

Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution

Peter Eirew^{1,2*}, Adi Steif^{1,2*}, Jaswinder Khattra^{1,2*}, Gavin Ha^{1,2}, Damian Yap^{1,2}, Hossein Farahani^{1,2}, Karen Gelmon³, Stephen Chia³, Colin Mar³, Adrian Wan¹, Emma Laks^{1,2}, Justina Biele^{1,2}, Kary Shumansky¹, Jamie Rosner¹, Andrew McPherson^{1,2}, Cydney Nielsen^{1,2}, Andrew J. L. Roth^{1,2}, Calvin Lefebvre^{1,2}, Ali Bashashati^{1,2}, Camila de Souza¹, Celia Siu¹, Radhouane Aniba^{1,2}, Jazmine Brimhall¹, Arusha Oloumi^{1,2}, Tomo Osako^{1,2}, Alejandra Bruna^{4,5}, Jose L. Sandoval^{4,5}, Teresa Algara^{1,2}, Wendy Greenwood^{4,5}, Kaston Leung^{6,7}, Hongwei Cheng^{8,9}, Hui Xue^{8,9}, Yuzhuo Wang^{8,9}, Dong Lin^{8,9}, Andrew J. Mungall¹⁰, Richard Moore¹⁰, Yongjun Zhao¹⁰, Julie Lorette¹¹, Long Nguyen^{12,13}, David Huntsman^{2,11}, Connie J. Eaves^{12,13}, Carl Hansen^{6,7}, Marco A. Marra¹⁰, Carlos Caldas^{4,5}, Sohrab P. Shah^{1,2,10,11} & Samuel Aparicio^{1,2,10,11}

Human cancers, including breast cancers, comprise clones differing in mutation content. Clones evolve dynamically in space and time following principles of Darwinian evolution^{1,2}, underpinning important emergent features such as drug resistance and metastasis^{3–7}. Human breast cancer xenoengraftment is used as a means of capturing and studying tumour biology, and breast tumour xenografts are generally assumed to be reasonable models of the originating tumours^{8–10}. However, the consequences and reproducibility of engraftment and propagation on the genomic clonal architecture of tumours have not been systematically examined at single-cell resolution. Here we show, using deep-genome and single-cell sequencing methods, the clonal dynamics of initial engraftment and subsequent serial propagation of primary and metastatic human breast cancers in immunodeficient mice. In all 15 cases examined, clonal selection on engraftment was observed in both primary and metastatic breast tumours, varying in degree from extreme selective engraftment of minor (<5% of starting population) clones to moderate, polyclonal engraftment. Furthermore, ongoing clonal dynamics during serial passaging is a feature of tumours experiencing modest initial selection. Through single-cell sequencing, we show that major mutation clusters estimated from tumour population sequencing relate predictably to the most abundant clonal genotypes, even in clonally complex and rapidly evolving cases. Finally, we show that similar clonal expansion patterns can emerge in independent grafts of the same starting tumour population, indicating that genomic aberrations can be reproducible determinants of evolutionary trajectories. Our results show that measurement of genetically defined clonal population dynamics will be highly informative for functional studies using patient-derived breast cancer xenoengraftment.

To evaluate xenograft clonal dynamics (see Supplementary Table 1 for definitions of terms used) we generated 30 xenograft lines by serially transplanting (up to 16 generations over 3 years) breast cancer tissue organoid suspensions from 55 patients (Extended Data Fig. 1, Supplementary Table 2 and Supplementary Fig. 1) into highly immunodeficient *NOD/SCID/Il2rg^{-/-}* (NSG) and *NOD/Rag1^{-/-}/Il2rg^{-/-}* (NRG) mice¹¹ (details in the Supplementary Information). We carried out massively parallel whole-genome shotgun sequencing (WGSS) on DNA from xenograft passages of 15 patient lines (10 primary tumour-derived and five pleural effusion-derived), along with matched patient tumour and

normal DNA (47 samples total, median sequencing depth 45.1, Supplementary Table 3). For these, plus 56 additional xenograft passage samples, we validated 3,187 somatic single nucleotide variant (SNV) positions (100–300 per tumour-xenograft series) and 132 structural variant positions by targeted-amplicon deep sequencing (Supplementary Tables 4–6), quantifying allele ratios to a high level of precision. We surveyed the copy-number alteration (CNA) and loss of heterozygosity (LOH) landscapes using Affymetrix SNP Array 6.0 (Supplementary Tables 7 and 8). The mutation load of somatic SNVs (range: 4.3–27.7 × 10³ genome-wide; 57–1,040 in coding regions), CNA and LOH (34–67% of genome), and structural variants in the 15 tumour-xenograft series (Supplementary Figs 2 and 3 and Supplementary Table 9) were consistent with previous genome-wide breast cancer studies^{4,12–17}, although low tumour cellularity hindered mutation discovery in case numbers SA429 and SA496 originating tumours. Tumour-xenograft pairs displayed comparable nucleotide substitution patterns (Supplementary Figs 2 and 4), suggesting that mutational processes are maintained post-engraftment.

To determine the extent of evolution in the SNV landscape, we first compared the genome-wide variant allele prevalences (the proportion of aligned reads at the SNV position with the variant base, see Supplementary Table 1) from WGSS data in xenograft relative to tumour (SA429 and SA496 excluded due to low tumour cellularity). As expected, sizeable proportions (range: 53.0–92.9%) of high-confidence SNVs are shared in tumour-xenograft pairs, with prevalences lying on a scatter plot diagonal indicating neutral dynamics (Extended Data Fig. 2a and Supplementary Figs 5a and 6). Notably, all 15 samples also show clusters of SNVs prevalent in the xenograft while at or below the limit of detection in the tumour (range: 6.5–32.1% of SNVs, see for example, SA494, SA495 and SA499) and vice versa (range: 0.2–19.4%, see for example, SA494, SA495 and SA500), implying clonal selection on initial engraftment. Tumours and xenografts from SA494, SA495, SA499, SA500 and SA530 also exhibited substantial differences in structural variant content (Supplementary Figs 3 and 7).

To resolve clonal dynamics and genotypes, we applied a Bayesian clustering model (PyClone^{4,18}) to SNV variant allele prevalences measured by targeted deep sequencing, accounting for the effect of copy number, LOH status and cellularity. SNVs with co-varying estimates of cellular prevalence (the proportion of tumour or xenograft cells bearing the

¹Department of Molecular Oncology, BC Cancer Agency, 675 West 10th Avenue, Vancouver, British Columbia V5Z 1L3, Canada. ²Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia V6T 2B5, Canada. ³Department of Medical Oncology, BC Cancer Agency, 600 West 10th Avenue, Vancouver, British Columbia V5Z 4E6, Canada. ⁴Department of Oncology, University of Cambridge, Hills Road, Cambridge CB2 2XZ, UK. ⁵Cancer Research UK Cambridge Research Institute, University of Cambridge, Li Ka Shing Centre, Cambridge CB2 0RE, UK. ⁶Centre for High-Throughput Biology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. ⁷Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada. ⁸Department of Experimental Therapeutics, BC Cancer Agency, 675 West 10th Avenue, Vancouver, British Columbia V5Z 1L3, Canada. ⁹The Vancouver Prostate Centre, Vancouver General Hospital and Department of Urologic Sciences, University of British Columbia, Vancouver, British Columbia V5Z 1M9, Canada. ¹⁰Michael Smith Genome Sciences Centre, Vancouver, British Columbia V5Z 1L3, Canada. ¹¹Centre for Translational and Applied Genomics, BC Cancer Agency, 600 West 10th Avenue, Vancouver, British Columbia V5Z 4E6, Canada. ¹²Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada. ¹³Terry Fox Laboratory, BC Cancer Agency, Vancouver, British Columbia V5Z 1L3, Canada.

*These authors contributed equally to this work.

mutation) across all time points are grouped into putative mutation clusters (Supplementary Table 1). Consistent with the raw variant allele prevalence measurements, several cases contained mutation clusters with high (75–100%) prevalences in the xenografts and low (0–15%) prevalences in the tumours, implying expansion of initially minor clones to dominate the xenograft (for example, clusters 3, 4, 3, 2, 8, 2 and 2 in SA494, SA495, SA500, SA530, SA532, SA533 and SA535, respectively) (Extended Data Fig. 2b and Supplementary Fig. 5b). Other series (SA493, SA499, SA501, SA531, SA534 and SA536) demonstrated non-neutral clonal dynamics but involving alleles occupying much smaller proportions of total cellular populations. Notably, polyclonal population structure specific to the xenograft was observed after initial expansion in SA493, SA494, SA495, SA500 and SA531, suggesting that initial selection on engraftment remains permissive to additional clonal evolution (Extended Data Fig. 2b and Supplementary Fig. 5b). Polyclonal engraftment was evident in SA493, SA501, SA531 and SA532, suggesting that multiple clones maintained their fitness post-engraftment.

Analogously, we analysed clonal dynamics using CNAs as clonal marks, applying a probabilistic model (TITAN¹⁹) that infers CNA and LOH from WGSS data, accounting for mixtures of tumour and normal cells and reporting estimates of mutation cellular prevalence and mutation cluster membership (Supplementary Table 10). Despite conservation of complex disruptions, such as chromothripsis in SA429 (Supplementary Fig. 8) and breakage–fusion–bridge cycles in SA429 and SA494 (Supplementary Figs 9 and 10), we identified substantial differences in copy-number architecture between tumour and xenograft in all cases (Extended Data Fig. 2c and Supplementary Fig. 5c). These included a xenograft-specific deletion event containing *TP53* (in SA500) that coincided with retention of a somatic SNV (Supplementary Fig. 11 and Supplementary Table 6). Notably, the predominant clonal dynamic (minor subclone

expansion in SA494, SA495, SA532 and SA533; polyclonal engraftment in SA493 and SA501) mirrored those seen in SNV space.

We next asked how clonal dynamics differ after initial engraftment, using PyClone predictions over serial passage generations spanning up to 3 years (Extended Data Fig. 1). We distinguished statistically significant directional clonal dynamics by testing the overlap of 90% credible intervals derived from Bayesian posterior probability distributions (Fig. 1). Cases showing strongest clonal dynamics in the first engraftment passages (for example, SA500, SA530, SA494 and SA535) exhibited more stable prevalence over subsequent passages. In contrast, cases showing moderate initial clonal dynamics showed more marked subsequent dynamics (for example, mutation clusters 2, 3 and 8 of SA501), in some cases leading to gradual expansion of a minor clone to dominate the xenograft over serial passages. We noted examples of all oestrogen receptor/HER2 subtypes and primary/metastatic cancers evolving by these two different modes. Some mutation clusters showed non-dynamic patterns over time (for example, clusters 1, 4 and 6 of SA500, clusters 1–3, 5, 7, 9 and 10 in SA532, as well as the highest prevalence clusters representing putative ancestral mutations that remained invariant, as expected). For two cases we noted preferential engraftment of initial transplants in mammary fat pad over subrenal sites (SA496 4 of 4 mammary fat pad versus 0 of 4 subrenal; SA429 2 of 4 mammary fat pad versus 0 of 4 subrenal, Extended Data Fig. 1). However, transplant site changes in established xenografts were not associated with unusually strong clonal dynamics (Fig. 1, see SA495 X3–4, SA499 X3–4, SA429 X1–2 and SA496 X1–2, where X denotes the xenograft passage).

To validate the population-based inference of mutation clusters and clonal genotypes directly, we carried out single-cell analyses of cases SA494 (an example of extreme initial selection) and SA501 (complex post-engraftment clonal dynamics). We performed multiplexed targeted

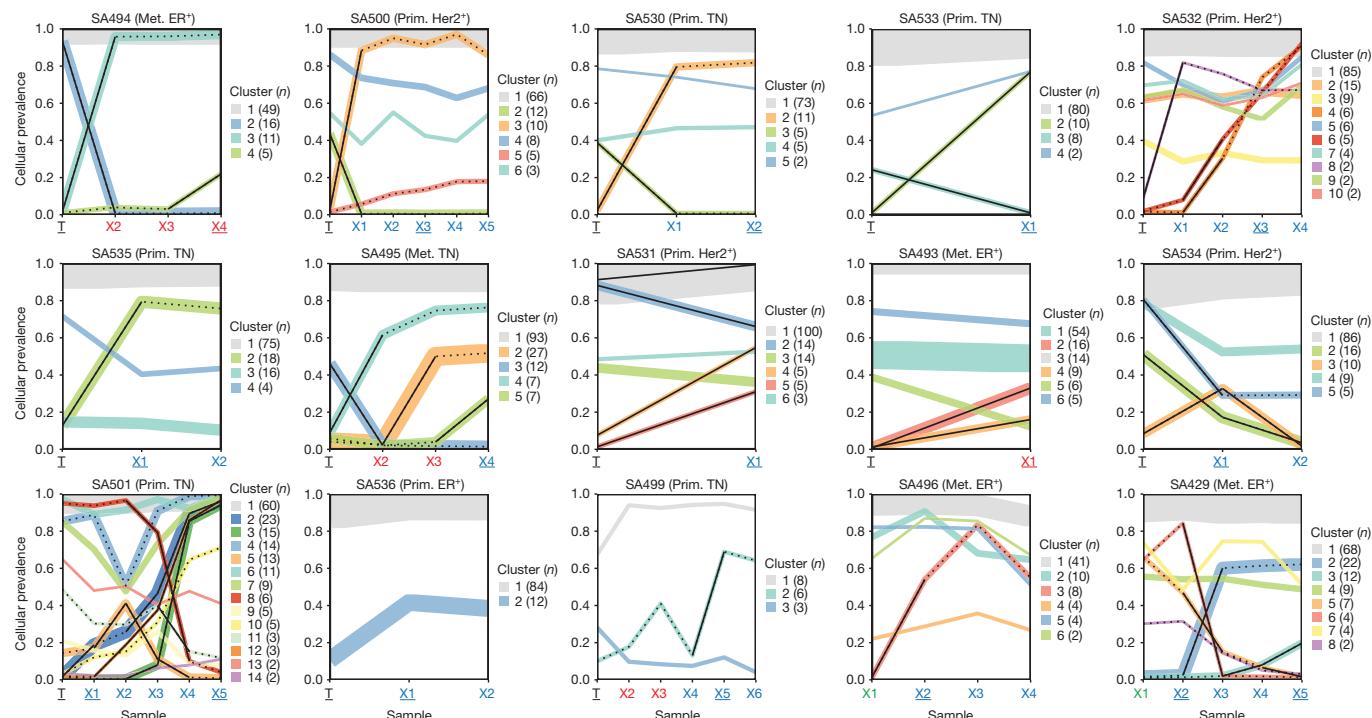


Figure 1 | Clonal dynamics over multiple passages in time. Plots display the mean cellular prevalence estimates of mutation clusters in originating tumours (T) and subsequent xenograft passages (X1, X2, etc.). The clusters and prevalences were inferred by PyClone from bulk population-level targeted deep sequence data. Line widths indicate the number of SNVs comprising each mutation cluster (numbers in brackets adjacent to each plot). Black lines indicate non-neutral dynamics, assessed by non-overlap of credible intervals derived from Bayesian posterior distributions (solid lines, non-neutral over indicated passage; dotted lines, non-neutral over cumulative passages since

initial transplant). All passages that underwent deep sequencing are shown. Transplant sites are represented by colour (blue, subcutaneous; red, subrenal; green, mammary fat pad), tumour and passages analysed by WGSS are underlined. The panels are ordered left to right and top to bottom by the degree of initial change in mutation cellular prevalence. Singleton clusters were not displayed for clarity. ER⁺, oestrogen receptor positive; Her2⁺, Her2 positive; Met., metastatic pleural effusion; Prim., primary breast; TN, triple negative breast cancer.

re-sequencing of SNVs in 210 isolated tumour and xenograft nuclei, using microfluidic devices. We determined evolutionary relationships between nuclei by Bayesian phylogenetic inference²⁰, deriving consensus genotypes for clades representing high probability branch points in the phylogenetic tree.

As predicted by PyClone, two major clades emerge in the SA494 phylogeny, comprising tumour and xenograft nuclei respectively, bearing mutually exclusive sets of alleles in addition to a set of shared alleles (Extended Data Fig. 3a–c and Supplementary Fig. 13). The ancestral clone SNVs (PyClone cluster 1) are common to nuclei from both clades, while SNVs in the predicted dominant tumour clone (cluster 2) and

minor engrafting clone (cluster 3) are restricted to tumour and xenograft nuclei, respectively (Extended Data Fig. 3d, genotypes A and B). This confirms the ancestral relationship between tumour and xenograft, verifies the expansion of a very minor clone (<5%), while also showing unambiguously that mutation clusters inferred by PyClone represent major clonal genotypes.

PyClone analysis of SA501 (Fig. 2 and Supplementary Fig. 12) revealed a dynamic and complex clonal architecture, with gradual expansion of minor mutation clusters observed over consecutive passages, and expansion followed by decline of other clusters (Fig. 2d). The major mutation clusters and their gradual change in prevalence over time

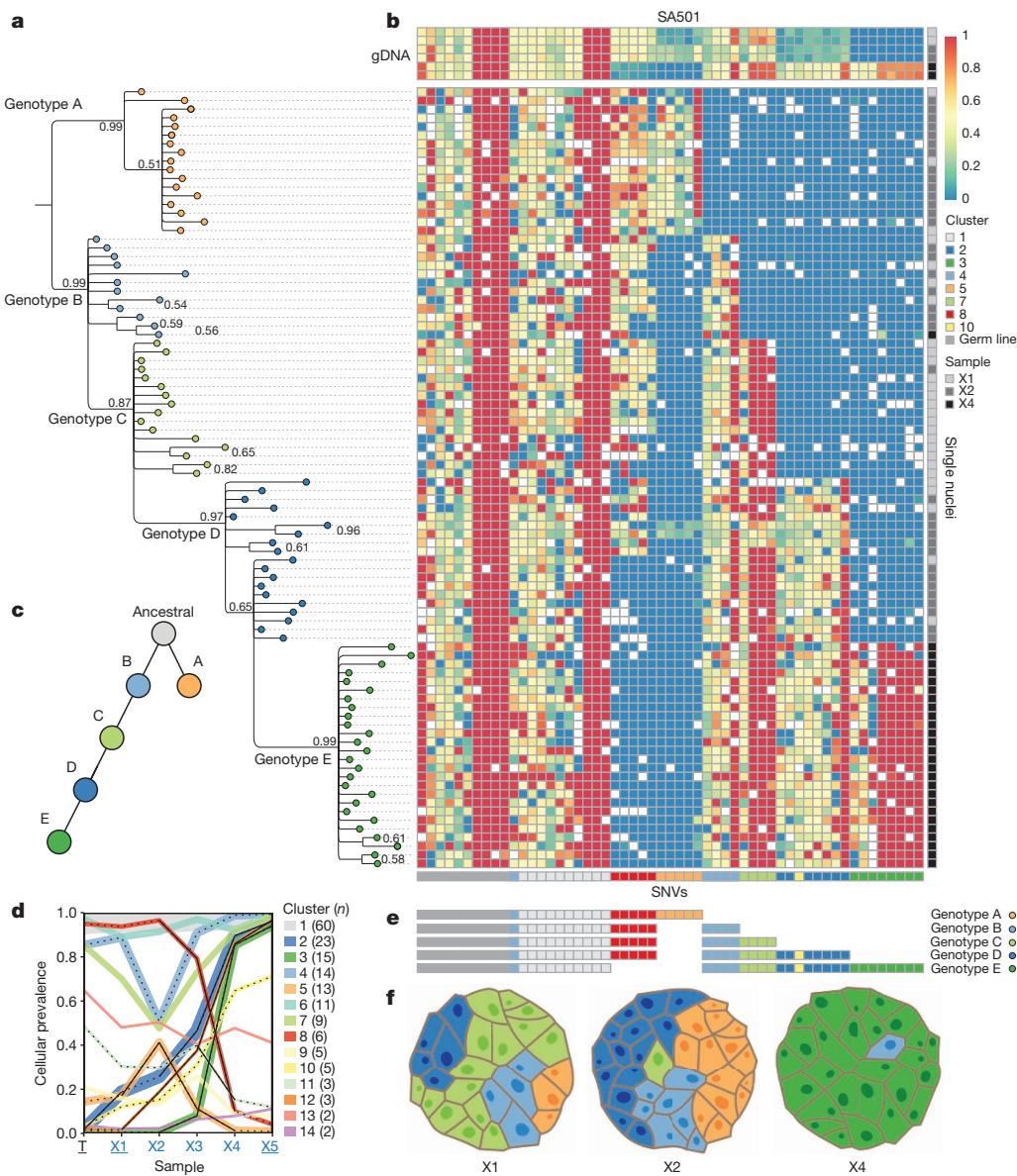


Figure 2 | Single-cell determination of clonal genotypes recapitulates population-based prediction of cascading subclonal evolution. DNA was prepared from 90 individual lysed SA501 xenograft nuclei from passages X1, X2 and X4, and the variant allele ratios were determined by targeted ultra-deep sequencing at 45 somatic SNV and 10 germline SNV positions. **a**, Bayesian phylogenetic tree derived from multi-locus genotypes of individual nuclei, depicting cascading evolution. **b**, Heat map depicting multilocus variant allele ratios (blue/yellow/red corresponds to wild-type/heterozygous/homozygous loci, respectively). Nuclei (y axis) are ordered according to the phylogenetic tree in **a**. gDNA, genomic DNA. Positions (x axis) are grouped according to the consensus genotypes derived from high-probability branch splits in a manner naïve to the PyClone clustering. The cluster groupings (horizontal bar below

horizontal axis) recapitulate the PyClone groupings inferred from bulk population measurements (**d**). **c**, Schematic of the phylogeny derived from single-cell genotyping depicts the sequential expansion of genomic subclones. Genotypes are coloured according to the last PyClone mutation cluster acquired at a given point in the phylogeny. **d**, PyClone inference of temporal clonal dynamics from bulk population measurements. **e**, Five consensus genotypes derived from high-probability splits in the phylogenetic tree. **f**, Schematic representations of xenografts X1, X2 and X4 based on single-cell genotypes. Cells are coloured according to their genotype in **c**, and the number of cells within each schematic corresponds to the number of sequenced nuclei with the given genotype in **b**. The relative proportions of cells with each genotype reflect predictions based on bulk measurements in **d**.

predicted by PyClone were confirmed by the clonal genotypes of single cells from SA501 passages X1, X2 and X4 (Fig. 2b and Supplementary Fig. 13). Phylogenetic inference resolved the clonal genotypes of five major clades (Fig. 2a, e), with cascading acquisition of mutations from parental to descendant clone (Fig. 2c). Genotypes A and B belong to sibling clades defined by the addition of cluster 5 and cluster 4 mutations, respectively, to the ancestral genotype defined by clusters 1 and 8; genotype C was derived from genotype B with the addition of mutations in cluster 7; genotype D derived from genotype C with the addition of mutations defined by cluster 2; and genotype E derived from genotype D with the addition of cluster 3 mutations and loss of cluster 8 mutations (Fig. 2a, c, e). The clonal dynamics measured in the population was reflected in the relative abundance of single-cell genotypes in each xenograft (Fig. 2f), mirroring bulk population predictions (Fig. 2d). Both X1- and X2-sampled nuclei show an admixture of clones defined by genotypes A, B, C and D (relatively rare in X1). Genotype E is confined exclusively to X4 nuclei, suggesting that by passage 4, this clone had nearly exhaustively outcompeted its ancestor and sibling clones. Its eventual dominance is mirrored by the decline of genotype A (initially present in X1 and X2), suggesting that the descendants of genotype B outcompeted those of genotype A over time.

Taken together, these single-cell genotyping experiments combined with phylogenetic inference have recapitulated population-level PyClone predictions in a simple (SA494) and a complex (SA501) clonal expansion model. Thus, single-cell genotyping validates PyClone mutation clusters as genomic markers of major clonal genotypes, while providing additional insight into the ancestral lineages of cell populations.

Finally, to determine whether directional clonal dynamics might be associated with deterministic as opposed to stochastic processes (such as random genetic drift), we tested whether similar clonal dynamics occurred when the same tumour population was multiply transplanted into different mice. In 4 of 5 series examined, parallel clonal dynamics of the same mutation cluster(s) were observed (arrows in Fig. 3a, b and Extended Data Fig. 4a, b: SA501 2 of 2 replicate mice at passage X3 and 4 of 4 at X4; SA535 3 of 3 at X1; SA532 3 of 3 at X1, 3 of 7 at X2 and 2 of 2 at X3; SA429 3 of 5 at X2). These include reproducible expansions of initially minor subclones, implying a high likelihood of a shared deterministic mechanism rather than repeated rare stochastic events (for example, arising from transplants close to limiting dilution). In SA501 the same pattern (expansion of cluster 3 mutations mirrored by a decline of cluster 5 mutations) was independently observed in transplants at passage 2, 3 and 4 (2B, 3B and 4A–D in Fig. 3a), suggesting shared clonal fitness but variable timing. We also observed instances of divergence, for example expansion of SA532 cluster 4 specific to branch 1A–2A–3A–4A (Extended Data Fig. 4a). SA535 (Fig. 3b) and SA532 showed examples of clonal expansion patterns replicated in related but different immunodeficient mouse strains (NSG, NRG). To control against shared clonal structure imposed through joint inference of the data sets, we also carried out independent PyClone analyses that excluded all but one transplant at each passage, and observed high correlations of inferred mutation prevalences between same-passage replicates (Extended Data Fig. 5; median Pearson correlations 0.94, 0.93, 0.91, 0.91 and 0.46 for SA501, SA535, SA532, SA429 and SA496, respectively). These data indicate that clonal genotypes defined by somatic aberrations (and/or closely co-segregating genomic factors) can be biologically meaningful determinants of fitness, leading to consistent and reproducible clonal dynamics.

We show here that patient-derived xenograft clonal dynamics on initial transplant vary from polyclonal engraftment with only moderate clonal selection, in which tumour and xenograft clonal prevalence are broadly similar (a minority of cases), to highly skewed dynamics in which initially minor prevalence clones expand to dominate the xenograft (the majority of cases). Expansion of minor subclones has been suggested in previous xenotransplantation studies using malignant epithelial^{10,21–23} or haematopoietic^{24,25} cells, without formal resolution of the clonal genotypes or pattern of subsequent clonal dynamics. In contrast with

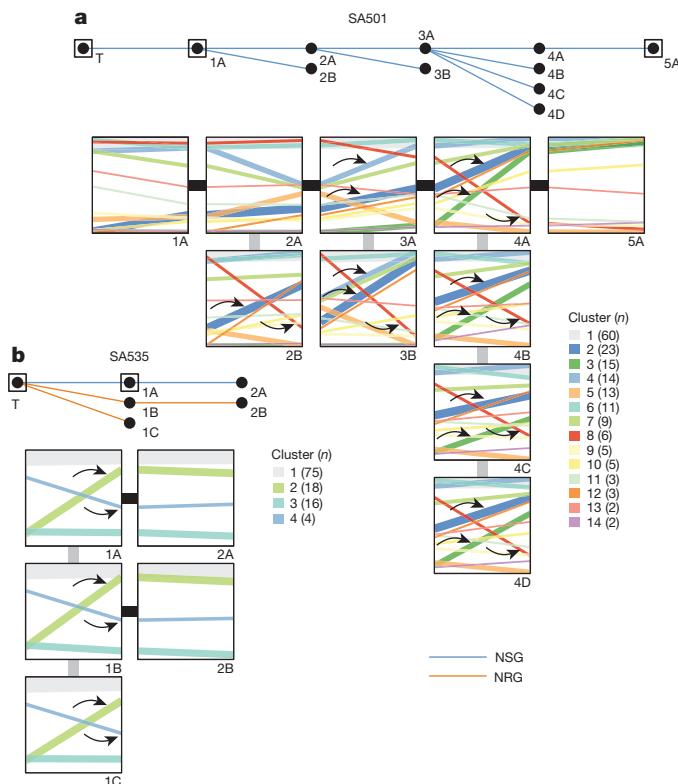


Figure 3 | Clonal dynamics are reproduced in replicate transplants. **a, b,** Upper panels, passaging history of SA501 and SA535 showing transplants that resulted in successful xenografts. The host mouse strains (blue, NSG; orange, NRG) are indicated. All transplants were in subcutaneous sites. Lower panels, change in cellular prevalence of mutation clusters over individual transplants. Plots correspond to passages in upper panels. The clusters are inferred by PyClone using grouped data from all passages, and correspond to those displayed in Fig. 1. Boxed nodes indicate passages analysed by WGS. Arrows show examples of parallel clonal dynamics of the same mutation cluster in multiple replicate transplants.

preliminary studies of xenoengraftment, we find correlated dynamics of clones defined by SNVs or copy-number aberrations as clonal marks. Expansion patterns are most often pronounced in the initial establishment passage; however, in cases where initial clonal selection is weak, subsequent evolution over passaging is more evident. Furthermore, polyclonal sub-structure may emerge even in xenografts that have undergone a modest population bottleneck on initial engraftment. These dynamic processes are not evident from histopathological or imaging characteristics, which remain broadly stable, consistent with previous reports^{8,9,23}.

Notably, we find that the population dynamics of genetically defined clones are replicated when transplants are carried out in multiple mice, implying that the basis of selection is non-random and probably closely linked to the particular mutation genotype (or epigenotype) that defines the clone. The most parsimonious explanation for repeated observation of these clonal dynamics is that the clones are mostly pre-existing, and variations in clonal fitness explain the dynamic behaviour, as opposed to *de novo* somatic mutation. Furthermore, cases in which conversion from minor to dominant clone occurs monotonically over multiple passages demonstrate that selective fitness can be persistent rather than transient. Thus, specific somatic genotypes are likely to act as genetic markers of clonal growth and fitness advantages, yielding predictable and reproducible clonal dynamics. Determination of the precise aberrations that give rise to selective clonal fitness still faces considerable challenges. In this regard, we believe that ascertainment of clonal dynamics will prove essential for fully informed future studies of drug response and tumour biology in xenografts of human breast cancers.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 August 2013; accepted 8 October 2014.

Published online 26 November 2014.

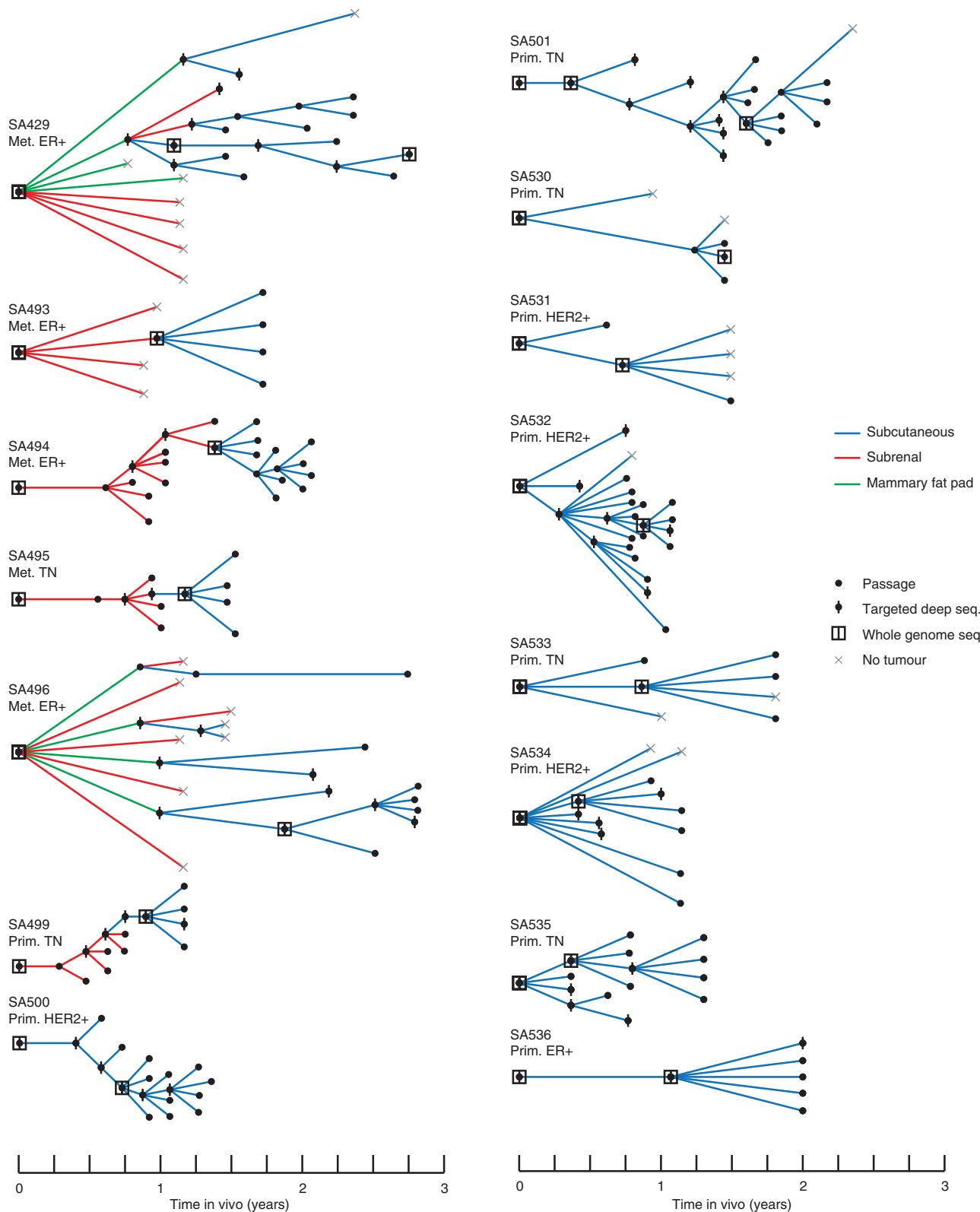
1. Aparicio, S. & Caldas, C. The implications of clonal genome evolution for cancer medicine. *N. Engl. J. Med.* **368**, 842–851 (2013).
2. Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194**, 23–28 (1976).
3. Diaz, L. A. Jr *et al.* The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486**, 537–540 (2012).
4. Shah, S. P. *et al.* The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **486**, 395–399 (2012).
5. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
6. Campbell, P. J. *et al.* The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467**, 1109–1113 (2010).
7. Bashashati, A. *et al.* Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. *J. Pathol.* **231**, 21–34 (2013).
8. DeRose, Y. S. *et al.* Tumor grafts derived from women with breast cancer authentically reflect tumor pathology, growth, metastasis and disease outcomes. *Nature Med.* **17**, 1514–1520 (2011).
9. 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 (2013).
10. Ding, L. *et al.* Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* **464**, 999–1005 (2010).
11. Pearson, T. *et al.* Non-obese diabetic-recombination activating gene-1 (NOD-Rag1^{null}) interleukin (IL)-2 receptor common gamma chain (IL2r^{null}) null mice: a radioresistant model for human lymphohaematopoietic engraftment. *Clin. Exp. Immunol.* **154**, 270–284 (2008).
12. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
13. Ha, G. *et al.* Integrative analysis of genome-wide loss of heterozygosity and monoallelic expression at nucleotide resolution reveals disrupted pathways in triple-negative breast cancer. *Genome Res.* **22**, 1995–2007 (2012).
14. Nik-Zainal, S. *et al.* The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).
15. Ellis, M. J. *et al.* Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* **486**, 353–360 (2012).
16. Banerji, S. *et al.* Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* **486**, 405–409 (2012).
17. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).
18. Roth, A. *et al.* PyClone: statistical inference of clonal population structure in cancer. *Nature Methods* **11**, 396–398 (2014).
19. Ha, G. *et al.* Titan: inference of copy number architectures in clonal cell populations from tumor whole genome sequence data. *bioRxiv* **24**, 1881–1893 (2014).
20. Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
21. Kreso, A. *et al.* Variable clonal repopulation dynamics influence chemotherapy response in colorectal cancer. *Science* **339**, 543–548 (2013).
22. Nolan-Stevaux, O. *et al.* Measurement of cancer cell growth heterogeneity through lentiviral barcoding identifies clonal dominance as a characteristic of *in vivo* tumor engraftment. *PLoS ONE* **8**, e67316 (2013).
23. Li, S. *et al.* Endocrine-therapy-resistant ESR1 variants revealed by genomic characterization of breast-cancer-derived xenografts. *Cell Rep.* **4**, 1116–1130 (2013).
24. Notta, F. *et al.* Evolution of human BCR-ABL1 lymphoblastic leukaemia-initiating cells. *Nature* **469**, 362–367 (2011).
25. Clappier, E. *et al.* Clonal selection in xenografted human T cell acute lymphoblastic leukemia recapitulates gain of malignancy at relapse. *J. Exp. Med.* **208**, 653–661 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to the staff of the CTAG Molecular Pathology facility, members of the Library Technical Development, Library Construction, Sequencing and Bioinformatics teams at the Michael Smith Genome Sciences Centre for technical assistance with data generation, and S. Kalloger for assistance with sample collection. S.A. and S.P.S. are supported by Canada Research Chairs. P.E. is supported by a Michael Smith Foundation for Health Research (MSFHR) Fellowship. A.S. is supported by an NSERC CREATE scholarship through the graduate program in Genome Science and Technology at UBC. S.P.S. is a MSFHR scholar. We acknowledge long-term funding support provided by the BC Cancer Foundation. The S.A., S.P.S. and C.H. groups receive operating funds from the Canadian Breast Cancer Foundation, Canadian Cancer Society Research Institute, Terry Fox Research Institute, Genome Canada and Canadian Institutes for Health Research (CIHR). We thank S. Mullaly for critical reading of the manuscript.

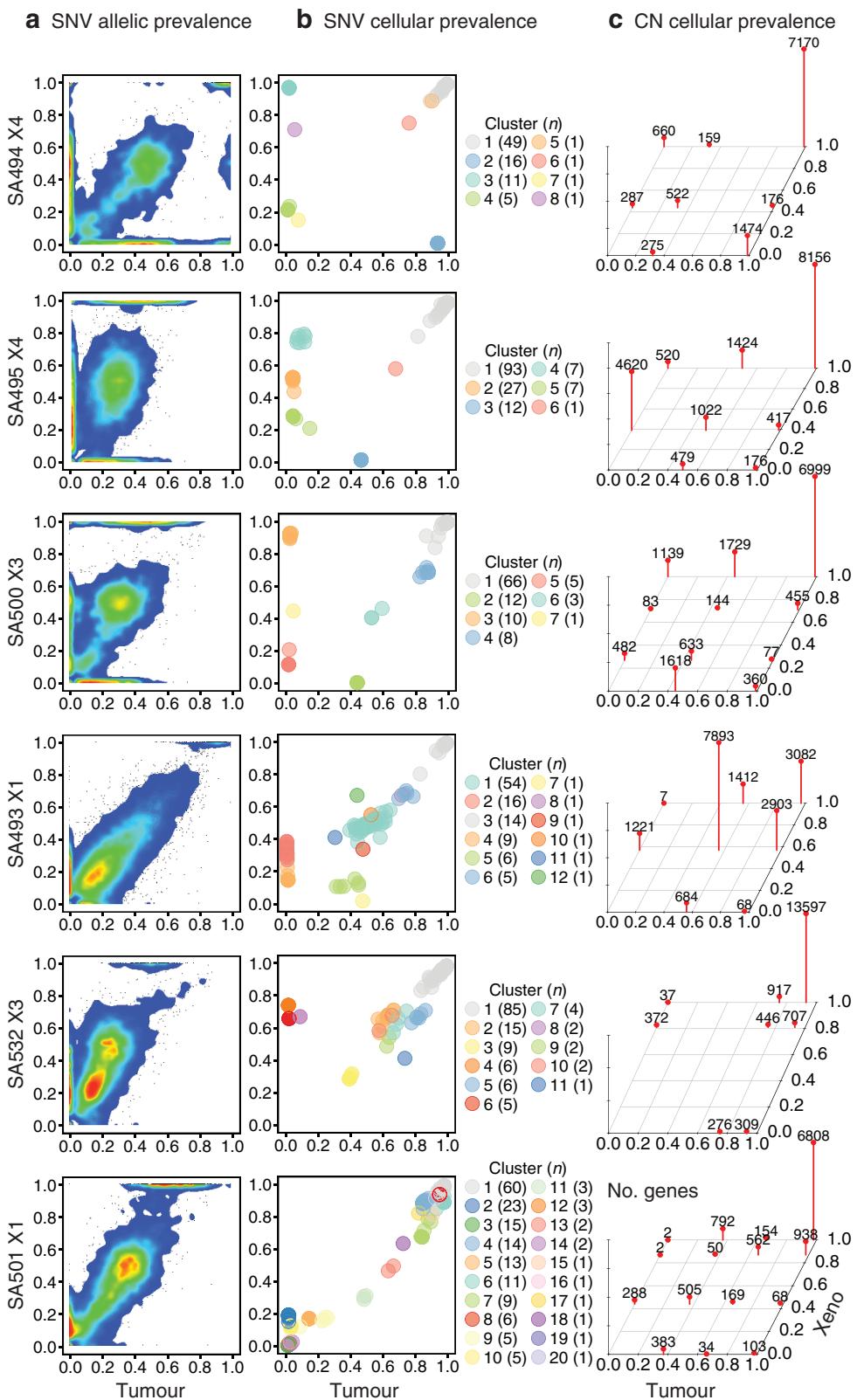
Author Contributions S.A. and S.P.S. designed the study and supervised the research. S.A., S.P.S., P.E. and A.S. wrote the paper. P.E., A.Br., J.S., T.A., W.G., H.C., H.X., L.N., Y.W. and D.L. performed transplants and passaging. K.G., S.C. and C.M. recruited patients and performed tissue biopsies. A.S., P.E., G.H., C.N., H.F., A.J.L.R., C.L., A.Ba., C.S., K.S., J.R., R.A., A.M., C.D.S., S.P.S. and S.A. carried out bioinformatics analyses. J.K., D.Y., E.L., J.Br., A.W., J.Bi., K.L., A.J.M., A.O., R.M., Y.Z., C.H. and M.A.M. assisted with sequence generations and single-cell experiments. T.O., J.L. and D.H. contributed to histological analysis. C.J.E., C.H., M.A.M., C.C., S.P.S. and S.A. provided intellectual contributions to design or interpretation.

Author Information Genome data have been deposited at the European Genome-phenome Archive (<http://www.ebi.ac.uk/ega/>) under accession number EGAS00001000952. Processed data can be viewed at <http://www.ncbi.nlm.nih.gov/geo/>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.A. (saparicio@bccrc.ca) or S.P.S. (sshah@bccrc.ca).



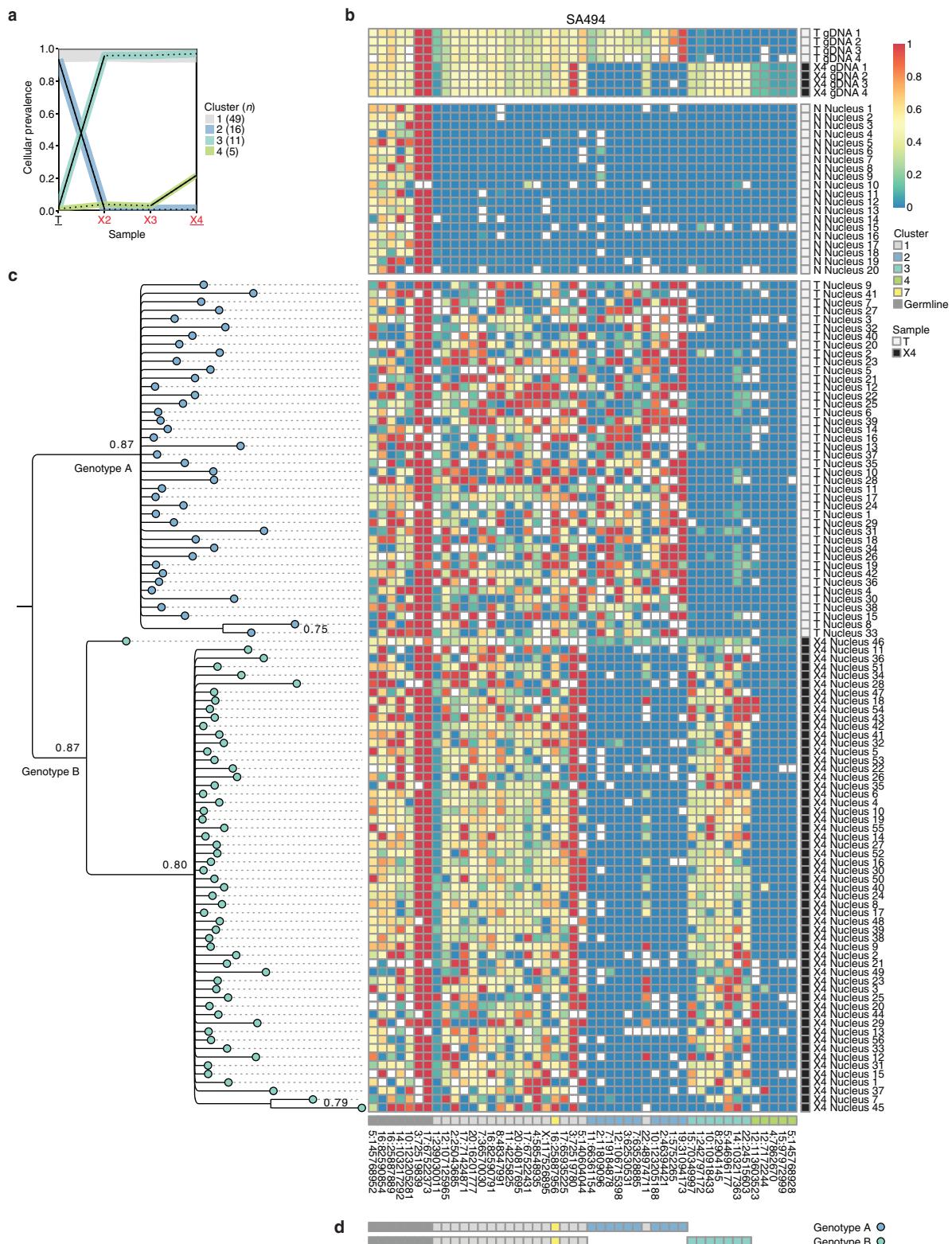
Extended Data Figure 1 | Transplant history. Diagrams show the transplant history of each xenograft line. Line segment colours represent the site used for each transplant (blue, subcutaneous; red, subrenal capsule; green, mammary fat pad). Black points indicate the passage of an engrafted xenograft to the next mouse generation. Grey crosses indicate transplants that did not result in palpable tumours. Samples analysed by whole-genome and/or targeted deep

sequencing are indicated (black squares and vertical lines, respectively). The cumulative time *in vivo* is shown on the *x* axis. The originating tumour site (Met., pleural effusion; Prim., primary breast) and immunohistochemical expression of biomarkers (ER, oestrogen receptor; PR, progesterone receptor; TN, triple negative for ER, PR and HER2) are shown.



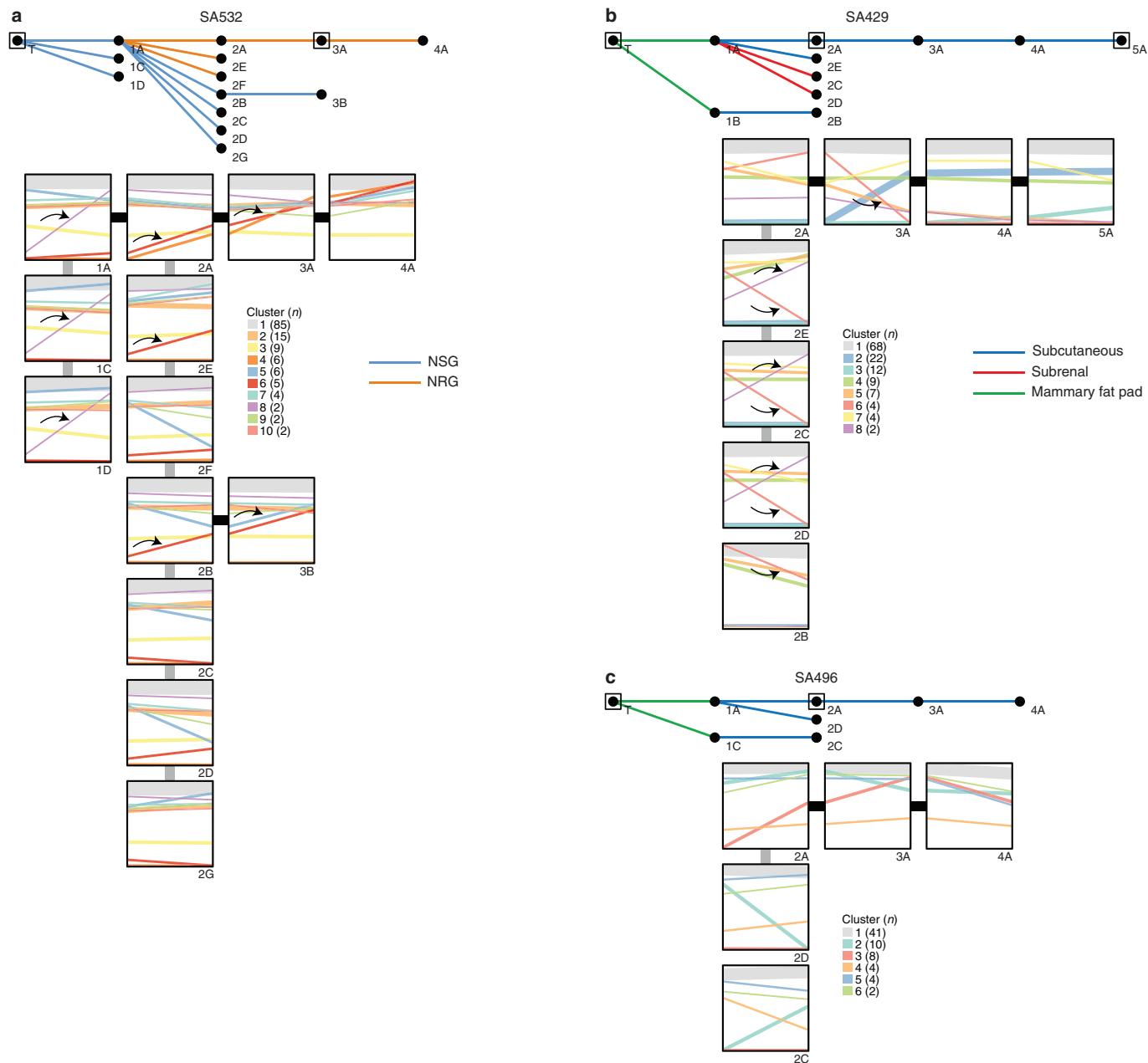
Extended Data Figure 2 | Comparison of the prevalence of mutations in six originating tumours and subsequent xenografts in SNV and CNA spaces. **a**, Density scatter plots showing the WGSS variant allele prevalence of genome-wide high-confidence SNVs in tumours (x axis) and xenografts (y axis). SNVs in clones undergoing neutral dynamics lie along a diagonal, and SNVs in clones undergoing expansion or contraction lie on/towards the y and x axes, respectively. **b**, Scatter plots showing the mutation cellular prevalence of selected SNVs in tumours and xenografts, inferred by PyClone from

population-level targeted deep sequencing. Circles represent individual SNVs, colours indicate clusters of mutations for which mutation cellular prevalences vary together over all sample time points. **c**, Scatter plots show co-occurrence of CNA/LOH events inferred by TITAN in tumours and xenografts. The z axis height of each bar shows the number of genes belonging to a unique mutation cluster and present at the indicated mutation cellular prevalence in tumour (x axis) and xenograft (y axis).



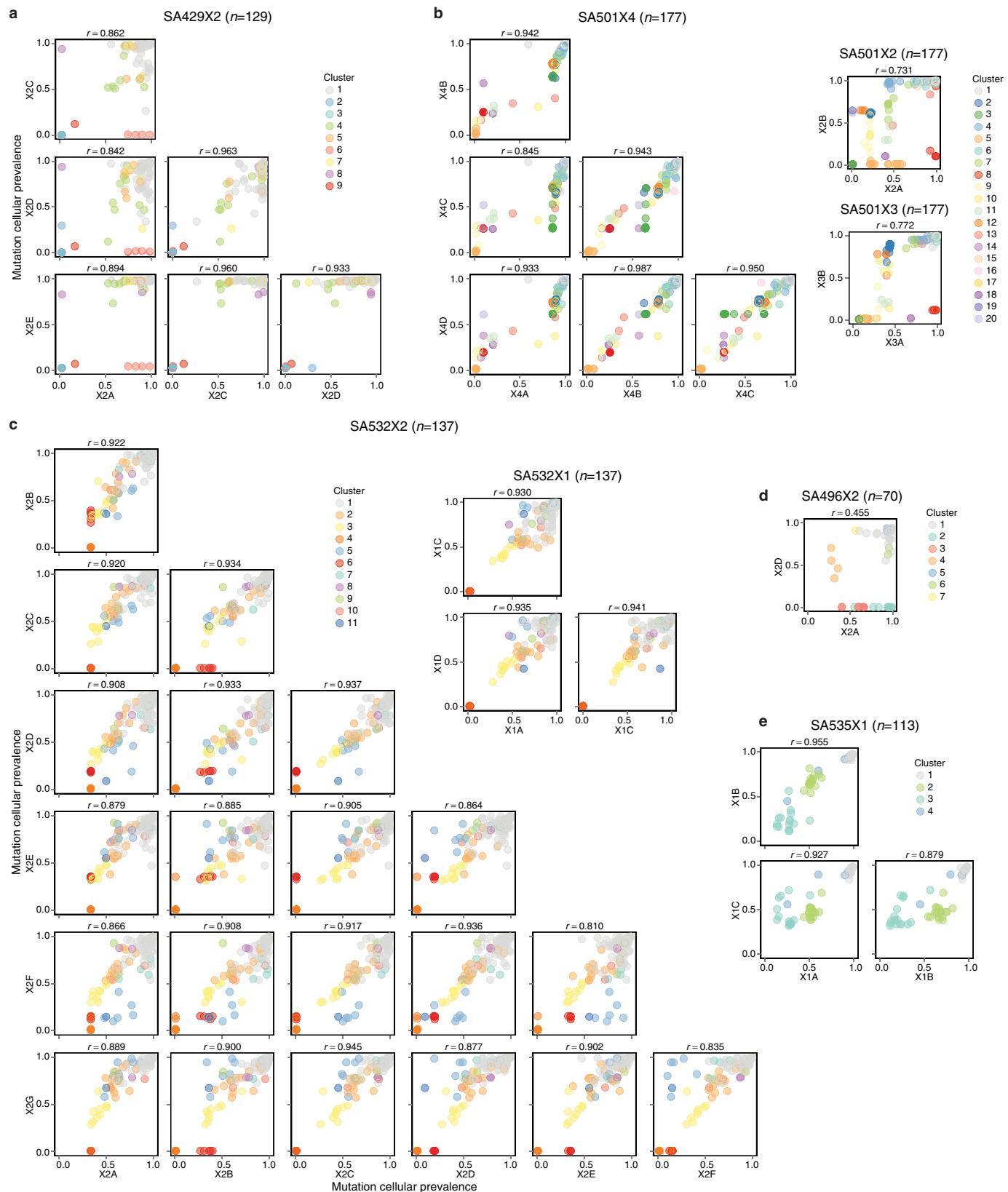
Extended Data Figure 3 | Single-cell determination of clonal genotypes recapitulates population-based prediction of minor clone selection. DNA prepared from 62 individual lysed SA494 tumour and 58 passage 4 lysed xenograft nuclei was amplified in single reactions using a panel of multiplexed PCR primer pairs targeting amplicons containing 40 SNV and 7 germline variants, and the variant allele ratios were determined by targeted deep sequencing. **a**, Mutation clusters inferred by the PyClone model from bulk population measurements. **b**, Bayesian phylogenetic tree derived from multi-locus genotypes of individual nuclei. The tumour and xenograft nuclei group in distinct clades. **c**, Heat map depicts the multi-locus variant allele prevalences

(blue/yellow/red corresponds to wild-type/heterozygous/homozygous loci, respectively) at variant positions (horizontal axis) in individual nuclei (vertical axis, ordered by phylogenetic grouping in **b**). Upper two blocks show gDNA controls and normal cell nuclei present in tumour samples. The PyClone mutation cluster corresponding to each SNV is indicated by colour in the lowermost horizontal bar. **d**, Consensus genotypes derived from high-probability splits in the phylogenetic tree confirm a set of high prevalence tumour-specific and xenograft-specific mutations, consistent with the expansion of a minor originating clone to dominance in the xenograft, as well as mutations shared in tumour and xenograft nuclei.



Extended Data Figure 4 | Clonal dynamics are reproduced in replicate transplants. **a–c,** Upper panels, passaging history of SA532, SA429 and SA496, showing transplants that resulted in successful xenografts. The transplant sites (blue, subcutaneous; red, subrenal; green, mammary fat pad; all subcutaneous for SA532) and host mouse strains (blue, NSG; orange, NRG; all NSG for SA429 and SA496) are shown. Boxed nodes indicate passages analysed by WGSS. Lower panels, change in cellular prevalence of mutation

clusters over individual transplants. Plots correspond to passages in upper panels. The clusters are inferred by PyClone using grouped data from all passages and correspond to those displayed in Fig. 1. Arrows in SA429 and SA532 show examples of parallel clonal dynamics of the same mutation cluster in multiple replicate transplants. SA496 exhibits less replicated evolution compared with other cases.



Extended Data Figure 5 | Correlation of clonal dynamics in replicate transplants of SA429, SA501, SA532, SA496 and SA535. **a–e**, Scatter plots display the inferred mutation cellular prevalence of all SNVs in pairs of same-passage replicates, for cases SA429, SA501, SA532, SA496 and SA535, respectively. For each replicate, prevalences are inferred by a separate PyClone

analysis that excludes data from other same-passage transplants. Colours indicate mutation clusters inferred in each individual PyClone analyses; the SNVs clustered and colours assigned may differ in each plot. The Pearson correlation coefficients are shown, indicating closely related evolution in most pairs.