, Kim HY, Seok JY, Jang HH, Park YH, Kim SY, Shin DB, Hong S. Genomic profiling of patient-derived colon cancer xenograft models. *Medicine* 2014;**93**: e298.

46. Williams SA, Anderson WC, Santaguida MT, Dylla SJ. Patient-derived xenografts, the cancer stem cell paradigm, and cancer pathobiology in the 21st century. *Lab Invest* 2013;**93**: 970-82.

47. Eirew P, Steif A, Khattra J, Ha G, Yap D, Farahani H, Gelmon K, Chia S, Mar C, Wan A, Laks E, Biele J, et al. Population dynamics of genomic clones in breast cancer patient xenografts at single cell resolution. *Nature* 2014;**In Publication**.

48. Ikebuchi F, Oka K, Mizuno S, Fukuta K, Hayata D, Ohnishi H, Nakamura T. Dissociation of c-Met phosphotyrosine sites in human cells in response to mouse hepatocyte growth factor but not human hepatocyte growth factor: the possible roles of different amino acids in different species. *Cell biochemistry and function* 2013;**31**: 298-304.

49. Drake AC, Chen Q, Chen J. Engineering humanized mice for improved hematopoietic reconstitution. *Cellular & molecular immunology* 2012;**9**: 215-24.

50. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. *Nature reviews Cancer* 2012;**12**: 252-64.

For Peer Review

Table 1. Demographics of patients from which PDX tumors originated.

		Profiled (n=36)
Age	Median(range)	65.2(44.1-79.9)
Sex	F	14 (38.9%)
	M	22 (61.1%)
Histology	AdC	15 (41.7%)
	SqCC	18 (50.0%)
	Sarcomatoid	1 (2.8%)
	LCNEC	2 (5.5%)
Grade	WD	1 (2.8%)
	MD	15 (41.7%)
	PD	20 (55.5%)
Smoking	No	4 (11.1%)
	Yes	32 (88.9%)
Stage	I	19 (52.8%)
	II	5 (13.9%)
	III	8 (22.2%)
	IV	4 (11.1%)
T category	T1	4 (11.4%)
	T2	26 (72.2%)
	T3-4	6 (17.1%)
N category	N0	24 (66.7%)
	N1-3	12 (33.3%)
M category	M0	32 (88.9%)
	M1	4 (11.4%)
Tumor size	Median(range)	4.5 (2.1-11)
Adjuvant Radiation	N	32 (88.9%)
	Y	4 (11.1%)
Adjuvant Chemo.	N	21 (58.3%)
	Y	15 (41.7%)
Recurrence rate	CIF at 3 years	38%
Survival	KM at 3 years	53%
Follow-up, years	Median(range)	5.3 (0.2-7.3)

WD, well differentiated; MD, moderately differentiated; PD, poorly differentiated; CIF, cumulative incidence function; KM, Kaplan-Meier.

For Peer Review

Figure legends

Figure 1. Establishment of PDX models from NSCLC patient tumors. (A) A flow diagram describes the steps taken to establish a stable PDX line. (B) Kaplan-Meier survival curves of stable PDX (XG) and no-PDX (no-XG) patients show a significant difference in disease-free survival. (C) Multivariate analysis shows xenograft engraftment as a prognostic feature. Representative histology of the four NSCLC subtypes show paired patient-PDX tumors: adenocarcinoma (D/E, model 82, A-patient/B-xeno), squamous cell carcinoma (F/G, model 57), sarcomatoid (H/I, model 134) and large cell neuroendocrine carcinoma (J/K, model 88). Slides were stained with hematoxylin and eosin and images were captured on an Aperio ScanScope XT (Leica Biosystems).

Figure 2. Somatic genomic alterations in NSCLC PDXs and in their corresponding patients. (A) Copy number alteration frequencies across the genome for PDXs and corresponding patient tumors. A more stringent threshold for copy gain and loss was applied to PDX copy number profiles compared to that of patient tumor. Gene alterations characteristic of adenocarcinoma (AdC) and squamous cell carcinoma (SqCC) are labelled.

Figure 3. Comparative phosphotyrosine profiles of NSCLC PDXs, cell lines and primary tumors. The activated tyrosine phosphorylation in 365 proteins in our PDXs was compared to publicly available NSCLC datasets of normal, patient, and cell line samples reported by Rikova, *et al.* [37]. (A) Samples along columns are sorted by hierarchical clustering across phosphorylated proteins. Bar plots underneath heatmaps show the relative proportions of proteins with a spectral count >5. (B) Tyrosine kinase phosphorylation profiles in our PDX are compared to normal,

For Peer Review

patient, and cell line samples from Rikova, *et al*. Samples are grouped by hierarchical clustering (dendogram), and the first two principal components (scatterplot). (C) A heatmap comparing phosphorylation status of receptor tyrosine kinase proteins and phosphatases. Phosphorylated proteins in yellow have a spectral count ≥ 1. Samples are ordered by hierarchical clustering showing PDXs (green), cell lines (pink) and patient tumor (blue).

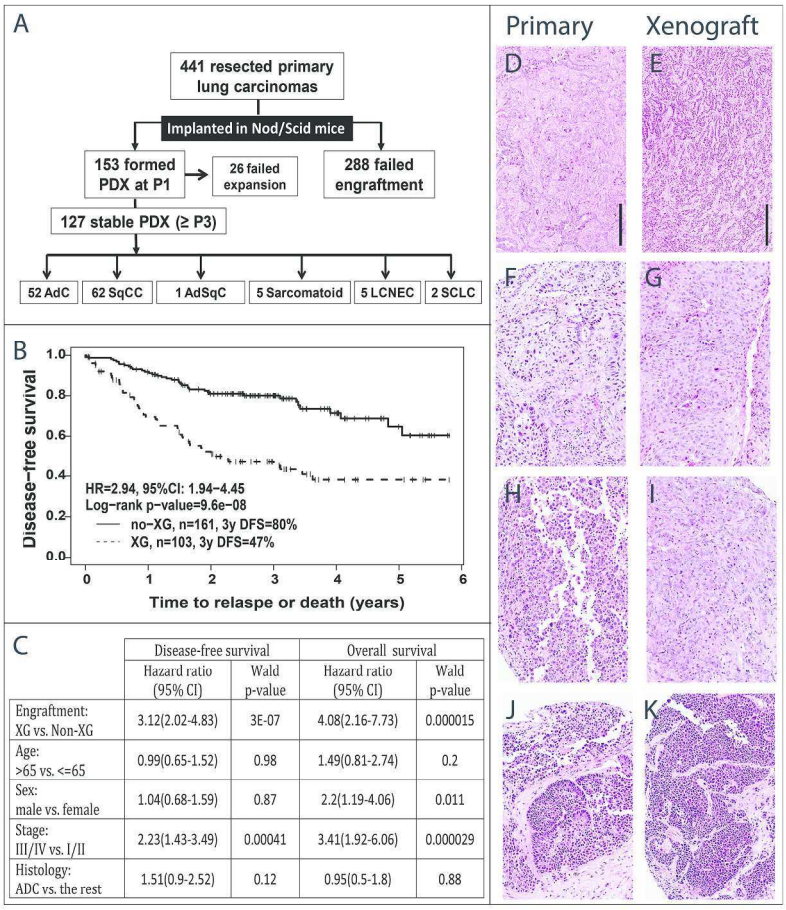
Figure 4. PDXs capture gene expression and methylation patterns in corresponding patient tumors. Heat maps show hierarchical clustering of (A) the 200 most variable DNA methylation sites at promoter regions, or (B) the 200 most variable mRNA expressions of xenograft tumors corresponds to the histology subtypes of matched patient tumors. (C) The principal component clustering of DNA methylation profiles in TCGA patient tumors and PDXs is separated by histological subtypes. (D) Gene set enrichment analysis was performed on genes highly expressed in either AdC or SqCC of PDXs and compared to patient AdC or SqCC of GSE14814 dataset in a hypergeometric distribution.

Figure 5. Key of aberrations in cancer associated genes and pathways. (A) Verified mutations in key cancer associated genes across PDX tumors. The percentage of lung xenograft models is shown by the bar graph. Samples are listed according to histology classification. Somatically mutated genes listed vertically are grouped as AdC (red), SqCC (blue), Other (green) based on their prevalence in human patient tumors. Within each group, the genes are listed in decreasing order of nonsilent somatic mutation prevalence in xenografts. Colored rectangles indicate mutation category in a given gene and tumor. White boxes indicate unknown status or silent mutations. (B) Genomic alterations (somatic mutation or copy number loss/gain) across 36 PDXs in cancer associated pathways that explained a > 20% genomic variation

For Peer Review

according to iClusterPlus. (C) Expression signature of 100 genes selected by RandomForest as correlated with the pattern of the DNA alterations in panel (B). High expression (> 2 z-score) are colored red and low expression (< -2 z-score) are colored blue.

For Peer Review



	Disease-free survival		Overall survival	
	Hazard ratio (95% CI)	Wald p-value	Hazard ratio (95% CI)	Wald p-value
Engraftment: XG vs. Non-XG	3.12(2.02-4.83)	3E-07	4.08(2.16-7.73)	0.000015
Age: >65 vs. <=65	0.99(0.65-1.52)	0.98	1.49(0.81-2.74)	0.2
Sex: male vs. female	1.04(0.68-1.59)	0.87	2.2(1.19-4.06)	0.011
Stage: III/IV vs. I/II	2.23(1.43-3.49)	0.00041	3.41(1.92-6.06)	0.000029
Histology: ADC vs. the rest	1.51(0.9-2.52)	0.12	0.95(0.5-1.8)	0.88

Figure 1. Establishment of PDX models from NSCLC patient tumors. (A) A flow diagram describes the steps taken to establish a stable PDX line. (B) Kaplan-Meier survival curves of stable PDX (XG) and no-PDX (no-XG) patients show a significant difference in disease-free survival. (C) Multivariate analysis shows xenograft engraftment as a prognostic feature. Representative histology of the four NSCLC subtypes show paired patient-PDX tumors: adenocarcinoma (D/E, model 82, A-patient/B-xeno), squamous cell carcinoma (F/G, model 57), sarcomatoid (H/I, model 134) and large cell neuroendocrine carcinoma (J/K, model 88). Slides were stained with hematoxylin and eosin and images were captured on an Aperio ScanScope XT (Leica Biosystems).

215x279mm (300 x 300 DPI)

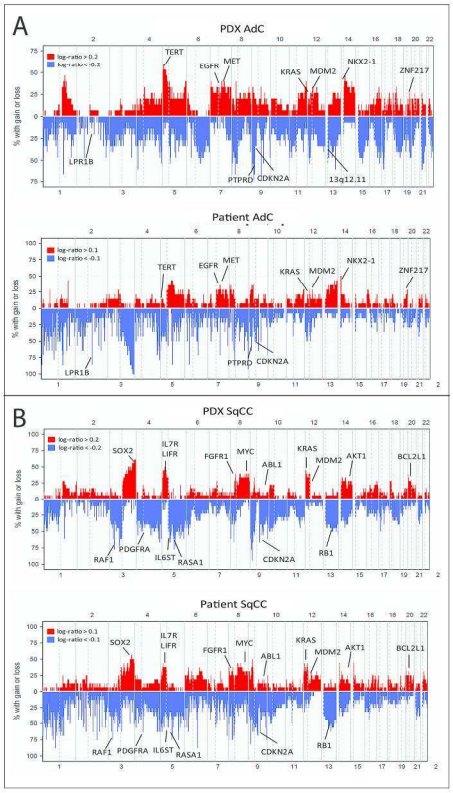


Figure 2. Somatic genomic alterations in NSCLC PDXs and in their corresponding patients. (A) Copy number alteration frequencies across the genome for PDXs and corresponding patient tumors. A more stringent threshold for copy gain and loss was applied to PDX copy number profiles compared to that of patient tumor. Gene alterations characteristic of adenocarcinoma (AdC) and squamous cell carcinoma (SqCC) are labelled.

215x279mm (300 x 300 DPI)

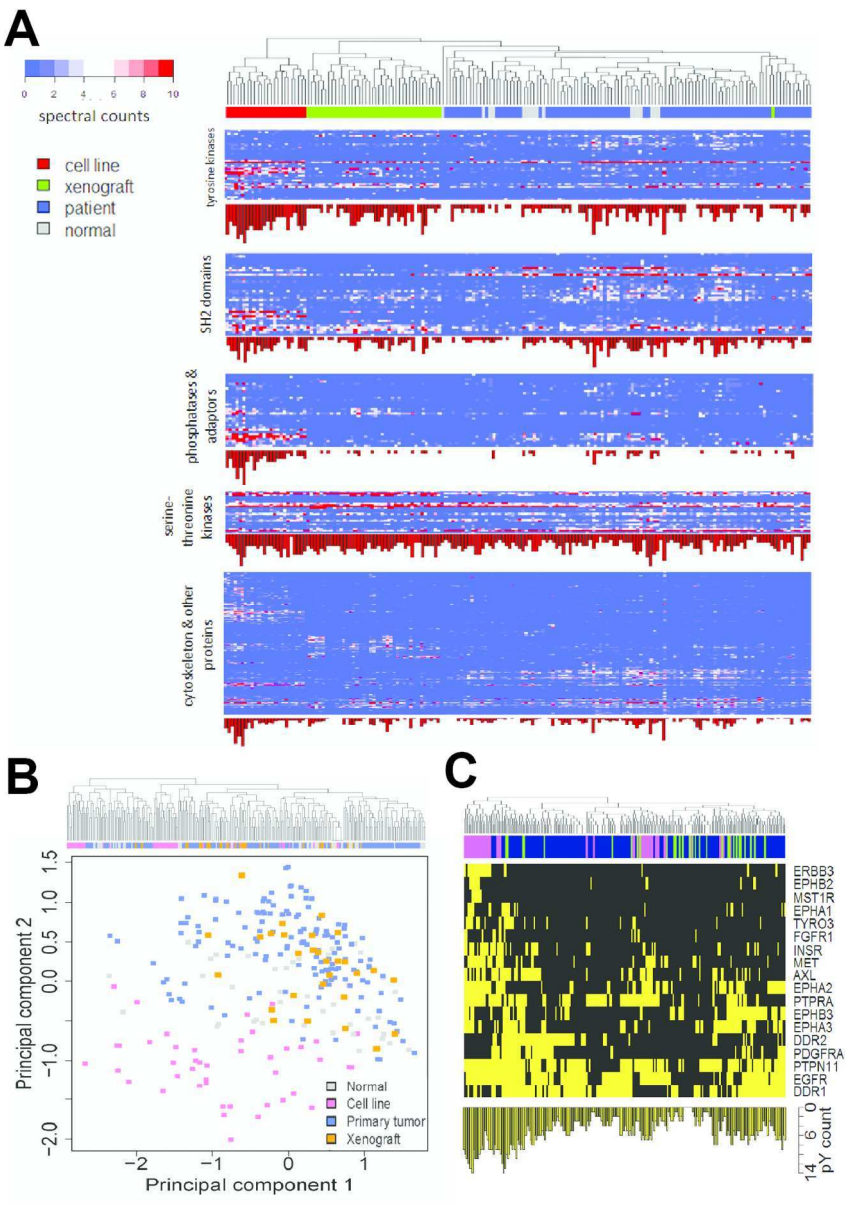


Figure 3. Comparative phosphotyrosine profiles of NSCLC PDXs, cell lines and primary tumors. The activated tyrosine phosphorylation in 365 proteins in our PDXs was compared to publicly available NSCLC datasets of normal, patient, and cell line samples reported by Rikova, et al. 37. (A) Samples along columns are sorted by hierarchical clustering across phosphorylated proteins. Bar plots underneath heatmaps show the relative proportions of proteins with a spectral count >5. (B) Tyrosine kinase phosphorylation profiles in our PDX are compared to normal, patient, and cell line samples from Rikova, et al. Samples are grouped by hierarchical clustering (dendogram), and the first two principal components (scatterplot). (C) A heatmap comparing phosphorylation status of receptor tyrosine kinase proteins and phosphatases. Phosphorylated proteins in yellow have a spectral count ≥ 1. Samples are ordered by hierarchical clustering showing PDXs (green), cell lines (pink) and patient tumor (blue).

125x179mm (300 x 300 DPI)

For Peer Review

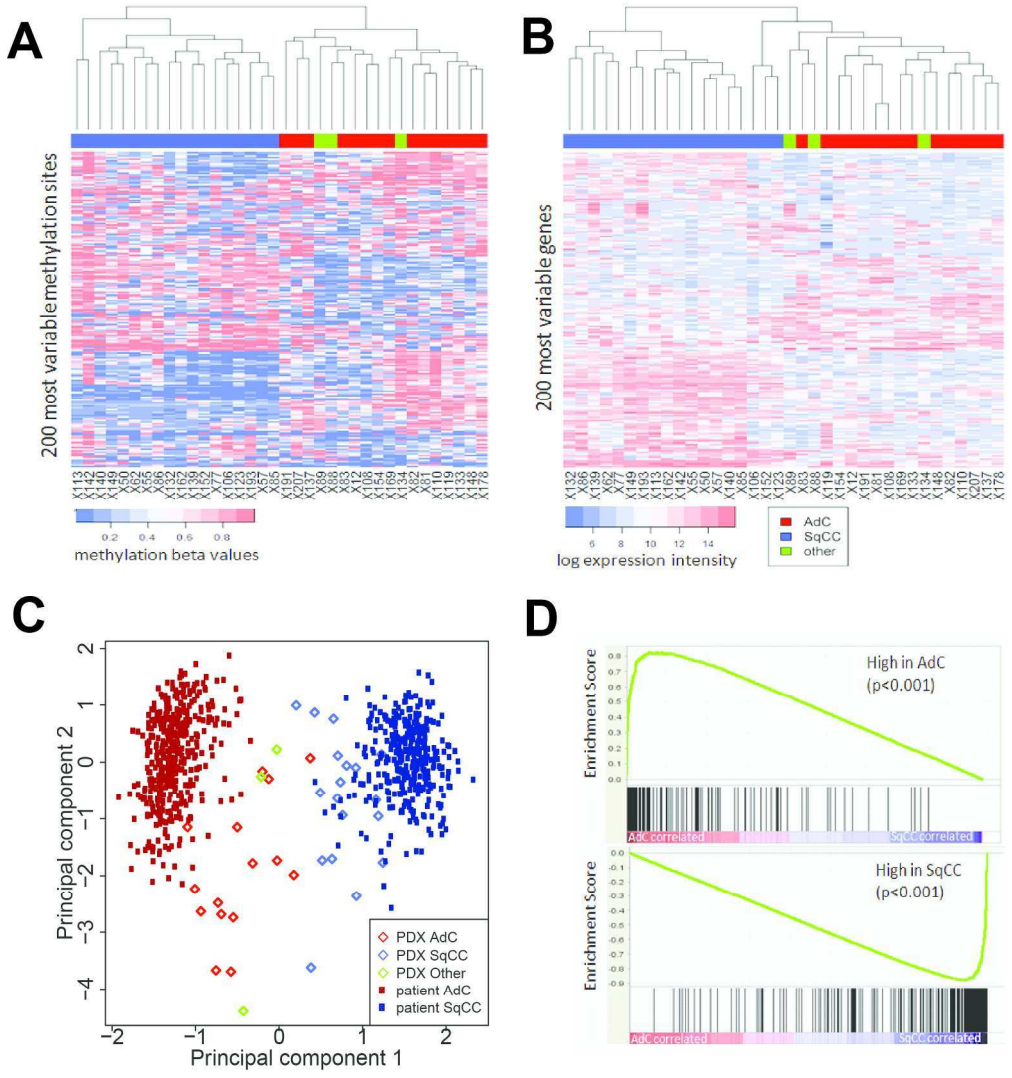


Figure 4. PDXs capture gene expression and methylation patterns in corresponding patient tumors. Heat maps show hierarchical clustering of (A) the 200 most variable DNA methylation sites at promoter regions, or (B) the 200 most variable mRNA expressions of xenograft tumors corresponds to the histology subtypes of matched patient tumors. (C) The principal component clustering of DNA methylation profiles in TCGA patient tumors and PDXs is separated by histological subtypes. (D) Gene set enrichment analysis was performed on genes highly expressed in either AdC or SqCC of PDXs and compared to patient AdC or SqCC of GSE14814 dataset in a hypergeometric distribution.

212x230mm (300 x 300 DPI)

Figure 5. Key of aberrations in cancer associated genes and pathways. (A) Verified mutations in key cancer associated genes across PDX tumors. The percentage of lung xenograft models is shown by the bar graph. Samples are listed according to histology classification. Somatically mutated genes listed vertically are grouped as AdC (red), SqCC (blue), Other (green) based on their prevalence in human patient tumors. Within each group, the genes are listed in decreasing order of nonsilent somatic mutation prevalence in xenografts. Colored rectangles indicate mutation category in a given gene and tumor. White boxes indicate unknown status or silent mutations. (B) Genomic alterations (somatic mutation or copy number loss/gain) across 36 PDXs in cancer associated pathways that explained a > 20% genomic variation according to iClusterPlus. (C) Expression signature of 100 genes selected by RandomForest as correlated with the pattern of the DNA alterations in panel (B). High expression (> 2 z-score) are colored red and low expression (< -2 z-score) are colored blue.

388x426mm (300 x 300 DPI)