Supplementary Information

Patient-derived xenografts undergo murine-specific tumor evolution

Uri Ben-David[1], Gavin Ha[1,2], Yuen-Yi Tseng[1], Noah F. Greenwald[1,3,4], Coyin Oh[1], Juliann Shih[1,2], James M. McFarland[1], Bang Wong[1], Jesse S. Boehm[1], Rameen Beroukhim[1,2,3,5,6,10] and Todd R. Golub[1,7,8,9,10]

[1] Cancer Program, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA
[2] Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA
[3] Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA
[4] Department of Neurosurgery, Brigham and Women's Hospital, Boston, Massachusetts, USA
[5] Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA
[6] Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA
[7] Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA
[8] Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, USA
[9] Howard Hughes Medical Institute, Chevy Chase, Maryland, USA
[10] These authors jointly supervised this work.

Correspondence should be addressed to R.B. (rameen_beroukhim@dfci.harvard.edu) or T.R.G. (golub@broadinstitute.org)

Table of contents

Supplementary Figures:

Supplementary Figure 1: CNA profiles from PDX gene expression data are highly similar to those from PDX SNP array data

Supplementary Figure 2: The CNA landscapes of PDXs are highly similar to those of primary tumors from matched tissues

Supplementary Figure 3: Gradual evolution of CNA landscapes throughout PDX passaging

Supplementary Figure 4: Shifts in the allelic fraction of point mutations during PDX derivation and propagation

Supplementary Figure 5: PDX models from metastases exhibit larger CNA fractions and higher CIN70 scores than PDX models from primary tumors

Supplementary Figure 6: Expansion of pre-existing subclones during PDX propagation demonstrated by identification of LOH "reversion"

Supplementary Figure 7: Genomic instability in PDXs correlates both the genomic instability and the heterogeneity levels of primary tumors

Supplementary Figure 8: Disappearance of recurrent CNAs throughout PDX propagation: opposite trends of patient-acquired and model-acquired CNAs

Supplementary Figure 9: Disappearance of recurrent CNAs throughout PDX propagation: prevalence differences between early and late passages

Supplementary Figure 10: Genomic instability of PDXs is comparable to that of cell lines and CLDXs

Supplementary Figure 11: Arm-level CNAs affect genetic dependencies and drug sensitivities

Supplementary Figure 12: Arm-level CNAs are associated with extensive gene expression changes in cell lines

Supplementary Figure 13: CNA-based and mutation-based phylogenetic trees are highly concordant

Supplementary Figure 14: *De novo* CNAs may play a role in PDX CNA dynamics as well

Supplementary Tables:

Supplementary Table 1: Summary of PDX datasets

Supplementary Table 2: Comparison of DNA- and RNA-based CNA profiles

Supplementary Table 3: Summary of advanced disease datasets

Supplementary Table 4: Summary of newly-derived cell lines

Supplementary Table 5: Comparisons of gene expression and genetic dependencies in cell lines with arm-level CNAs and cell lines without them

Supplementary Table 6: Comparison of drug response in cell lines with arm-level CNAs and cell lines without them

Legends to Supplementary Datasets:

Supplementary Dataset 1: CNA profiles of PDX samples

Supplementary Dataset 2: Model-acquired CNAs in PDX samples

Supplementary Dataset 3: CNA profiles of CLDX samples

Supplementary Dataset 4: Model-acquired CNAs in CLDX samples

Supplementary Note:

1. Supplementary Introduction: Previous evidence for genomic instability of PDXs

2. Supplementary Introduction: Somatic copy number alterations (CNAs) in cancer

3. Supplementary Results: Comparison of DNA- and RNA-derived copy number profiles

4. Supplementary Results: Changes in the allelic fraction of point mutations throughout PDX passaging

5. Supplementary Results: Deviation of melanoma from the observed correlation between PDX instability and primary tumor heterogeneity

6. Supplementary Results: Association between arm-level CNAs and cell line gene expression, genetic dependencies and drug response

7. Supplementary Discussion: Comparison of our findings to previous studies

8. Supplementary Discussion: The relevance of our CNA-based analysis to other types of genetic alterations

9. Supplementary Discussion: Combination of pre-existing and *de novo* occurring CNAs in PDXs

10. Supplementary Discussion: Distinct selection pressures in PDXs and in patients

11. Supplementary Discussion: Implications of PDX genomic instability for their use in drug testing

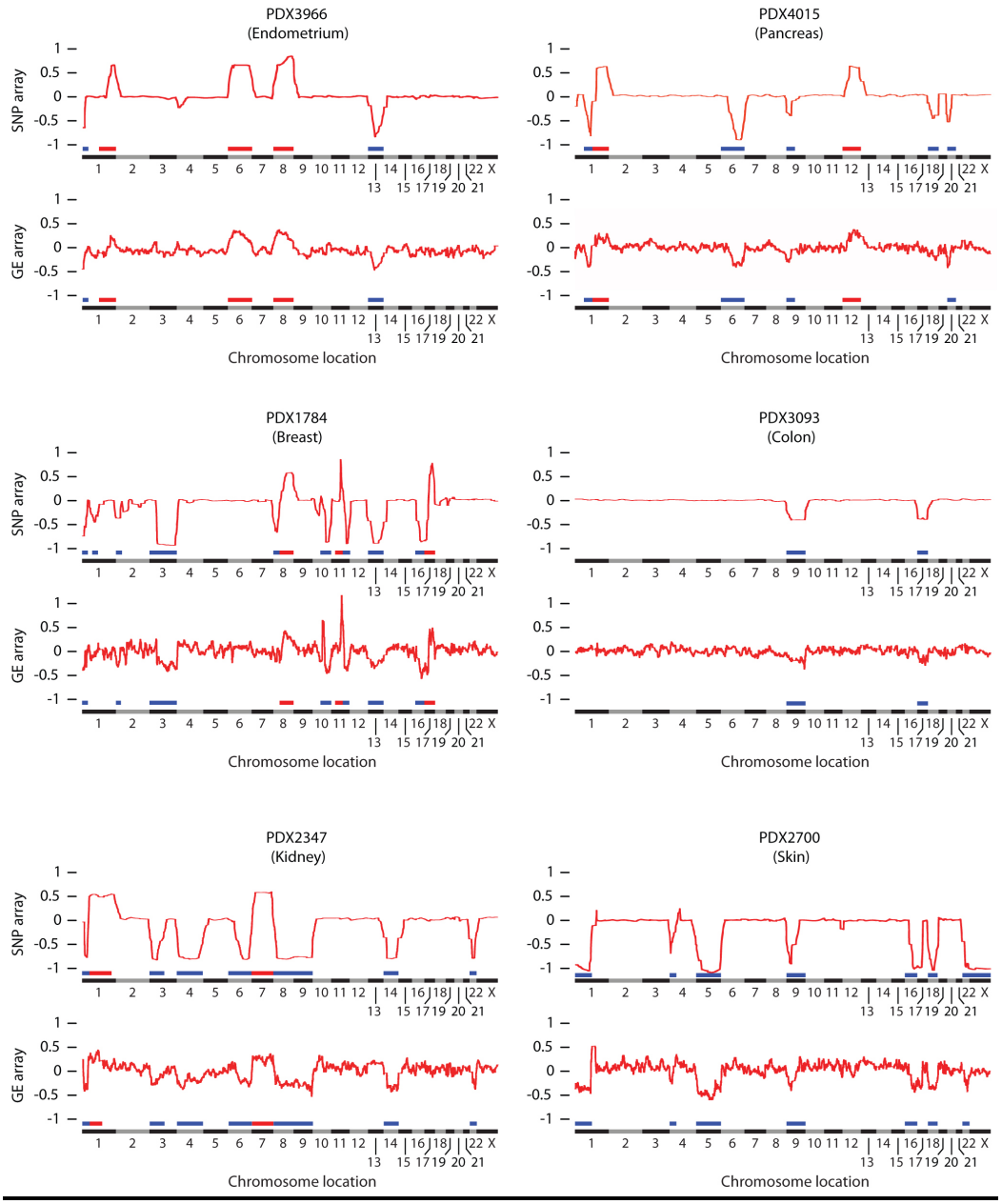
12. Supplementary Discussion: Possible explanations for the difference in CNA acquisition rate between PDXs and CLDXs

13. Supplementary Discussion: Implications beyond cancer model systems

Supplementary References

Supplementary Figures

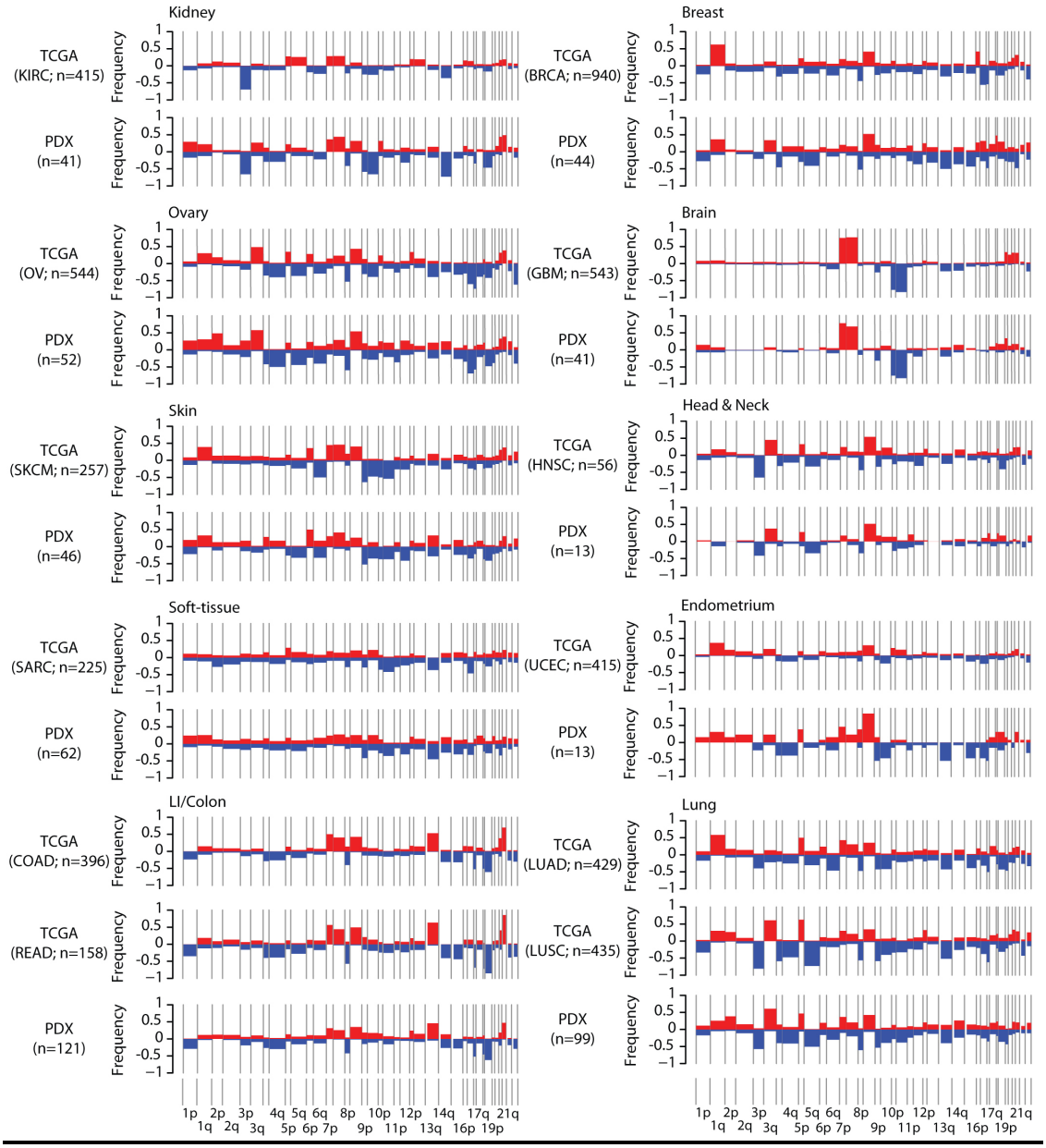
Figure S1



Supplementary Figure 1: CNA profiles from PDX gene expression data are highly similar to those from PDX SNP array data

Moving average plots of PDX SNP arrays (upper panels) and their corresponding gene expression arrays (lower panels) in six representative cancer types. The CNAs identified in each sample by our pipeline are depicted as rectangles above the affected genomic regions. Gains are shown in red, losses in blue.

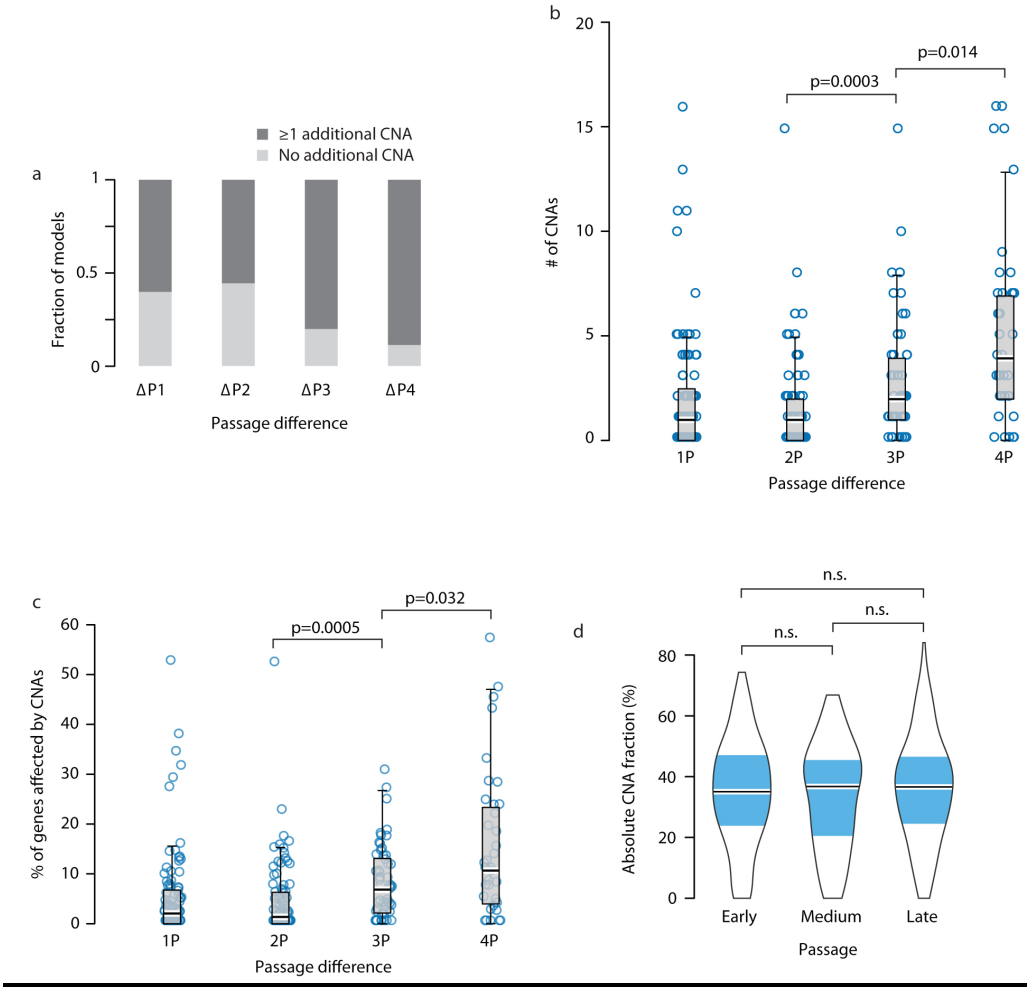
Figure S2



Supplementary Figure 2: The CNA landscapes of PDXs are highly similar to those of primary tumors from matched tissues

CNA frequency plots of PDX model types and the respective primary tumor types from TCGA, showing that PDXs generally exhibit the aneuploidies and CNAs that are characteristic of each tissue type. Gains are shown in red, losses in blue.

Figure S3



Supplementary Figure 3: Gradual evolution of CNA landscapes throughout PDX passaging

PDX models acquire CNAs throughout their *in vivo* propagation. **(a)** Bar plots present the fraction of the PDX models with at least one model-acquired CNA, as a function of the number of passages between measurements. (**b**) Box plots present the number of discrete CNAs as a function of the number of passages between measurements. Bar, median; box, 25th and 75th percentiles; whiskers, data within 1.5*IQR of lower or upper quartile; circles: all data points. P-values indicate significance from a Wilcoxon rank-sum test. (**c**) Box plots present the proportion of genes affected by CNAs as a function of the number of passages between measurements. Bar, median; box, 25th and 75th percentiles; whiskers, data within 1.5*IQR of lower or upper quartile; circles: all data points. P-values indicate significance from a Wilcoxon rank-sum test. (**d**) Equal rates of acquiring new CNAs and losing existing ones in PDXs. Violin plots present the absolute CNA fraction of PDX models at early ($p<2$), medium ($2<p<4$) and late ($p>=4$) passages. Bar, median; colored rectangle, 25th and 75th percentiles; width of the violin indicates frequency at that CNA fraction level. n.s., non-significant (Wilcoxon rank-sum test).

Figure S4

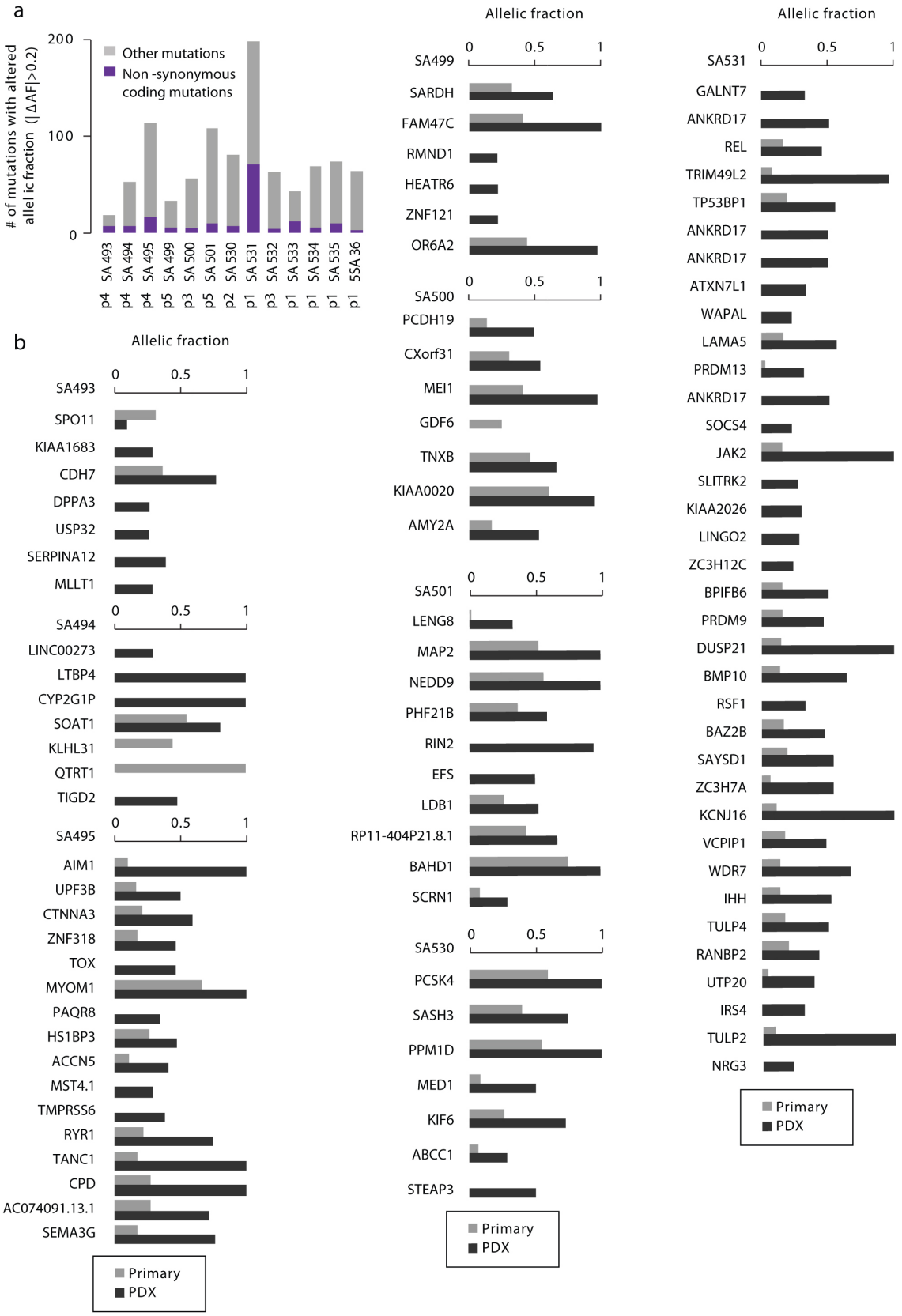
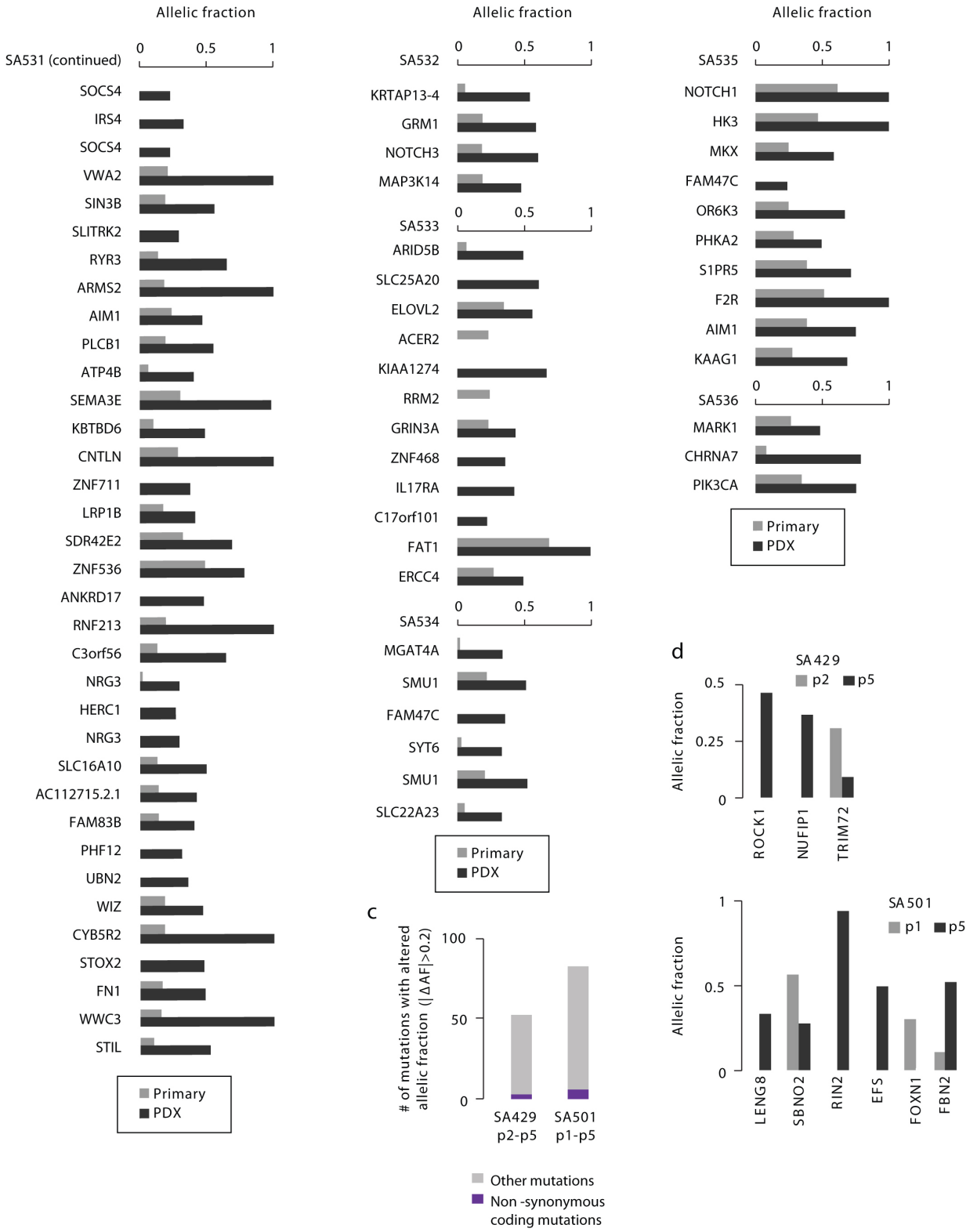

a
# of mutations with altered allelic fraction (|ΔAF|>0.2)
200
100
0
Other mutations
Non -synonymous coding mutations
SA 493
SA 494
SA 495
SA 499
SA 500
SA 501
SA 530
SA 531
SA 532
SA 533
SA 534
SA 535
5SA 36
p4
p4
p4
p5
p3
p5
p2
p1
p3
p1
p1
p1
p1
b
Allelic fraction
SA493
SPO11
KIAA1683
CDH7
DPPA3
USP32
SERPINA12
MLLT1
SA494
LINC00273
LTBP4
CYP2G1P
SOAT1
KLHL31
QTRT1
TIGD2
SA495
AIM1
UPF3B
CTNNA3
ZNF318
TOX
MYOM1
PAQR8
HS1BP3
ACCN5
MST4.1
TMPRSS6
RYR1
TANC1
CPD
AC074091.13.1
SEMA3G
Primary
PDX
Allelic fraction
SA499
SARDH
FAM47C
RMND1
HEATR6
ZNF121
OR6A2
SA500
PCDH19
CXorf31
MEI1
GDF6
TNXB
KIAA0020
AMY2A
SA501
LENG8
MAP2
NEDD9
PHF21B
RIN2
EFS
LDB1
RP11-404P21.8.1
BAHD1
SCRN1
SA530
PCSK4
SASH3
PPM1D
MED1
KIF6
ABCC1
STEAP3
Primary
PDX
Allelic fraction
SA531
GALNT7
ANKRD17
REL
TRIM49L2
TP53BP1
ANKRD17
ANKRD17
ATXN7L1
WAPAL
LAMA5
PRDM13
ANKRD17
SOCS4
JAK2
SLITRK2
KIAA2026
LINGO2
ZC3H12C
BPIFB6
PRDM9
DUSP21
BMP10
RSF1
BAZ2B
SAYSD1
ZC3H7A
KCNJ16
VCPIP1
WDR7
IHH
TULP4
RANBP2
UTP20
IRS4
TULP2
NRG3
Primary
PDX

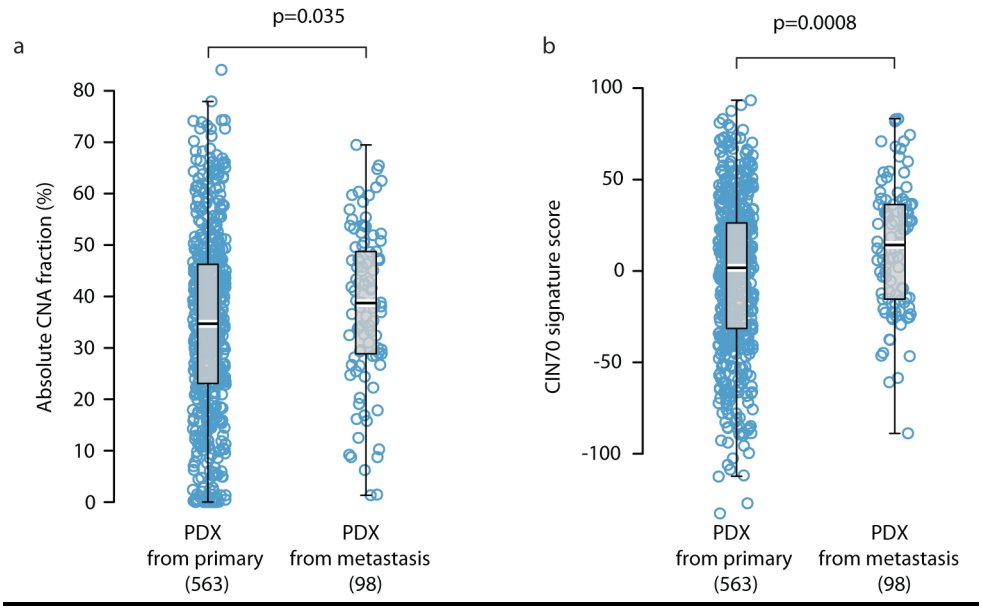
Figure S4 (continued)



Supplementary Figure 4: Shifts in the allelic fraction of point mutations during PDX derivation and propagation

(**a**) The allelic fractions (AF) of point mutations can change throughout PDX passaging. Plots present the number of point mutations with an AF shift (|ΔAF|>0.2) between the primary tumor and its derived xenograft, in 13 matched pairs of primary breast cancer tumors and PDXs from Eirew et al. [1]. (**b**) AF shifts in non-synonymous missense and nonsense coding mutations in each of the matched pairs described in (**a**). (**c**) The allelic fractions (AF) of point mutations can change throughout PDX passaging. Plots present the number of point mutations with an AF shift (|ΔAF|>0.2) between early (p=1 or p=2) and late (p=5) PDXs, in two matched pairs of tumor xenografts from Eirew et al. [1]. (**d**) AF shifts in non-synonymous missense and nonsense coding mutations in each of the matched pairs described in (**c**).

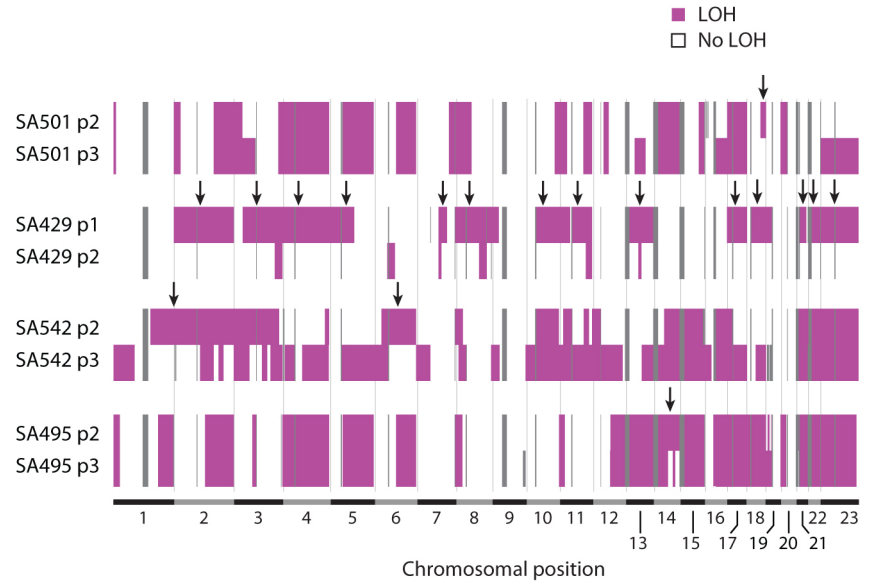
Figure S5



Supplementary Figure 5: PDX models from metastases exhibit larger CNA fractions and higher CIN70 scores than PDX models from primary tumors

PDX models from metastases are more aneuploid than those from primary tumors. (**a**) Box plots present the absolute CNA fraction of PDX models from primary tumors (n=563) and from metastases (n=98). (**b**) Box plots present chromosomal instability (CIN70) signature scores of PDX models from primary tumors (n=563) and from metastases (n=98). P-values indicate significance from Wilcoxon rank-sum tests.

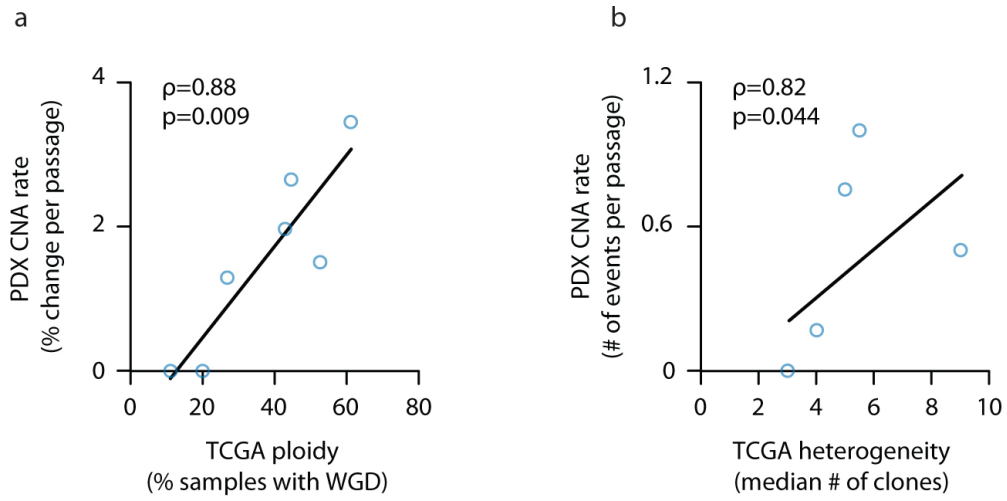
Figure S6



Supplementary Figure 6: Expansion of pre-existing subclones during PDX propagation demonstrated by identification of LOH "reversion"

Alleles that seem to have been lost in early-passage PDX tumors can "re-appear" in later passages of the same PDX models, demonstrating expansion of rare pre-existing subclones throughout PDX propagation. Plots present the loss of heterozygosity (LOH) status along the genomes of four PDX models from Eirew et al. [1]. LOH events are shown in purple. For each model, shown are two passages. Arrows mark large (>10Mb) chromosomal segments for which LOH was identified at the earlier passage, but both alleles were present at the later passage.

Figure S7



Supplementary Figure 7: Genomic instability in PDXs correlates both the genomic instability and the heterogeneity levels of primary tumors

(**a**) The DGI of PDXs and that of primary tumors correlate extremely well. In PDXs, tissue DGI was defined as the median CNA fraction affected per passage. In TCGA tumors, tissue DGI was defined as the fraction of samples with whole-genome duplication (WGD). (**b**) The DGI of PDXs also correlates extremely well with intra-tumor heterogeneity (ITH) of primary tumors (including skin tissue). The DGI of PDXs was defined as the median number of arm-level CNAs per passage. The heterogeneity of primary tumors was defined as the median number of clones per tumor. Spearman's rho values and p-values indicate the strength and significance of the correlations, respectively.

Figure S8

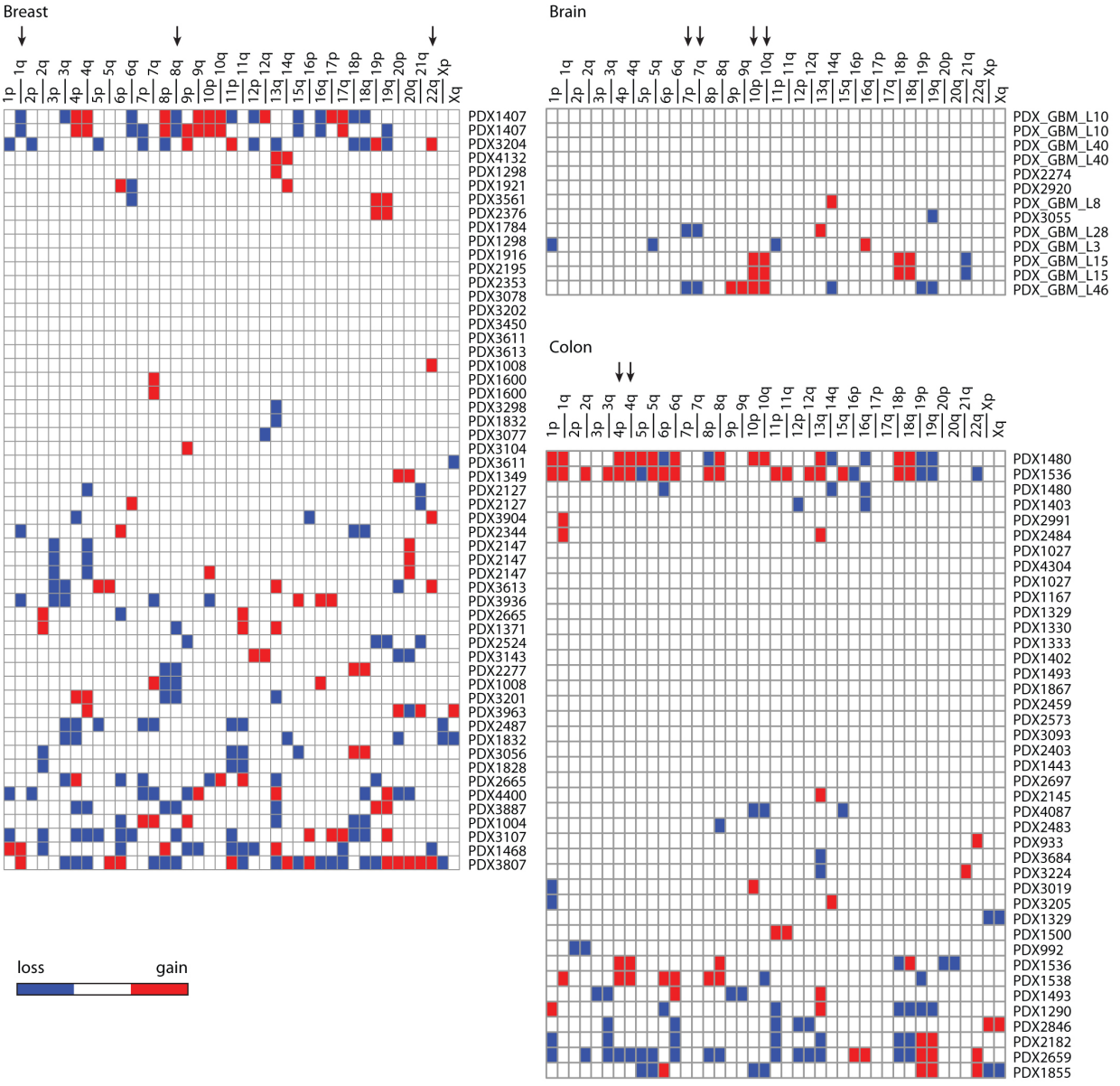
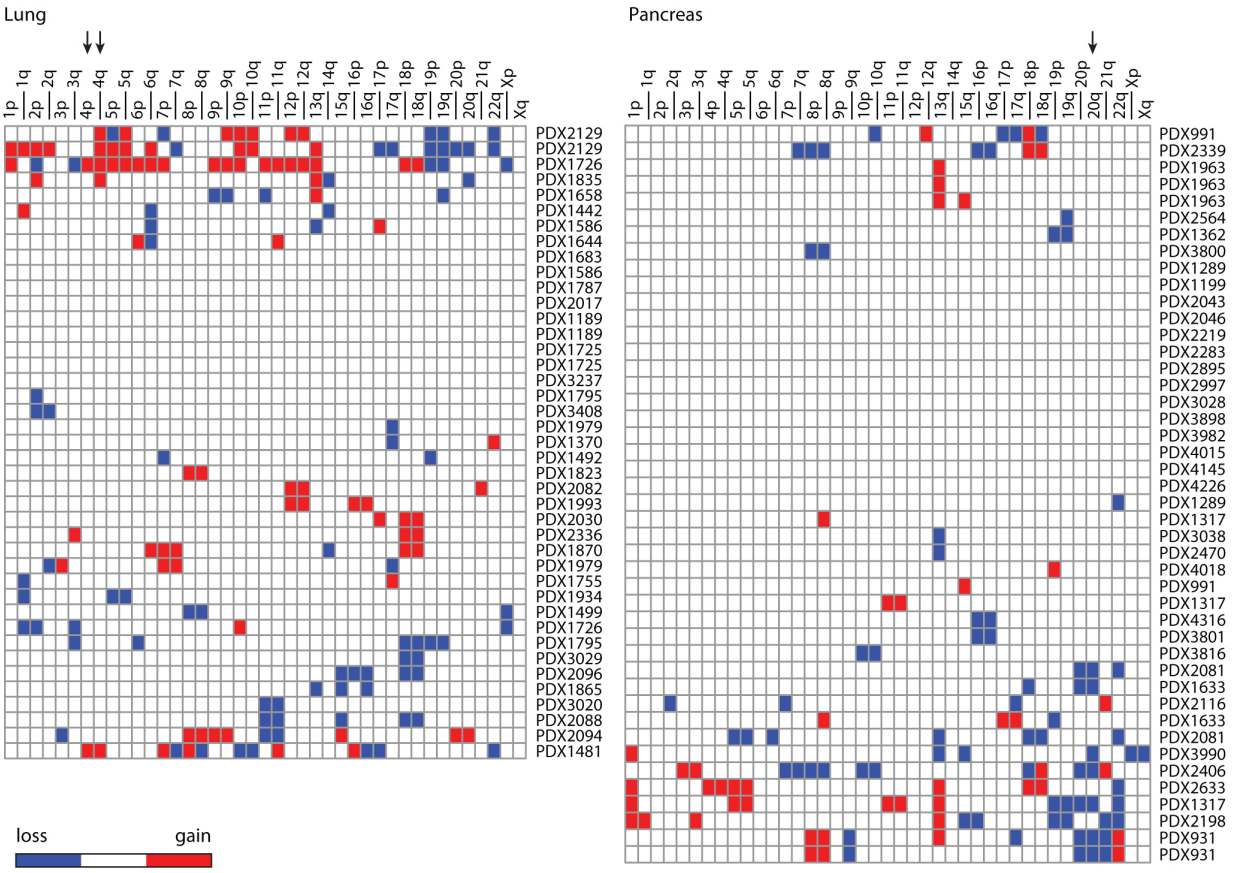

Breast
1p 1q 2p 2q 3p 3q 4p 4q 5p 5q 6p 6q 7p 7q 8p 8q 9p 9q 10p 10q 11p 11q 12p 12q 13q 14q 15q 16p 16q 17p 17q 18p 18q 19p 19q 20p 20q 21q 22q Xp Xq
PDX1407
PDX1407
PDX3204
PDX4132
PDX1298
PDX1921
PDX3561
PDX2376
PDX1784
PDX1298
PDX1916
PDX2195
PDX2353
PDX3078
PDX3202
PDX3450
PDX3611
PDX3613
PDX1008
PDX1600
PDX1600
PDX3298
PDX1832
PDX3077
PDX3104
PDX3611
PDX1349
PDX2127
PDX2127
PDX3904
PDX2344
PDX2147
PDX2147
PDX2147
PDX3613
PDX3936
PDX2665
PDX1371
PDX2524
PDX3143
PDX2277
PDX1008
PDX3201
PDX3963
PDX2487
PDX1832
PDX3056
PDX1828
PDX2665
PDX4400
PDX3887
PDX1004
PDX3107
PDX1468
PDX3807
loss
gain
Brain
1p 1q 2p 2q 3p 3q 4p 4q 5p 5q 6p 6q 7p 7q 8p 8q 9p 9q 10p 10q 11p 11q 12p 12q 13q 14q 15q 16p 16q 17p 17q 18p 18q 19p 19q 20p 20q 21q 22q Xp Xq
PDX_GBM_L10
PDX_GBM_L10
PDX_GBM_L40
PDX_GBM_L40
PDX2274
PDX2920
PDX_GBM_L8
PDX3055
PDX_GBM_L28
PDX_GBM_L3
PDX_GBM_L15
PDX_GBM_L15
PDX_GBM_L46
Colon
1p 1q 2p 2q 3p 3q 4p 4q 5p 5q 6p 6q 7p 7q 8p 8q 9p 9q 10p 10q 11p 11q 12p 12q 13q 14q 15q 16p 16q 17p 17q 18p 18q 19p 19q 20p 20q 21q 22q Xp Xq
PDX1480
PDX1536
PDX1480
PDX1403
PDX2991
PDX2484
PDX1027
PDX4304
PDX1027
PDX1167
PDX1329
PDX1330
PDX1333
PDX1402
PDX1493
PDX1867
PDX2459
PDX2573
PDX3093
PDX2403
PDX1443
PDX2697
PDX2145
PDX4087
PDX2483
PDX933
PDX3684
PDX3224
PDX3019
PDX3205
PDX1329
PDX1500
PDX992
PDX1536
PDX1538
PDX1493
PDX1290
PDX2846
PDX2182
PDX2659
PDX1855

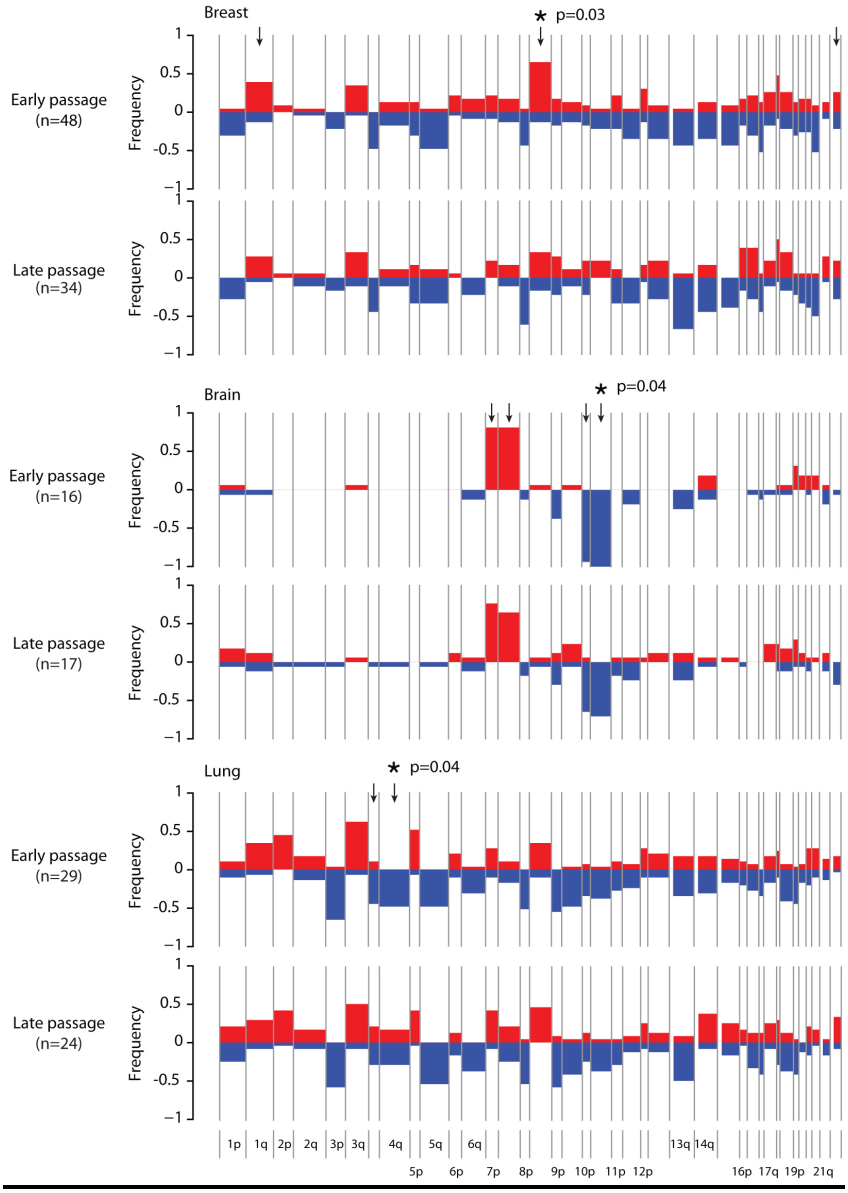
Figure S8 (continued)



Supplementary Figure 8: Disappearance of recurrent CNAs throughout PDX propagation: opposite trends of patient-acquired and model-acquired CNAs

Twelve recurrent arm-level CNAs, which were observed in >40% of TCGA samples, were found to be preferentially lost during PDX passaging. Heatmaps present the model-acquired arm-level CNAs identified in five PDX tumor types: breast, brain, colon, lung and pancreas. Gains are shown in red, losses in blue. The chromosome arms that show an opposite acquisition trend to that seen in human patients are highlighted with arrows. 86% of these events represent the disappearance of a CNA that existed at an earlier passage, rather than the acquisition of the opposite CNA.

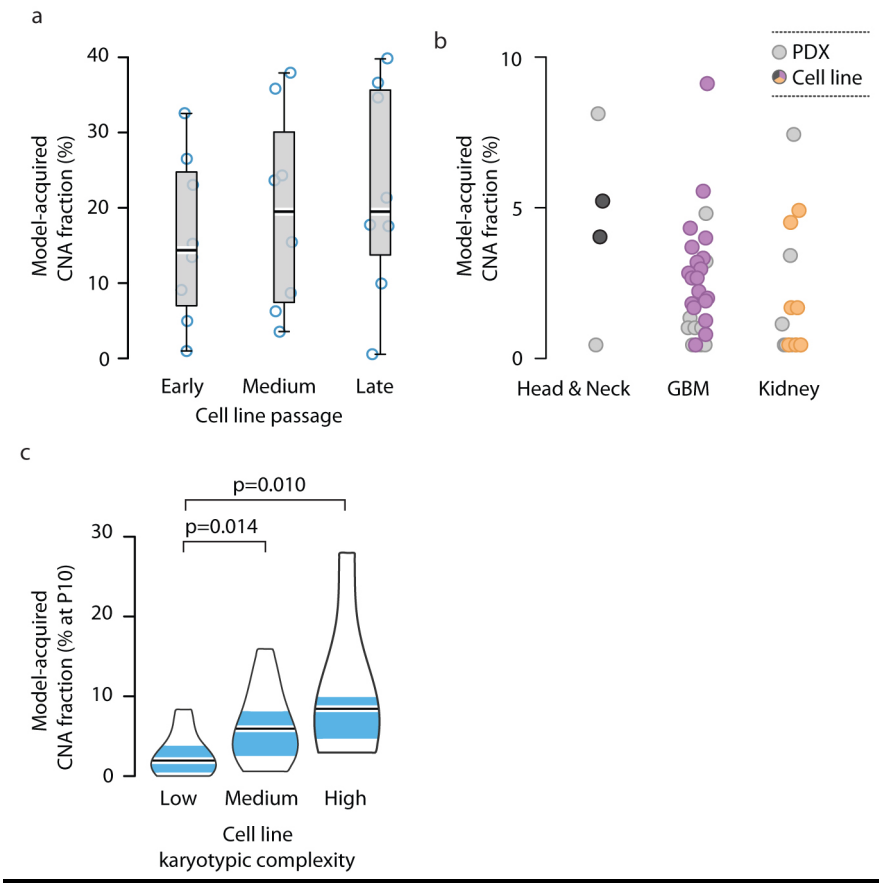
Figure S9



Supplementary Figure 9: Disappearance of recurrent CNAs throughout PDX propagation: prevalence differences between early and late passages

Recurrent CNAs that tend to disappear during PDX passaging are less commonly identified in late compared to early passage PDX samples. Absolute CNA frequency plots of three PDX model types (breast, brain and lung) at early and late passage numbers are presented. Gains are shown in red, losses in blue. Nine of the twelve events that tend to disappear in PDXs are less common in high passage PDXs (highlighted by arrows). P-values indicate significance from a Fisher's exact test.

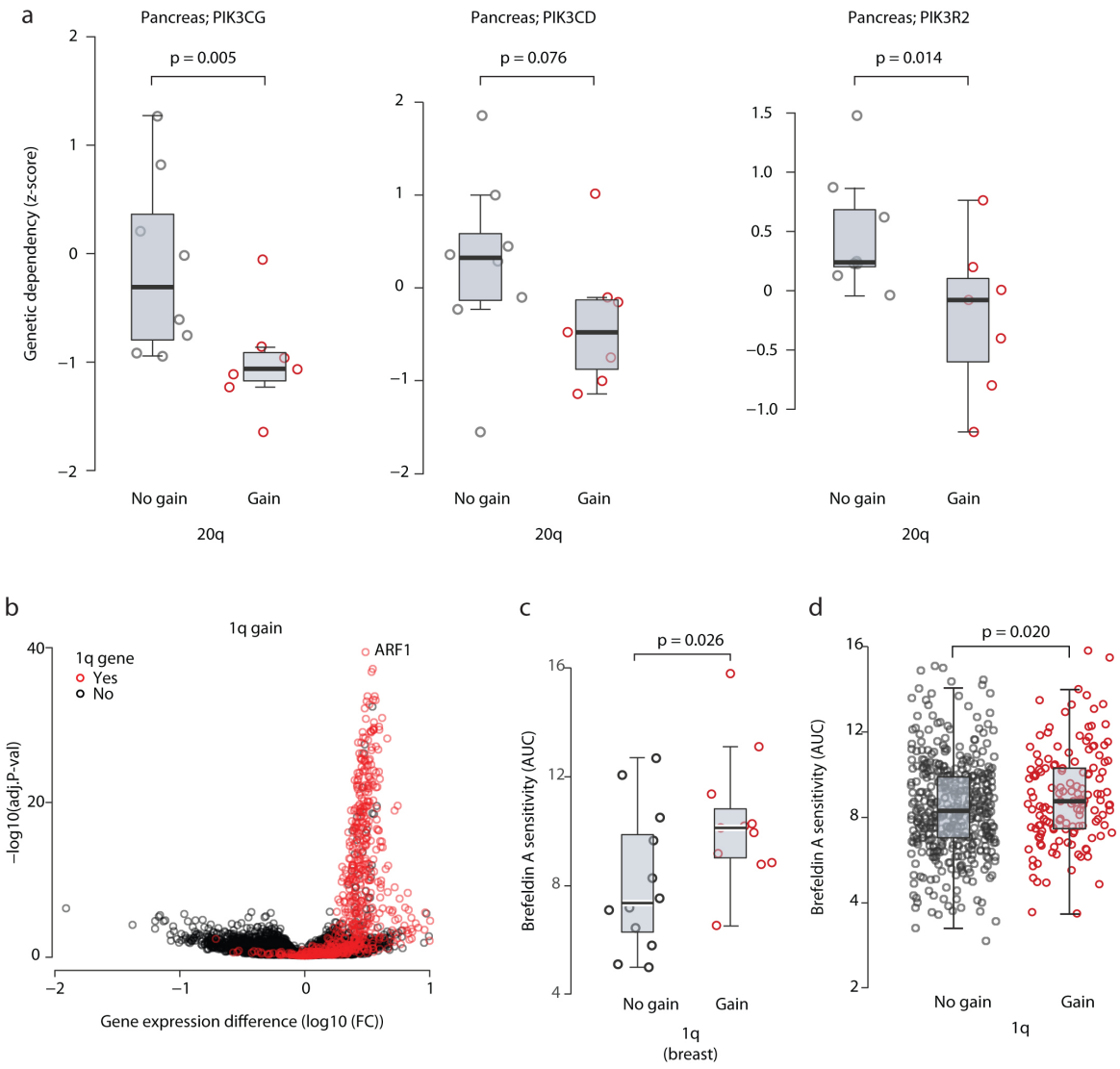
Figure S10



Supplementary Figure 10: Genomic instability of PDXs is comparable to that of cell lines and CLDXs

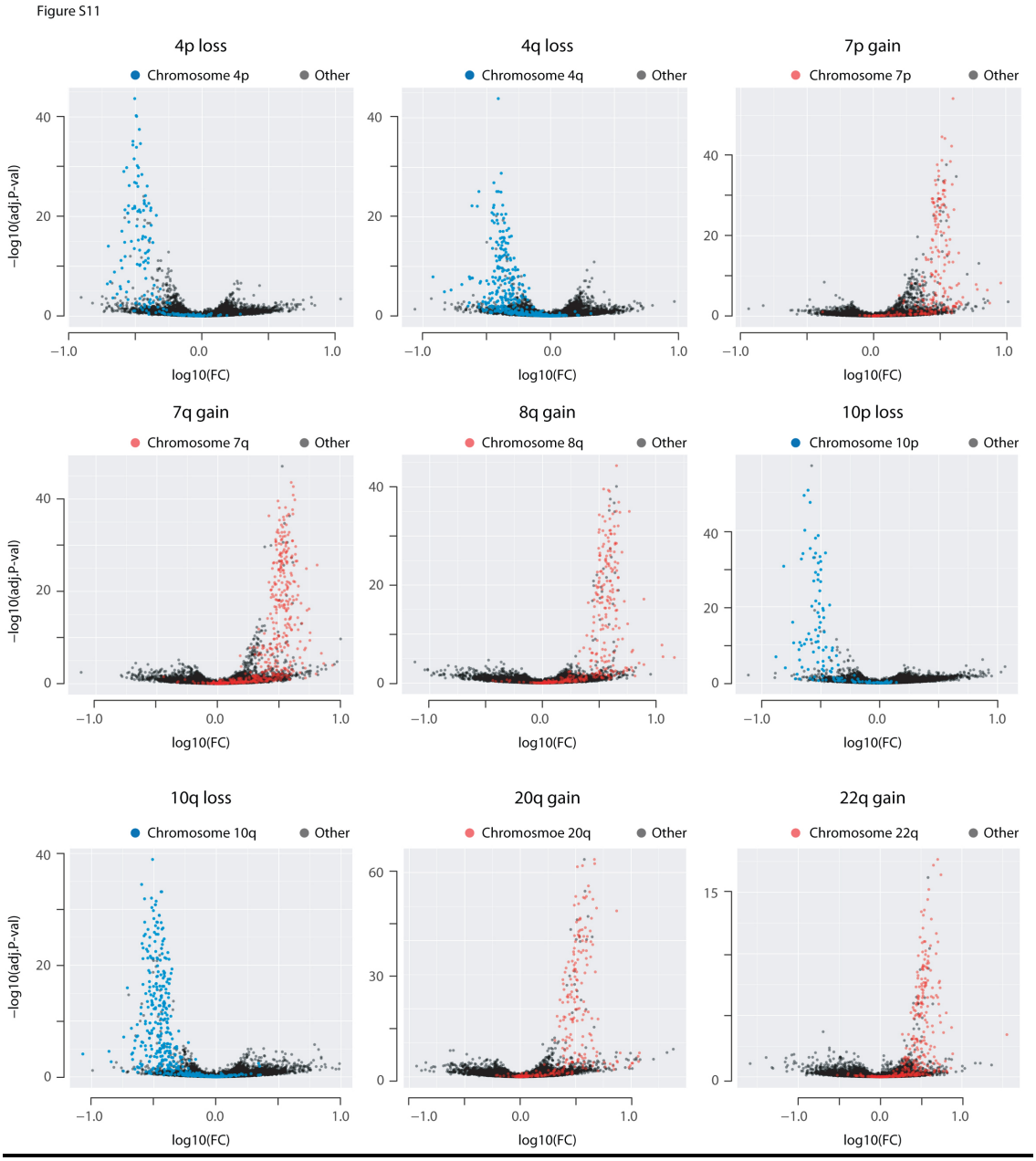
(**a**) Gradual evolution of CNA landscapes throughout passaging of newly-derived cell lines. Box plots present model-acquired CNA fraction as a function of *in vitro* passage number. (**b**) Similar rates of CNA acquisition in PDXs and in newly-derived cell lines. Dot plots present the distribution of model-acquired CNA fractions across three available cancer types. Cell lines used for this analysis are listed in **Supplementary Table 3**. (**c**) The CNA acquisition rate of CLDXs is associated with the numerical karyotypic complexity of the parental cell lines. Violin plots present the fraction of CNAs acquired by passage 10 as a function of numerical karyotypic complexity. P-values indicate significance from a Wilcoxon rank-sum test.

Figure S11



Supplementary Figure 11: Arm-level CNAs affect genetic dependencies and drug sensitivities

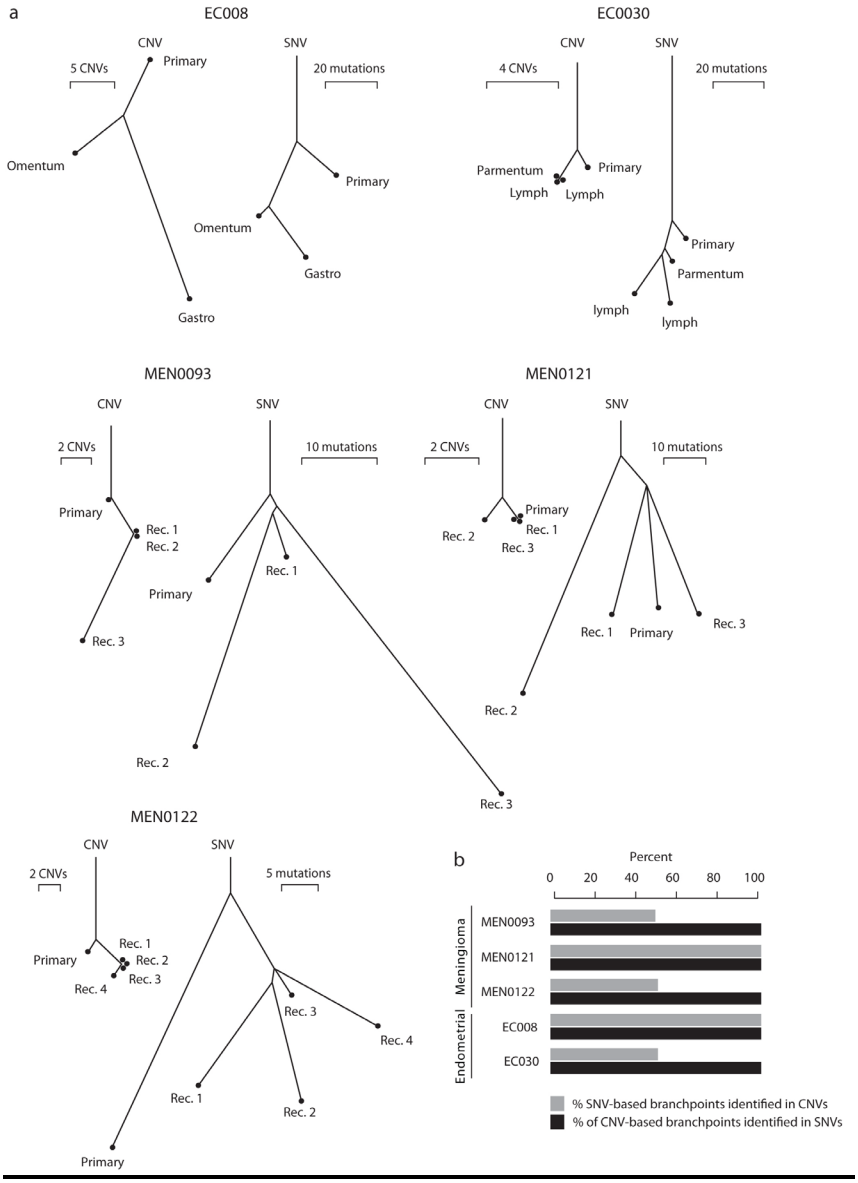
(**a**) Pancreatic cancer cell lines with chromosome 20q gain are more sensitive to RNAi-mediated knockdown of multiple *PI3K* genes. Box plots present the dependency scores to RNAi-mediated knockdown of the indicated genes. (**b**) Arm-level CNAs lead to significant gene expression changes in cell lines. Volcano plot shows the differential gene expression between cell lines with and without a gain of chromosome 1q. Genes that reside within chromosome 1q are highlighted in red. (**c**) Breast cancer cell lines with a chromosome 1q gain are less sensitive to the ARF1 inhibitor brefeldin A. (**d**) Cell lines with a chromosome 1q gain are less sensitive to the ARF1 inhibitor brefeldin A. Box plots present the area under the curve (AUC) values for cell lines with and without 1q gain. P-values indicate significance from a Wilcoxon rank-sum test.



Supplementary Figure 12: Arm-level CNAs are associated with extensive gene expression changes in cell lines

The expression of most genes that reside within arm-level CNAs changes significantly, in the expected direction of the aberration. Volcano plots present the differential gene expression analysis between cell lines with and without arm-level CNAs, for all the recurrent arm-level CNAs that tend to disappear during PDX passaging. Genes that reside within the affected arm are highlighted in red (for gains) or blue (for losses).

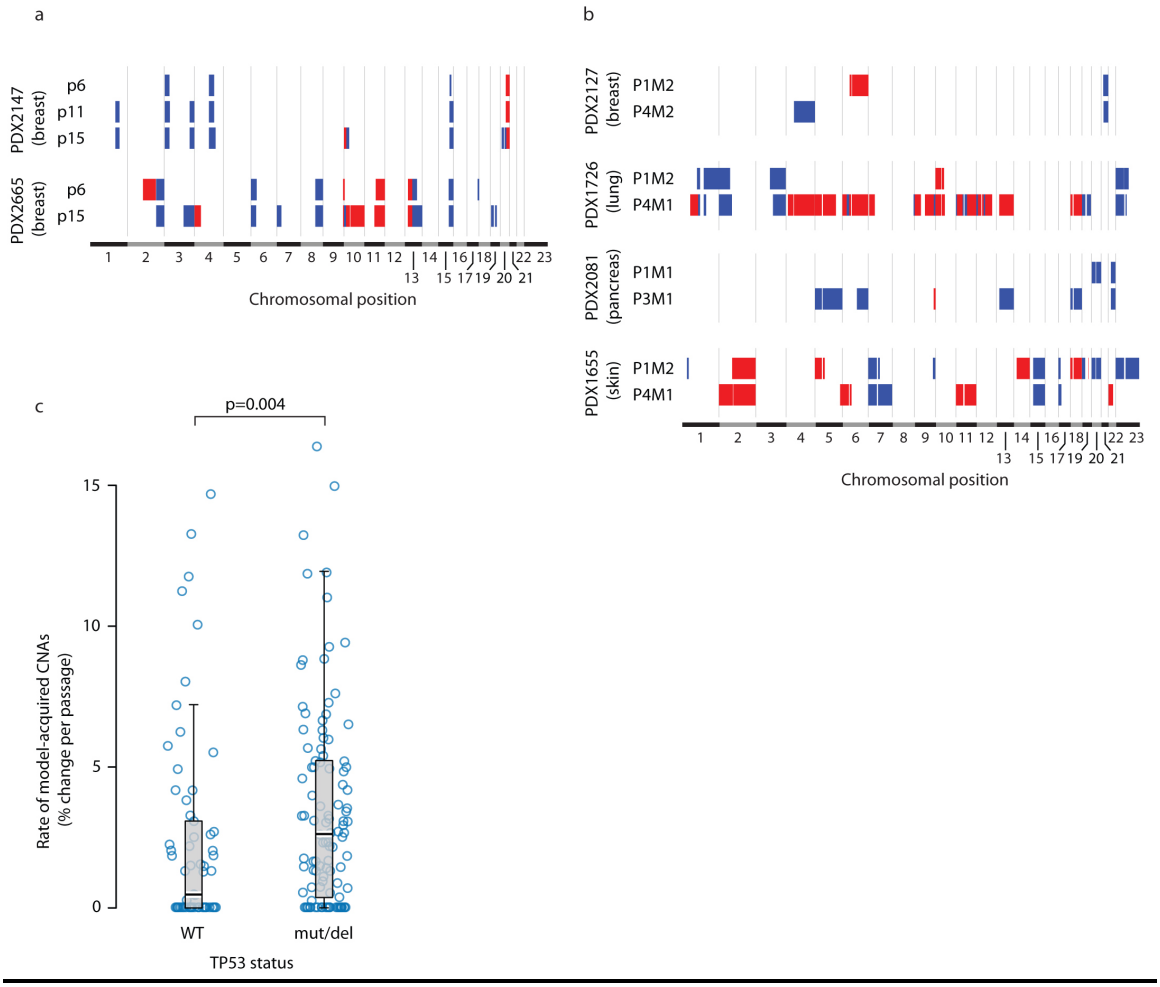
Figure S13



Supplementary Figure 13: CNA-based and mutation-based phylogenetic trees are highly concordant

Arm-level CNA-based phylogenetic trees are less informative than, but as accurate as, mutation-based trees. **(a)** Phylogenetic trees constructed from arm-level CNAs (left) or point mutations (right) data from five patients[2,3]. Branch lengths correspond to genetic distance. (**b**) Quantification of the sensitivity and specificity of the arm-level CNA-based trees, compared to the mutation-based trees. Bar plots present the percentage of branch points from mutation-based trees that were also identified in CNA-based trees (gray); and the percentage of branch points identified in CNA-based trees that were also identified in mutation-based trees (black).

Figure S14



Supplementary Figure 14: *De novo* CNAs may play a role in PDX CNA dynamics as well

(**a**) Model-acquired CNAs keep emerging at high *in vivo* passages. Plots present the model-acquired CNAs in multiple passages of two breast PDX models [4]. (**b**) Unique events also emerge in "sibling" PDXs, which were derived from the same primary tumor and propagated independently in mice. Plots present the model-acquired CNAs in pairs of passages from breast (PDX2127), lung (PDX1726), pancreas (PDX2081) and skin (PDX1655) PDX models [5]. Gains are shown in red, losses in blue. (**c**) PDXs with a mutant or deleted p53 present a significantly higher rate of CNA acquisition throughout passaging, compared to their WT counterparts. Box plots present the rate of model-acquired CNAs in PDX models without (PDX-WT; n=65) and with (TP53 mut/del; n=110) a TP53 perturbation. P-value indicates significance from a Wilcoxon rank-sum test.

Supplementary Tables

Study ID	Tumor type (by primary tissue)	# of PDX samples	# of PDX models	# of models with multiple time points	# of models with primary tumors available	Platform type	Platform	Reporting paper (PUBMED ID)	Analyses in which data were used
GSE14804	Brain cancer (GBM)	21	21	0	0	SNP array	Affymetrix Human Mapping 50K Xba240 SNP Array	19139420	Fig. 1a, 1e
GSE32530	Breast cancer	12	12	12	12	SNP array	Affymetrix Genome-Wide Human SNP 6.0 Array	22019887	Fig. 1a, 1e, Table S2
GSE41188	Prostate cancer	21	7	7	3	CGH array	Agilent-022060 SurePrint G3 Human CGH Microarray 4x180K (Feature Number version)	24356420	Fig. 1a, 1e, Table S2
MEXP-2576	Bone cancer (osteosarcoma)	27	15	15	8	CGH array	NMC Human Genomic Array 20K v1	21713766	Fig. 1a, 1e
EGAS00001000952	Breast cancer	79	15	15	15	SNP array	Affymetrix Genome-Wide Human SNP 6.0 Array	25470049	Fig. 1a, 1e, 2f, S5
		17				DNA sequencing	Illumina MiSeq; Illumina HiSeq 2000		Fig. 1a, 1e
GSE78806	Biliary tract	1	1	0	0	GE array	Affymetrix HG-U133 Plus 2.0 Array	26479923	Fig. 1a, 1e
	Bone cancer	10	6	4	0				Fig. 1a, 1d, 1e, 2a, 2b, 2c, 3a, S3, S5, S14
	Breast cancer	79	40	27	0				Fig. 1a, 1b, 1d, 1e, 2a, 2b, 2c, 2e, 3a, 3b, 3c, 4a, 4b, 4c, 6a, 6b, S1, S2, S3, S5, S7, S8, S9, S14
	Brain cancer	7	4	3	0				Fig. 1a, 1b, 1d, 1e, 2a, 2b, 2c, 3a, 3b, 3c, 3d, 4a, 4b, 4c, 5b, S2,S3, S5, S7, S8, S9, S10, S14
	Endometrium cancer	13	8	5	0				Fig. 1a, 1b, 1d, 1e, 2a, 2b, 2c, 3a, 3b, 3c, 3d, S1, S2,S3, S5, S7, S14
	Lymphoma	4	2	2	0				Fig. 1a, 1d, 1e, 2a, 2b, S3, S14
	Kidney cancer	41	22	13	0				Fig. 1a, 1b, 1d, 1e, 2a, 2b, 2c, 3a, 3b, 3c, 3d, S1, S2, S3, S5, S7, S10, S14
	Large Intestine (colorectal)	121	64	36	0				Fig. 1a, 1b, 1d, 1e, 2a, 2b, 2c, 3a, 3b, 3c, 4a, 4b, 5b, S1, S3, S5, S7, S8, S14
	Liver cancer	3	2	1	0				Fig. 1a, 1d, 1e, 2a, 2b, 2c, S3, S5, S14
	Non-small-cell lung cancer	99	50	35	0				Fig. 1a, 1b, 1d, 1e, 2a, 2b , 2e, 3a, 3b, 3c, 3d, 4a, 4b, 4c, 6a, 6b, S2, S3, S7, S8, S9, S14
	Meninges cancer	2	1	1	0				Fig. 1a, 1d, 1e, 2a, 2b, S3, S14
	Oesophagus cancer	15	10	5	0				Fig. 1a, 1d, 1e, 2a, 2b, 2c, 3a, 5b, S3, S5, S14
	Ovarian cancer	52	36	16	0				Fig. 1a, 1b, 1d, 1e, 2a, 2b, 2c, 3a, 3b, 3c, 4a, S2, S3, S5, S7, S14
	Pancreatic cancer	94	46	35	0				Fig. 1a, 1d, 1e, 2a, 2b, 2c, 2e, 3a, 3c, 4b, 5b, 6a, 6b, S1, S3, S5, S7, S8, S14
	Skin cancer	46	29	15	0				Fig. 1a, 1b, 1d, 1e, 2a, 2b, 2c, 2e, 3a, 3c, 4a, 6a, S1, S2, S3, S5, S7, S14
	Small intestine	2	2	0	0				Fig. 1a, 1e, 2b, S3, S14
	Soft-tissue cancer	62	41	19	0				Fig. 1a, 1b, 1d, 1e, 2a, 2b, 2c, 3a, 4a, S2, S3, S5, S14
	Stomach cancer	8	5	3	0				Fig. 1a, 1d, 1e, 2a, 2b, 3a, 3c, 3d, S3, S5, S7, S14
	Aerodigestive tumor	1	1	0	0				Fig. 1a, 1e, 2b, S3
	Endocrine tumor	1	1	0	0				Fig. 1a, 1e, 2b, S3

GSE14805 *	Brain cancer (GBM)	34	22	8	0	GE array	Affymetrix HT Human Genome U133A Array	19139420	Fig. 1a, 1b, 1c, 1e, 4a, 4b, 4b, S2, S3, S8, S9, S10
GSE32531 **	Breast cancer	15	11	11	11	GE array	Agilent-014850 Whole Human Genome Microarray 4x44K G4112F	22019887	Fig. 1a, 1b, 1e, 2a, 4a, 4b, 4c, 5b, S2, S3, S8, S9, Table S2
GSE41192 ***	Prostate cancer	21	7	5	0	GE array	Agilent-028004 SurePrint G3 Human GE 8x60K Microarray (Probe Name Version)	24356420	Fig. 1a, 1e, Table S2
GSE6465 / GSE72981	Liver cancer	38	7	6	7	GE array	Affymetrix HG-U133 Plus 2.0 Array	18490075	Fig. 1a, 1e
GSE66187	Prostate cancer	24	19	0	0	GE array	Agilent-016162 PEDB Whole Human Genome Microarray 4x44K	26071481	Fig. 1a, 1e
GSE28570	Breast cancer	15	5	0	0	GE array	Illumina HumanHT-12 V4.0 expression beadchip	21768359	Fig. 1a, 1b, 1e
GSE67312	Bladder cancer	5	5	5	5	GE array	Affymetrix Human Exon 1.0 ST Array	26041878	Fig. 1a, 1e
GSE51130	Soft-tissue cancer (rhabdomyosarcoma)	5	1	1	1	GE array	Affymetrix HG-U133 Plus 2.0 Array	24687871	Fig. 1a, 1b, 1e
GSE57633/ GSE57491	Acute lymphoblastic leukaemia	9	9	9	9	GE array	Illumina HumanWG-6 v3.0 expression beadchip	24885906	Fig. 1a, 1e
GSE46106	Breast cancer	44	25	15	0	GE array	Affymetrix HG-U133 Plus 2.0 Array	23737486	Fig. 1a, 1b, 1e, 4a, 4b, 4c, S2, S3, S8, S9
GSE55828	Liver cancer	9	9	9	9	GE array	[PrimeView] Affymetrix Human Gene Expression Array	26062443	Fig. 1a, 1e
GSE42975	Soft-tisseue cancer (liposarcoma)	14	3	3	3	GE array	Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (Probe Name version)	23416162	Fig. 1a, 1b, 1e, S2, S3
GSE28860	Breast cancer	11	3	3	2	GE array	Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (Probe Name version)	22446188	Fig. 1a, 1b, 1e, S2, S3
GSE15621	Acute lymphoblastic leukaemia	5	5	5	5	GE array	llumina HumanWG-6 v3.0 expression beadchip	20739953	Fig. 1a, 1e
GSE45153	Head and neck cancer	13	5	5	0	GE array	Affymetrix HG-U133 Plus 2.0 Array	23981300	Fig. 1a, 1b, 1e, S2, S3, S10
SRP028952	Large Intestine (colon)	10	4	4	4	RNA sequencing	Illumina HiSeq 2000	24278200	Fig. 1a, 1b, 1e, S2, S3

* Partial (20 models) overlap with the tumors described in GSE14804

** Partial (11 models) overlap with the tumors described in GSE32530

*** Partial (7 models) overlap with the tumors described in GSE41188

Supplementary Table 1: Summary of PDX datasets

A list of the datasets included in this study, together with their accession numbers, tumor types, the number of PDX models and samples included in them, the experimental platform used, and the Pubmed ID number of the original study that generated them.

Study	PDX model	SNP/CGH array		GE array		SNP/CGH array		GE array	
		Primary tumor to earliest passage				Early to late passage			
		Gains	Losses	Gains	Losses	Gains	Losses	Gains	Losses
GSE32530/1	HCI001	2p, 9p, Xp, Xq	-	2p, 9p, Xp, Xq	-	2p, 9q	9p, 10p, 14q, 19q, Xp, Xq	2q, 5q, 9q	9p, 10p, 14q, 19q, Xp, Xq
	HCI002	-	-	-	-	1q	-	1q	-
	HCI003	8q, 16p, 17q, 20p, 20q	1p, 4p, 4q, 8p	16p,17q,20p,20q	4p,4q,8p,13q	na	na	na	na
	HCI004	2q, 12q, 13q,21q	16q	2q, 12q, 13q,21q	16q	na	na	na	na
	HCI007	8q,17q,19q	-	8q,19q	-	na	na	na	na
	HCI008	-	-	-	-	na	na	na	na
	HCI009	7p, 10p,14q,20p,20q	19q	7p, 10p, 14q,20p,20q	-	na	na	na	na
	HCI10	9p	-	-	-	na	na	na	na
	HCI11	1q, 8q, 13q	1p, 8p, 9p, 11p,17p, 19q, 22q	8q,13q	1p, 9p, 11p, 17p, 19p	na	na	na	na
	HCI12	7p, 10q, 11q, 13q, 14q	3p, 4p, 10p, 11p, 12q, 22q, Xq	7p, 9q, 10q, 12p, 13q, 14q	3p, 4p, 10p, 11p, 12q, 22q	na	na	na	na
GSE41188/92	LTL418	-	-	-	-	2p, 4q	11q	2p, 4q	11q
	LTL412	-	1p	-	1p	-	-	-	-
	LTL311	na	na	na	na	-	12p	-	12p,19q
	LTL313B	na	na	na	na	-	-	-	14q
	LTL313H	na	na	na	na	13q,18q,Xq	-	13q,18q,Xq	-

Supplementary Table 2: Comparison of DNA- and RNA-based CNA profiles

A comparison of model-acquired CNAs inferred from DNA and RNA data from the same tumor samples.

Reporting paper (PUBMED ID)	Tumor type	# of primary tumor samples	# of advanced disease samples	Platform type	Platform	Accession ID
25979483	Colorectal cancer	5	17	CGH array	Agilent SurePrint G3 Human CGH Microarray 180 K	GSE58512
26109429	Colorectal cancer	19	19	DNA sequencing	Illumina HiSeq 2000	GSE50760
25979483	Colorectal cancer	11	11	DNA sequencing	Illumina HiSeq 2000	GSE53799
27472274	Colorectal cancer	5	17	SNP array	Affymetrix Genome-Wide Human SNP 6.0 Array	GSE63490
27348297	Endometrial cancer	23	31	DNA sequencing and SNP array	Illumina HiSeq 2000 & Affymetrix SNP 6.0 Array	phs001127.v1.p1 (dbGaP)
26618343	Brain cancer (glioma)	44	44	DNA sequencing and SNP array	Illumina HiSeq 2500 & Illumina HumanOmniExpress-24 v1.1 BeadChip	EGAS00001001588
27634761	Lung cancer (NSCLC)	6	16	DNA sequencing	Illumina HiSeq 2500	n.a.
26619122	Head and neck cancer	19	19	DNA sequencing	Illumina HiSeq 2500	phs0001007.v1.p1 (dbGaP)

Supplementary Table 3: Summary of advanced disease datasets

A list of the advanced disease datasets included in this study, together with their accession numbers, tumor types, the number of primary tumor and advanced disease samples, the experimental platform used, and the Pubmed ID number of the original study that generated them.

Reporting paper (PUBMED ID)	Cancer type	Tumor / Cell line ID	Passage_number	Accession_number
Tseng et al. (in preparation)	Pancreatic	AA01T	1	n.a.
			3	
			10	
			15	
			21	
		AA02T	0	
			6	
			10	
			20	
	Esophagus	AB002_ATT	1	
			5	
			10	
			20	
		AB002_BTT	1	
			9	
			10	
			15	
			19	
		AB018T	0	
			7	
			10	
			20	
	Metastatic Colon Cancer to Brain	BT584T	1	
			5	
			10	
			20	
	GBM	BT607T	0	
			5	
			10	
			20	
		BT632T	0	
			5	
			10	
			20	
	Anaplastic Thyroid	JL16T	0	
			5	
			10	
			20	
23981300	Head and neck	HOSC1	n.s.	GSE45153
			n.s.	

18037961	GBM	GS1	19	GSE8049
			28	
		GS2	10	
			13	
		GS3	12	
		GS3	23	
		GS4	11	
			12	
		GS5	9	
			12	
		GS6	10	
		GS7	11	
			25	
		GS8	8	
			18	
		GS9	6	
			13	
21294158	GBM	GS10	12	GSE23806
		GS11	11	
		GS12	11	
21668985	Kidney	50PC	n.s.	n.a.
		59RG	n.s.	
		60CC	n.s.	
		61FG	n.s.	
		66SML	n.s.	
		70LS	n.s.	
		73PG	n.s.	
		80MLa	n.s.	
		81BPG	n.s.	

Supplementary Table 4: Summary of newly-derived cell lines

A list of the newly-derived cell lines included in this study, together with their accession numbers, tumor types, passage numbers, and the Pubmed ID number of the original study that generated them.

Gene expression comparison		
Recurrent arm-level CNA	**enrichment_stat**	**p_val (one-sided)**
chr1q_gain	36.84934863	9.01E-47
chr4p_loss	-11.03874633	8.60E-51
chr4q_loss	-20.05261324	5.08E-36
chr7p_gain	18.678073	4.00E-65
chr7q_gain	33.94634954	3.54E-59
chr8q_gain	23.68555794	1.07E-58
chr10p_loss	-8.341567128	6.29E-49
chr10q_loss	-22.16795676	1.86E-50
chr20q_gain	21.82304448	3.07E-71
chr22q_gain	7.672067713	1.40E-18

Genetic dependency (RNAi) comparison		
Recurrent arm-level CNA	**enrichment_stat**	**p_val (one-sided)**
chr1q_gain	2.658194034	3.25E-05
chr4p_loss	-1.045604187	2.05E-05
chr4q_loss	-1.759585902	0.002660013
chr7p_gain	1.759644212	2.62E-05
chr7q_gain	4.365535663	3.02E-08
chr8q_gain	2.767599408	1.08E-06
chr10p_loss	-1.021092902	0.000237148
chr10q_loss	-1.588799628	0.002168448
chr20q_gain	2.552646879	3.74E-06
chr22q_gain	1.789044702	0.000281875

Supplementary Table 5: Comparisons of gene expression and genetic dependencies in cell lines with arm-level CNAs and cell lines without them

A lineage-controlled comparison of cell lines with and without recurrent arm-level CNAs, across a panel of 936 cell lines (for gene expression) or 446 cell lines (for gene dependencies). In the gene expression comparison, genes that are differentially expressed between cell lines with the CNA and cell lines without it were subjected to gene set enrichment analysis (GSEA), and found to be significantly enriched for genes that reside within that chromosome arm. In the genetic dependency comparison, genes that are differentially depleted in a pooled genome-wide RNAi screen between cell lines with the CNA and cell lines without it were subjected to GSEA, and found to be significantly enriched for genes that reside within that chromosome arm. Presented are the effect magnitudes (enrichment statistics) and the p-values of the enrichments. Note that the enrichment statistics are positive in the expression comparison and negative in the dependency comparison. Therefore, a gain of an arm results (on average) in increased expression and reduced dependency on the genes within the arm, and a loss of an arm results (on average) in decreased expression and increased dependency on the genes within the arm.

cpd_name	broad_cpd_id	Associated CNA	target	logFC	p_left	p_right	q_left	q_right
VX-680	BRD-K59369769	1q_amp	AURKA;AURKB;AURKC	-3.24	0.00	1.00	0.05	0.95
BRD-K01737880	BRD-K01737880	1q_amp	na	-2.94	0.00	1.00	0.05	0.95
navitoclax:pluripotin (1:1 mol/mol)	BRD-M07227555	1q_amp	BCL2;BCL2L1;BCL2L2;MAPK1;RASAL1	0.71	1.00	0.00	0.98	0.02
decitabine:navitoclax (2:1 mol/mol)	BRD-M43399311	1q_amp	DNMT1;BCL2;BCL2L1;BCL2L2	0.69	1.00	0.00	0.78	0.22
brefeldin A	BRD-A31107743	1q_amp	ARF1	0.67	1.00	0.00	0.82	0.18
navitoclax	BRD-K82746043	1q_amp	BCL2;BCL2L1;BCL2L2	0.61	1.00	0.00	0.76	0.24
UNC0638:navitoclax (1:1 mol/mol)	BRD-M97749165	1q_amp	EHMT1;EHMT2;BCL2;BCL2L1;BCL2L2	0.58	1.00	0.00	0.76	0.24
BRD-A02303741:navitoclax (2:1 mol/mol)	BRD-M76874421	1q_amp	DOT1L;BCL2;BCL2L1;BCL2L2	0.56	1.00	0.00	0.78	0.22
pluripotin	BRD-K98538768	1q_amp	MAPK1;RASAL1	0.54	1.00	0.00	0.78	0.22
navitoclax:piperlongumine (1:1 mol/mol)	BRD-M43334969	1q_amp	BCL2;BCL2L1;BCL2L2	0.53	0.99	0.01	0.76	0.24
bardoxolone methyl	BRD-K59437938	1q_amp	na	-0.38	0.01	0.99	0.24	0.76
bendamustine	BRD-K17068645	1q_amp	na	-0.28	0.00	1.00	0.24	0.76
clofarabine	BRD-K34022604	4q_del	na	0.97	1.00	0.00	0.99	0.01
nutlin-3	BRD-A12230535	4q_del	MDM2	0.43	1.00	0.00	0.93	0.07
carboplatin:etoposide (40:17 mol/mol)	BRD-M63575423	4q_del	TOP2A	0.40	1.00	0.00	0.83	0.17
TGX-221	BRD-A41692738	4q_del	PIK3CB	0.29	1.00	0.00	0.83	0.17
brefeldin A	BRD-A31107743	10q_del	ARF1	-0.83	0.00	1.00	0.06	0.94
BRD-A05715709	BRD-A05715709	10q_del	IDH1	0.41	1.00	0.00	0.91	0.09
BRD-K01737880	BRD-K01737880	20q_amp	na	0.97	0.95	0.05	0.76	0.24
SB-743921	BRD-K62358710	20q_amp	KIF11	0.84	1.00	0.00	0.96	0.04
RITA	BRD-K00317371	20q_amp	MDM2;TP53	0.77	1.00	0.00	0.99	0.01
omacetaxine mepesuccinate	BRD-K76674262	20q_amp	na	0.74	1.00	0.00	0.97	0.03
teniposide	BRD-A35588707	20q_amp	TOP2A;TOP2B	0.74	1.00	0.00	0.97	0.03
barasertib	BRD-K63923597	20q_amp	AURKB	0.71	1.00	0.00	0.97	0.03
mitomycin	BRD-K59670716	20q_amp	na	0.71	1.00	0.00	0.99	0.01
rigosertib	BRD-K55187425	20q_amp	PIK3CA;PIK3CB;PLK1	0.63	1.00	0.00	0.96	0.04
tanespimycin:gemcitabine (1:1 mol/mol)	BRD-M41695494	20q_amp	HSP90AA1;CMPK1;RRM1;TYMS	0.62	1.00	0.00	0.96	0.04
doxorubicin	BRD-K92093830	20q_amp	TOP2A	0.61	1.00	0.00	0.99	0.01
vincristine	BRD-K82109576	20q_amp	na	0.61	0.99	0.01	0.93	0.07
paclitaxel	BRD-A28746609	20q_amp	na	0.58	0.98	0.02	0.86	0.14
indisulam	BRD-K17610631	20q_amp	CA9	0.57	1.00	0.00	1.00	0.00
docetaxel:tanespimycin (2:1 mol/mol)	BRD-M60574774	20q_amp	HSP90AA1	0.57	0.99	0.01	0.91	0.09
GSK461364	BRD-K92428232	20q_amp	PLK1	0.57	0.99	0.01	0.91	0.09
tanespimycin	BRD-K81473043	20q_amp	HSP90AA1	0.55	1.00	0.00	0.96	0.04
nutlin-3	BRD-A12230535	20q_amp	MDM2	0.54	1.00	0.00	1.00	0.00

navitoclax:gemcitabine (1:1 mol/mol)	BRD-M68535767	20q_amp	BCL2;BCL2L1;BCL2L2;CMPK1;RRM1;TYMS	0.54	0.99	0.01	0.92	0.08
parbendazole	BRD-K02407574	20q_amp	na	0.54	1.00	0.00	0.95	0.05
PF-3758309	BRD-K37764012	20q_amp	PAK4	0.54	0.97	0.03	0.81	0.19
clofarabine	BRD-K34022604	20q_amp	na	0.51	0.99	0.01	0.91	0.09
KX2-391	BRD-K29968218	20q_amp	SRC	0.48	0.99	0.01	0.88	0.12
bleomycin A2	BRD-A42083487	20q_amp	na	0.48	1.00	0.00	0.93	0.07
Ki8751	BRD-K47150025	20q_amp	KDR;KIT;PDGFRA	0.48	1.00	0.00	0.96	0.04
BI-2536	BRD-K64890080	20q_amp	PLK1	0.46	0.98	0.02	0.85	0.15
triazolothiadiazine	BRD-K05402890	20q_amp	PDE4A;PDE4B;PDE4D	0.45	0.99	0.01	0.93	0.07
doxorubicin:navitoclax (2:1 mol/mol)	BRD-M01915239	20q_amp	TOP2A;BCL2;BCL2L1;BCL2L2	0.45	1.00	0.00	0.96	0.04
tanespimycin:bortezomib (250:1 mol/mol)	BRD-M09675615	20q_amp	HSP90AA1;PSMB1;PSMB2;PSMB5;PSMD1;PSMD2	0.45	1.00	0.00	0.95	0.05
etoposide	BRD-K37798499	20q_amp	TOP2A	0.44	1.00	0.00	0.97	0.03
nakiterpiosin	BRD-K38264551	20q_amp	na	0.44	1.00	0.00	0.97	0.03
BRD-K70511574	BRD-K62825658	20q_amp	PLK1	0.44	1.00	0.00	0.97	0.03
carboplatin:etoposide (40:17 mol/mol)	BRD-M63575423	20q_amp	TOP2A	0.43	1.00	0.00	0.98	0.02
PHA-793887	BRD-K64800655	20q_amp	CDK1;CDK2;CDK4;CDK5;CDK7;CDK9	0.42	1.00	0.00	0.95	0.05
ML210	BRD-K01877528	20q_amp	na	-0.42	0.05	0.95	0.24	0.76
birinapant	BRD-K18589165	20q_amp	DIABLO;XIAP	0.41	0.98	0.02	0.85	0.15
CD-437	BRD-K28907958	20q_amp	RARG	0.41	1.00	0.00	0.98	0.02
cytarabine hydrochloride	BRD-K33106058	20q_amp	na	0.41	0.98	0.02	0.86	0.14
sirolimus:bortezomib (250:1 mol/mol)	BRD-M02488208	20q_amp	MTOR;PSMB1;PSMB2;PSMB5;PSMD1;PSMD2	0.40	0.98	0.02	0.86	0.14
FQI-2	BRD-K35498412	20q_amp	na	0.40	1.00	0.00	0.97	0.03
narciclasine	BRD-K06792661	20q_amp	RHOA	0.40	0.99	0.01	0.91	0.09
BRD-K44224150	BRD-K44224150	20q_amp	na	0.40	0.98	0.02	0.85	0.15
BRD9876	BRD-K89329876	20q_amp	na	0.39	0.99	0.01	0.93	0.07
alisertib:navitoclax (2:1 mol/mol)	BRD-M07476339	20q_amp	AURKA;AURKB;BCL2;BCL2L1;BCL2L2	0.38	0.97	0.03	0.80	0.20
SNX-2112	BRD-K71281111	20q_amp	HSP90AA1;HSP90B1	0.36	0.97	0.03	0.81	0.19
selumetinib:tretinoin (2:1 mol/mol)	BRD-M45185124	20q_amp	MAP2K1;MAP2K2;RARA;RARB;RARG	0.35	0.95	0.05	0.76	0.24
ceranib-2	BRD-K31514534	20q_amp	ACER1;ACER2;ACER3;ASAH1;ASAH2;ASAH2B	0.34	1.00	0.00	0.94	0.06
ciclosporin	BRD-A38030642	20q_amp	PPID	0.34	1.00	0.00	0.95	0.05
ELCPK	BRD-K44847641	20q_amp	na	0.33	0.99	0.01	0.90	0.10
MK-1775	BRD-K54256913	20q_amp	WEE1	0.32	0.99	0.01	0.89	0.11
topotecan	BRD-K55696337	20q_amp	TOP1	0.32	0.96	0.04	0.79	0.21
oxaliplatin	BRD-M14820059	20q_amp	na	0.32	1.00	0.00	0.96	0.04
quizartinib	BRD-K93918653	20q_amp	FLT3	0.32	0.99	0.01	0.89	0.11
cucurbitacin I	BRD-A28105619	20q_amp	na	0.31	0.96	0.04	0.77	0.23

AT-406	BRD-K02834582	20q_amp	XIAP	0.31	0.98	0.02	0.85	0.15
marinopyrrole A	BRD-K57261999	20q_amp	MCL1	0.31	1.00	0.00	0.93	0.07
TPCA-1	BRD-K51575138	20q_amp	IKBKB	0.30	1.00	0.00	0.96	0.04
AM-580	BRD-K06854232	20q_amp	RARA	0.30	0.97	0.03	0.81	0.19
BRD-K34222889	BRD-K34222889	20q_amp	na	0.30	1.00	0.00	0.95	0.05
avrainvillamide	BRD-K10466330	20q_amp	NPM1	0.30	1.00	0.00	0.93	0.07
chlorambucil	BRD-K29458283	20q_amp	na	0.29	0.99	0.01	0.93	0.07
BYL-719	BRD-K54997624	20q_amp	PIK3CA	0.29	0.96	0.04	0.77	0.23
cimetidine	BRD-K34157611	20q_amp	HRH2	0.29	0.99	0.01	0.88	0.12
ML312	BRD-K70809902	20q_amp	SCARB1	0.28	0.95	0.05	0.76	0.24
FSC231	BRD-K24784241	20q_amp	PICK1	0.28	0.99	0.01	0.89	0.11
AZD1480	BRD-K65928735	20q_amp	JAK1;JAK2	0.28	0.97	0.03	0.80	0.20
ML311	BRD-A40802033	20q_amp	MCL1	0.28	1.00	0.00	0.93	0.07
BRD-K66453893	BRD-K66453893	20q_amp	na	0.28	1.00	0.00	0.97	0.03
3-Cl-AHPC	BRD-A09890259	20q_amp	NR0B2	0.28	0.98	0.02	0.86	0.14
HBX-41108	BRD-K50501969	20q_amp	USP7	0.27	0.98	0.02	0.86	0.14
YK 4-279	BRD-A62182663	20q_amp	DHX9;ERG;ETV1	0.27	1.00	0.00	0.96	0.04
SB-225002	BRD-K61323504	20q_amp	CXCR2	0.27	1.00	0.00	0.94	0.06
CHM-1	BRD-K94145482	20q_amp	na	0.26	0.99	0.01	0.90	0.10
FQI-1	BRD-A91658086	20q_amp	na	0.26	0.97	0.03	0.81	0.19
KU-0063794	BRD-K67566344	20q_amp	MTOR	0.26	0.97	0.03	0.81	0.19
foretinib	BRD-K03449891	20q_amp	KDR;MET	0.26	0.95	0.05	0.76	0.24
BRD-K30019337	BRD-K30019337	20q_amp	na	0.26	0.96	0.04	0.79	0.21
LY-2183240	BRD-K37865504	20q_amp	FAAH	0.25	1.00	0.00	0.96	0.04
JQ-1:carboplatin (1:1 mol/mol)	BRD-M11510342	20q_amp	BRDT	0.25	0.96	0.04	0.79	0.21
XL765	BRD-K75308783	20q_amp	MTOR;PIK3CA;PIK3CB;PIK3CD;PIK3CG;PRKDC	0.25	1.00	0.00	0.97	0.03
crizotinib	BRD-K78431006	20q_amp	ALK;MET	0.25	0.99	0.01	0.92	0.08
lenvatinib	BRD-K39974922	20q_amp	FLT1;FLT3;KDR;KIT;PDGFRA;PDGFRB	0.25	0.96	0.04	0.80	0.20
methylstat	BRD-K52560704	20q_amp	KDM3A;KDM4A;KDM4B;KDM4C;KDM4D	0.25	0.98	0.02	0.87	0.13
serdemetan	BRD-K60219430	20q_amp	MDM2	0.24	0.99	0.01	0.91	0.09
AZ-3146	BRD-K59146805	20q_amp	TTK	0.24	0.98	0.02	0.85	0.15
R428	BRD-K14870255	20q_amp	AXL	0.23	0.97	0.03	0.81	0.19
gossypol	BRD-K19295594	20q_amp	BCL2;BCL2L1;LDHA;LDHB;LDHC	0.22	1.00	0.00	0.94	0.06
phloretin	BRD-K15563106	20q_amp	SLC5A1	0.22	1.00	0.00	0.96	0.04
ML320	BRD-K81491172	20q_amp	GSK3B	-0.22	0.05	0.95	0.24	0.76
palmostatin B	BRD-K42137908	20q_amp	LYPLA1	0.22	0.97	0.03	0.80	0.20

pitstop2	BRD-K90570971	20q_amp	CLTA;CLTB;CLTC;CLTCL1	0.21	0.97	0.03	0.81	0.19
NSC632839	BRD-K37392901	20q_amp	USP13;USP5	0.21	0.99	0.01	0.91	0.09
PLX-4720	BRD-K16478699	20q_amp	BRAF	-0.20	0.03	0.97	0.18	0.82
PL-DI	BRD-K83336168	20q_amp	na	0.20	0.98	0.02	0.85	0.15
BMS-270394	BRD-K04905989	20q_amp	RARG	0.20	0.99	0.01	0.92	0.08
KHS101	BRD-K78978711	20q_amp	TACC3	0.20	0.99	0.01	0.91	0.09
PIK-93	BRD-K29395450	20q_amp	PIK3CG	0.19	0.97	0.03	0.80	0.20
ML334 diastereomer	BRD-K93367411	20q_amp	KEAP1;NFE2L2	-0.19	0.03	0.97	0.20	0.80
prochlorperazine	BRD-K19352500	20q_amp	DRD2	0.18	1.00	0.00	0.94	0.06
isoevodiamine	BRD-A68631409	20q_amp	na	0.18	0.96	0.04	0.79	0.21
RO4929097	BRD-K22024824	20q_amp	APH1A;NCSTN;PSEN1;PSENEN	0.17	0.95	0.05	0.76	0.24
Compound 1541A	BRD-K30064966	20q_amp	CASP3;CASP6;CASP7	0.17	0.98	0.02	0.85	0.15
BRD-K50799972	BRD-K50799972	20q_amp	na	0.16	0.97	0.03	0.80	0.20
PYR-41	BRD-K60750172	20q_amp	UBA1	0.16	0.99	0.01	0.92	0.08
IU1	BRD-K45841694	20q_amp	USP14	0.16	0.99	0.01	0.89	0.11
GW-405833	BRD-K10705233	20q_amp	CNR2	0.15	0.99	0.01	0.89	0.11
NVP-BSK805	BRD-K02017404	20q_amp	JAK2	0.14	0.95	0.05	0.77	0.23
piperlongumine:MST-312 (1:1 mol/mol)	BRD-M37545453	20q_amp	TERT	0.14	0.96	0.04	0.79	0.21
JW-480	BRD-K99655327	20q_amp	NCEH1	0.14	0.98	0.02	0.86	0.14
PDMP	BRD-K05653692	20q_amp	UGCG	0.12	0.98	0.02	0.85	0.15
StemRegenin 1	BRD-K36739687	20q_amp	AHR	0.12	0.96	0.04	0.79	0.21
UNC0321	BRD-K74236984	20q_amp	EHMT2	0.11	0.97	0.03	0.81	0.19
ML029	BRD-K59962020	20q_amp	na	0.09	0.97	0.03	0.80	0.20

Supplementary Table 6: Comparison of drug response in cell lines with arm-level CNAs and cell lines without them

A lineage-controlled comparison of cell lines with and without recurrent arm-level CNAs, across a panel of 804 cell lines and 545 compounds. Compounds that are differentially active/inactive towards cell lines with CNAs vs. cell lines without them are listed (one-tailed $p<0.05$; one-tailed $q<0.25$). Presented are the compound names, the associated CNAs, the known compound targets, the average response differences (log fold-change), and the statistical significance of these differences.

Legends to Supplementary Datasets

Supplementary Dataset 1: CNA profiles of PDX samples

PDX CNA profiles generated in this study from gene expression data. The first tab provides a full description of the samples. The second tab provides a segmental aberration matrix, in a format readily visualized by the Integrative Genomics Viewer (IGV; https://www.broadinstitute.org/igv/).

Supplementary Dataset 2: Model-acquired CNAs in PDX samples

PDX model-acquired CNAs identified in this study from gene expression data. The first tab provides a full description of the samples. The second tab provides a segmental aberration matrix, in a format readily visualized by the Integrative Genomics Viewer (IGV; https://www.broadinstitute.org/igv/).

Supplementary Dataset 3: CNA profiles of CLDX samples

CLDX CNA profiles generated in this study from gene expression data. The first tab provides a full description of the samples. The second tab provides a segmental aberration matrix, in a format readily visualized by the Integrative Genomics Viewer (IGV; https://www.broadinstitute.org/igv/).

Supplementary Dataset 4: Model-acquired CNAs in CLDX samples

CLDX model-acquired CNAs identified in this study from gene expression data. The first tab provides a full description of the samples. The second tab provides a segmental aberration matrix, in a format readily visualized by the Integrative Genomics Viewer (IGV; https://www.broadinstitute.org/igv/).

Supplementary Note

1. Supplementary Introduction: Previous evidence for genomic instability of PDXs

Hints that PDXs may be more genomically unstable than assumed have begun to emerge, with a recent study showing that the clonal composition of breast cancer PDXs evolves during serial passaging *in vivo*[1]. Another study recently extended this analysis to additional breast cancer PDXs, showing that while there was overall similarity of PDX models to their tumors of origin, the clonal composition of the tumors could change dramatically throughout PDX derivation and propagation[6]. Importantly, both studies presented a deep characterization of PDXs from a total of 83 models of a single tissue type (breast), with no systematic assessment of the rate, prevalence or recurrence patterns of genomic changes during *in vivo* passaging of PDXs. Additionally, whether the observed clonal dynamics have any functional importance remains an open question.

2. Supplementary Introduction: Somatic copy number alterations (CNAs) in cancer

Somatic copy number alterations (CNAs) are detectable in the vast majority of cancers[7,8], and therefore represent a powerful strategy to track the clonal evolution of tumors. Moreover, CNAs are often drivers of tumorigenesis and have been associated with drug response and prognosis in human patients[9-14]. Despite the importance of CNAs in cancer, they are rarely characterized in PDX models, and comprehensive analysis of CNA dynamics during *in vivo* PDX passaging has yet to be reported[6,15-17].

3. Supplementary Results: Comparison of DNA- and RNA-derived copy number profiles

To validate the accuracy of inferred CNAs, we analyzed PDXs from which both gene expression and SNP array data (a more direct measurement of DNA copy number) were available. Because in most cases DNA and RNA were obtained from different PDX passages, we focused on the 59 PDX models that had stable CNAs over time. The DNA- and RNA-derived profiles were highly concordant both when comparing the proportion of the genome affected by CNAs (Pearson's r = 0.86) and when comparing the concordance of affected genes (median concordance = 0.82) (**Supplementary Fig. 1**). Moreover, for 15 breast and prostate PDXs, we could directly compare the changes that occurred during their engraftment and/or passaging (hereinafter called 'model-acquired CNAs'), using DNA and RNA data from the same samples. These DNA- and RNA-derived profiles were highly concordant (Pearson's r = 0.95; median concordance = 0.91). These results thus confirmed that gene expression accurately identifies model-acquired CNAs (**Supplementary Table 2**).

4. Supplementary Results: Changes in the allelic fraction of point mutations throughout PDX passaging

To assess whether model-acquired clonal evolution affected genes known to play important roles in cancer, we analyzed the whole-genome sequencing data of 13 matched breast cancer models[1]. We first focused on changes in the allelic fraction (AF) of mutations when comparing primary tumors to their PDXs. We found a median of 64 mutations with a substantial AF shift ($|\Delta AF|>0.2$) (**Supplementary Fig. 4a**). Importantly, a median of 7 of these mutations were non-synonymous missense or nonsense coding mutations, many involving oncogenes and tumor suppressor genes (**Supplementary Fig. 4a-b**). For example, the AF of a *TP53BP1* missense mutation increased from 0.188 to 0.562, and the AF of a *PIK3CA* missense mutation increased from 0.339 to 0.740 during the PDX evolution of tumors SA531 and SA536, respectively (**Supplementary Fig. 4b**). Similarly, substantial coding mutation AF shifts were observed when comparing early passage (P1 or P2) to late passage (P5) PDXs (**Supplementary Fig. 4c-d**). For example, the Rho-associated kinase *ROCK1* has been previously shown to promote tumor cell invasion and metastasis[18,19]. A missense *ROCK1* mutation was not detected at all in a PDX model at passage 2, but was detected at a high AF (AF=0.46) at passage 5 of the same model (**Supplementary Fig. 4d**). These results indicate that clonal dynamics quickly alter the prevalence of functional mutations in cancer genes, with potentially important functional consequences.

5. Supplementary Results: Deviation of melanoma from the observed correlation between PDX instability and primary tumor heterogeneity

Melanoma is the only tumor type that deviated from the tight correlation between intra-tumor heterogeneity and model-acquired genomic instability (**Supplementary Fig. 7a**). A potential explanation for this discrepancy is that the extent of genetic heterogeneity in melanoma may be overestimated compared to other cancer types due to the unusually high mutation load of this tumor.

6. Supplementary Results: Association between arm-level CNAs and cell line gene expression, genetic dependencies and drug response

To further assess the potential functional relevance of model-acquired chromosomal changes we turned to the Cancer Cell Line Encyclopedia (CCLE) project and its associated datasets of genomic features, genetic dependencies and drug response[20-22]. Interestingly, arm-level CNAs significantly affected the genetic dependencies for genes that reside within them: cell lines with an arm-level gain were less sensitive on average to RNAi-mediated depletion of genes within the gained arm, whereas cell lines with an arm-level loss were hyper-sensitive to such perturbations

(**Supplementary Table 5**). Next, we queried the drug response data for the CCLE cell lines and asked whether there were differential drug responses associated with the twelve hallmark arm-level CNAs that tend to disappear in PDXs. Out of 545 drugs tested, each arm-level CNA had on average 13 (range, 0-112) differentially active drugs associated with it. That is, the responses to these drugs were significantly associated with the copy number status of that arm ($p<0.05$, $q<0.25$; **Supplementary Table 6**). For example, the gene *ARF1* is located on chromosome 1q, and is the most significantly over-expressed gene in cell lines with a 1q gain (p=9.8E-45) (**Supplementary Fig. 11b**); this rendered cell lines with a gain of this chromosome arm more resistant to the ARF1 inhibitor brefeldin A (p=0.01), and this strong association remained significant when breast cancer cell lines were considered alone (p=0.026) (**Supplementary Fig. 11c-d**).

7. Supplementary Discussion: Comparison of our findings to previous studies

Our findings may be surprising in light of previous works that emphasized the relative stability of PDXs throughout their propagation. However, most prior studies either focused on comparisons at the cohort level[5,15,16]; inferred stability from the increased similarity between PDXs derived from the same patient compared to PDXs from different patients[5,6,17]; or used only a handful of markers to assess stability[23] (reviewed in [24]). Our large-scale, genome-wide, pairwise comparison of early vs. late PDXs exposed previously-underappreciated CNA dynamics, similar to the SNV dynamics recently seen in breast cancer PDXs[1].

8. Supplementary Discussion: The relevance of our CNA-based analysis to other types of genetic alterations

While we focused our analysis on CNAs, we confirmed the accuracy of CNA-based phylogenetic trees (**Supplementary Fig. 13**) using matched point mutation and CNA data from primary tumors and their metastases[2,3]. It seems very likely that PDX models (as all other models) also acquire other types of aberrations, including point mutations, small insertions and deletions, translocations, and epigenetic modifications. Large-scale datasets do not exist at present to experimentally confirm these other forms of selection, but an analysis of a small cohort of breast cancer PDXs[7] confirmed that the allelic fraction of cancer genes can drastically change within the course of very few passages.

9. Supplementary Discussion: Combination of pre-existing and *de novo* occurring CNAs in PDXs

Our study strongly suggests that clonal dynamics play a major role in model-acquired CNAs. In particular, the acquisition of identical events in "sibling" PDXs, and the detection of LOH "reversion" throughout PDX passaging, strongly point towards expansion of pre-existing subclones. However, our analysis suggests that *de novo* events also occur. First, model-acquired CNAs are not limited to the early passages and keep emerging, albeit at a lower rate, even at high passages (**Supplementary Fig. 14a** and **Supplementary Data 2**). Second, although "sibling" PDXs exhibit high similarity of model-acquired CNAs, most of them also acquire unique events (**Supplementary Figure 14b**). Third, we found that PDXs with a mutant or deleted p53 present a significantly higher rate of CNA acquisition throughout passaging, compared to their WT counterparts (**Supplementary Figure 14c**). All three of these findings, however, could also be potentially explained by extensive pre-existing heterogeneity.

10. Supplementary Discussion: Distinct selection pressures in PDXs and in patients

Unique, context-dependent selection pressures shape tumor evolution, giving rise to recurrent cancer type-specific CNAs[7]. While genetic drift or "founder effects" may underlie some of the changes observed in PDXs, we provide evidence that selection plays an important role. The increase in proliferation signatures and decrease in cell death signature throughout passaging; the independent emergence of the same events in "sibling" PDX models; and the tendency of recurrent arm-level CNAs to disappear – all support the notion of selection. In line with selection for increased fitness, a recent work demonstrated increased tumor growth rate with PDX passaging[25]. At least three important parameters may account for the different selection pressures between patients and PDXs: the species (human vs. mouse), the anatomical and physiological context (a specific organ vs. subcutaneous growth), and the interaction with the immune system (immunocompetent patients vs. immunodeficient animals). We note that relaxation of selection, followed by genetic drift, may also play a role in the observed dynamics. In the future, comparisons of orthotopic vs. subcutaneous PDXs, and of mouse-derived xenografts in "humanized" immunocompetent vs. immunodeficient recipients, may help delineate the contribution of each of these parameters to shaping tumor evolutionary pressures.

11. Supplementary Discussion: Implications of PDX genomic instability for their use in drug testing

PDX collections are generally used for drug testing in two different ways: to predict, at the cohort level, the relationship between genotype and dependency; and to predict, at the individual level, a therapeutic response[24]. Our findings have several practical implications for both of these uses. The rapid genomic divergence that we identify on the individual tumor level suggests that PDXs may often not be faithful representations of their parental tumors. If individual PDXs are to be utilized as avatar models for personalized medicine, it will be necessary to ensure that the

model retains the relevant genomic features of the primary tumor from which it was derived, before PDX drug response is used to guide clinical treatment decisions. It will also be advisable to use such avatar models at the earliest passage possible and avoid their prolonged propagation, especially in the context of a 1x1x1 (one animal per model per treatment) experimental design[6]. For population level analyses, our findings highlight the need to document the molecular properties of the models at the same passage as that used for drug testing, rather than relying on an early passage characterization. They also emphasize the importance of large cohorts of PDX models, similar to the large cell line collections that were recently established[20,26], in order to average out random effects when performing drug screens and biomarker studies. Finally, the gradual loss of recurrent primary CNAs suggests that prolonged propagation could lead to under-representation of some hallmark cancer events in late passage PDX cohorts.

12. Supplementary Discussion: Possible explanations for the difference in CNA acquisition rate between PDXs and CLDXs

The comparison of PDXs to CLDXs showed a lower CNA acquisition rate in CLDXs than in PDXs. There are three potential explanations for this difference: a lower degree of heterogeneity in established cancer cell lines, a reduced bottleneck upon cell line transplantation, or a reduced rate of ongoing instability. As cell lines are generally more clonal than primary tumors[27], and as we could attribute much of the CNA dynamics observed in PDXs to expansion of pre-existing clones, we speculate that the reduced heterogeneity of established cell lines explains most of the observed difference, although this question remains to be addressed experimentally. In any case, this difference suggests that although established cell lines don't represent primary tumors as faithfully as newly-derived cell lines and PDXs, their genomic landscapes are more stable.

13. Supplementary Discussion: Implications beyond cancer model systems

Our study may have implications beyond cancer model systems. Recent single cell RNAseq studies used hallmark arm-level CNAs as genetic markers to distinguish between tumor and non-tumor cells[28,29]. The finding that some of these events, such as trisomy 7 and monosomy 10 in GBM, can disappear in PDXs, suggests that minor subclones without these aberrations probably exist in primary tumors; therefore, cells should not be classified as non-tumor cells solely based on the absence of a single hallmark event.

Supplementary References

1 Eirew, P. *et al.* Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature* **518**, 422-426, doi:10.1038/nature13952 (2015).

2 Bi, W. L. *et al.* Genomic landscape of high-grade meningiomas. *npj Genom Med* **2**, doi:10.1038/s41525-017-0014-7 (2017).

3 Gibson, W. J. *et al.* The genomic landscape and evolution of endometrial carcinoma progression and abdominopelvic metastasis. *Nat Genet* **48**, 848-855, doi:10.1038/ng.3602 (2016).

4 Zhang, X. *et al.* A renewable tissue resource of phenotypically stable, biologically and ethnically diverse, patient-derived human breast cancer xenograft models. *Cancer Res* **73**, 4885-4897, doi:10.1158/0008-5472.CAN-12-4081 (2013).

5 Gao, H. *et al.* High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat Med* **21**, 1318-1325, doi:10.1038/nm.3954 (2015).

6 Bruna, A. *et al.* A Biobank of Breast Cancer Explants with Preserved Intra-tumor Heterogeneity to Screen Anticancer Compounds. *Cell*, doi:10.1016/j.cell.2016.08.041 (2016).

7 Zack, T. I. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat Genet* **45**, 1134-1140, doi:10.1038/ng.2760 (2013).

8 Ciriello, G. *et al.* Emerging landscape of oncogenic signatures across human cancers. *Nat Genet* **45**, 1127-1133, doi:10.1038/ng.2762 (2013).

9 Gordon, D. J., Resio, B. & Pellman, D. Causes and consequences of aneuploidy in cancer. *Nat Rev Genet* **13**, 189-203, doi:10.1038/nrg3123 (2012).

10 Solimini, N. L. *et al.* Recurrent hemizygous deletions in cancers may optimize proliferative potential. *Science* **337**, 104-109, doi:10.1126/science.1219580 (2012).

11 Davoli, T. *et al.* Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell* **155**, 948-962, doi:10.1016/j.cell.2013.10.011 (2013).

12 Carter, S. L., Eklund, A. C., Kohane, I. S., Harris, L. N. & Szallasi, Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet* **38**, 1043-1048, doi:10.1038/ng1861 (2006).

13 Silk, A. D. *et al.* Chromosome missegregation rate predicts whether aneuploidy will promote or suppress tumors. Proc Natl Acad Sci U S A **110**, E4134-4141, doi:10.1073/pnas.1317042110 (2013).

14 Andor, N. *et al.* Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat Med* **22**, 105-113, doi:10.1038/nm.3984 (2016).

15 Hodgson, J. G. *et al.* Comparative analyses of gene copy number and mRNA expression in glioblastoma multiforme tumors and xenografts. *Neuro Oncol* **11**, 477-487, doi:10.1215/15228517-2008-113 (2009).

16 DeRose, Y. S. *et al.* Tumor grafts derived from women with breast cancer authentically reflect tumor pathology, growth, metastasis and disease outcomes. *Nat Med* **17**, 1514-1520, doi:10.1038/nm.2454 (2011).

17 Lin, D. *et al.* High fidelity patient-derived xenografts for accelerating prostate cancer discovery and drug development. *Cancer Res* **74**, 1272-1283, doi:10.1158/0008-5472.CAN-13-2921-T (2014).

18 Croft, D. R. *et al.* Conditional ROCK activation in vivo induces tumor cell dissemination and angiogenesis. *Cancer Res* **64**, 8994-9001, doi:10.1158/0008-5472.CAN-04-2052 (2004).

19 Vigil, D. *et al.* ROCK1 and ROCK2 are required for non-small cell lung cancer anchorage-independent growth and invasion. *Cancer Res* **72**, 5338-5347, doi:10.1158/0008-5472.CAN-11-2373 (2012).

20 Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603-607, doi:10.1038/nature11003 (2012).
21 Rees, M. G. *et al.* Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat Chem Biol* **12**, 109-116, doi:10.1038/nchembio.1986 (2016).
22 Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564-576 e516, doi:10.1016/j.cell.2017.06.010 (2017).
23 Rubio-Viqueira, B. *et al.* An in vivo platform for translational drug development in pancreatic cancer. *Clin Cancer Res* **12**, 4652-4661, doi:10.1158/1078-0432.CCR-06-0113 (2006).
24 Hidalgo, M. *et al.* Patient-derived xenograft models: an emerging platform for translational cancer research. *Cancer Discov* **4**, 998-1013, doi:10.1158/2159-8290.CD-14-0001 (2014).
25 Pearson, A. T. *et al.* Patient-derived xenograft (PDX) tumors increase growth rate with time. *Oncotarget* **7**, 7993-8005, doi:10.18632/oncotarget.6919 (2016).
26 Garnett, M. J. *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570-575, doi:10.1038/nature11005 (2012).
27 Gillet, J. P., Varma, S. & Gottesman, M. M. The clinical relevance of cancer cell lines. *J Natl Cancer Inst* **105**, 452-458, doi:10.1093/jnci/djt007 (2013).
28 Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396-1401, doi:10.1126/science.1254257 (2014).
29 Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189-196, doi:10.1126/science.aad0501 (2016).