Copy-number signatures and mutational processes in ovarian carcinoma

Geoff Macintyre[1 †], Teodora E. Goranova[1 †], Dilrini De Silva[1], Darren Ennis[2], Anna M. Piskorz[1], Matthew Eldridge[1], Daoud Sie[3], Liz-Anne Lewsley[4], Aishah Hanif[4], Cheryl Wilson[4], Suzanne Dowson[2], Rosalind M. Glasspool[5], Michelle Lockley[6,7], Elly Brockbank[6], Ana Montes[8], Axel Walther[9], Sudha Sundar[10], Richard Edmondson[11], Geoff D. Hall[12], Andrew Clamp[13], Charlie Gourley[14], Marcia Hall[15], Christina Fotopoulou[16], Hani Gabra[16], James Paul[4], Anna Supernat[1], David Millan[17], Aoisha Hoyle[17], Gareth Bryson[17], Craig Nourse[2], Laura Mincarelli[2], Luis Navarro Sanchez[2], Bauke Ylstra[3], Mercedes Jimenez-Linan[18], Luiza Moore[18], Oliver Hofmann[2], Florian Markowetz[1]*, Iain A. McNeish[2,5]*, James D. Brenton[1,18,]*

1. Cancer Research UK Cambridge Institute, University of Cambridge, CB2 0RE, UK
2. Institute of Cancer Sciences, University of Glasgow, G61 1QH, UK
3. VU University Medical Centre, Amsterdam 1007 MB, The Netherlands
4. Cancer Research UK Clinical Trials Unit, Institute of Cancer Sciences, University of Glasgow, G12 0YN, UK
5. Beatson West of Scotland Cancer Centre, Glasgow, G12 0YN, UK
6. Barts Cancer Institute, London, EC1M 6BQ, UK
7. University College London Hospital, London, WC1E 6BD, UK
8. Guy's Hospital, London, SE1 9RT, UK
9. Bristol Haematology and Oncology Centre, Bristol, BS2 8ED, UK
10. City Hospital, Birmingham, B18 7QH, UK
11. St Mary's Hospital, Manchester, M13 9WL, UK
12. St James Hospital, Leeds, LS9 7TF, UK
13. The Christie Hospital, Manchester, M20 4BX, UK
14. Edinburgh Cancer Research Centre, Edinburgh, EH4 2XR, UK
15. Mount Vernon Cancer Centre, Northwood, HA6 2RN, UK
16. Imperial College, London, W12 0HS, UK
17. Queen Elizabeth University Hospital, Glasgow G51 4TF, UK
18. Addenbrooke's Hospital, Cambridge, CB2 0QQ, UK.

[†] These authors contributed equally to this work.
* Co-corresponding authors: Florian Markowetz (Florian.Markowetz@cruk.cam.ac.uk), Iain McNeish (iain.mcneish@glasgow.ac.uk), James Brenton (James.Brenton@cruk.cam.ac.uk)

Tumours with profound copy-number aberration elude molecular stratification due to their genomic complexity. By representing this complexity as a mixture of copy-number signatures, we provide molecular explanations for differing clinical outcomes. Here we present a method for copy-number signature identification, deriving eight signatures in 117 shallow whole-genome sequenced high-grade serous ovarian cancers (HGSOC), which validated on independent cohorts of 95 deep whole-genome sequenced, and 402 SNP array-profiled cases. Three copy-number signatures predicted longer overall survival, while the others predicted poorer outcome. We found evidence for the mutational processes giving rise to copy-number change for six of the eight signatures via correlations with other genomic features. Our results provide insights into the pathogenesis of HGSOC by uncovering multiple mutational processes that shape genomes following *TP53* mutation. Importantly, our work shows that most HGSOC have a mixture of mutational processes suggesting that targeting a single mutator phenotype may be therapeutically suboptimal.

Introduction

The discrete mutational processes that drive copy-number change in human cancers are not readily identifiable from genome-wide sequence data. This presents a major challenge for the development of precision medicine for high-grade serous ovarian (HGSOC), oesophageal, small-cell lung and triple negative breast cancers, which are strongly dominated by copy number changes (Ciriello et al., 2013). These tumours have low frequency of recurrent oncogenic mutations, few recurrent copy number alterations and highly complex genomic profiles (Hoadley et al., 2014).

HGSOCs are poor prognosis carcinomas with ubiquitous *TP53* mutations (Ahmed et al., 2010). Despite efforts to develop subtype classification approaches, overall survival has not improved over two decades (Vaughan et al., 2011). Current genomic stratification is limited to defining homologous recombination-deficient (HRD) tumours (Fong et al., 2010; Gelmon et al., 2011; Swisher et al., 2017), and gene expression classifiers identified by large-scale consortia have not yet demonstrated therapeutic utility (Etemadmoghadam et al., 2009; Verhaak et al., 2013). Deep whole genome sequencing has revealed gene breakage events targeting *RB1, NF1* and *PTEN,* but has not provided further insights into the underlying mutational processes driving HGSOC (Patch et al., 2015). Integrated molecular classification has recapitulated the separation of HRD tumours via the enrichment of amplification associated fold-back inversions in non-HRD tumours (Wang et al., 2017).

Recent algorithmic advances have facilitated interpretation of tumour genome complexity via the identification of *mutational signatures* - patterns that reflect the mutational processes active during the life of a tumour (Alexandrov et al., 2013). For example, signatures encoded by single nucleotide variants (SNVs) can represent mutations caused by UV exposure or mismatch repair defects (Alexandrov et al., 2013), whilst signatures encoded by structural variants (SVs) can represent different types of HRD (Nik-Zainal et al., 2016). Importantly, these studies show that tumours typically have multiple active mutational processes. Quantification of the exposure of a tumour to a particular signature may provide a rational framework to personalise therapy (Davies et al., 2017). However, accurate quantification currently requires costly whole-genome sequencing for SVs or exome sequencing for SNVs. By contrast, copy-number changes can be interrogated in a cost-effective manner using shallow whole-genome sequencing (Macintyre et al., 2016). Thus, we have developed a method for the identification of copy-number signatures from these data to test the hypothesis that copy-number signatures reflect underlying mutational processes.

Results

Identification and validation of copy-number signatures

To identify copy-number (CN) signatures, we sequenced 300 primary and relapsed HGSOC samples of 142 patients from the BriTROC-1 cohort (Goranova et al., 2017) (Figure 1A). We combined low-cost shallow whole-genome sequencing (sWGS; 0.1×) with targeted amplicon sequencing of *TP53* to generate absolute copy-number profiles. High-quality copy-number profiles from 117 patients were used for copy-number signature identification (Figure 1B).

For each sample, we computed the genome-wide distributions of six fundamental CN features: the number of breakpoints per 10MB, the copy-number of segments, the size of segments, the difference in CN between adjacent segments, the distance of breakpoints from the centromere, and the lengths of oscillating CN segment chains. The selection of these features was motivated by hallmarks of previously reported genomic aberrations, including breakage-fusion-bridge cycles (Murnane, 2012), chromothripsis (Korbel and Campbell, 2013) and tandem duplication (Menghi et al., 2016; Ng et al., 2012). We summarised each feature using measures of the shape of its distribution: mean, variance, skewness (asymmetry), kurtosis (weight of tails) and modality (number of peaks) resulting in each sample being described by 30 variables.

To identify copy-number signatures from these data, we used non-negative matrix factorisation (NMF) (Lee et al., 2017), a method previously used for analysing SNV signatures (Alexandrov et al., 2013). NMF identified eight CN signatures (Figure 1B), as well as their defining features and their exposures in each sample. The optimal number of signatures was chosen using a consensus from 1000 initialisations of the algorithm and 1000 random permutations of the data combining four model selection measures (Figure 2). We found highly similar signature variable weights in two independent cohorts of 95 whole-genome sequenced HGSOC samples from ICGC (Campbell et al., 2017) and 402 SNP array profiled HGSOC samples (TCGA, 2011), thus demonstrating the robustness of our approach (Figure 1B, P<0.005, median r^2=0.85. Table S1).

Using copy-number signatures to predict overall survival

Using a combined dataset of 602 samples with full clinical annotation, we explored the association between signature exposures and overall survival (Figure 3A). In a multivariate Cox proportional hazards model trained on 356 cases and tested on the remaining 246, CN signatures 1, 7 and 8 had a positive influence on survival, while the remaining five CN signatures had negative influences (Training: P=0.003, log-rank test; stratified by age and cohort; Test: P=0.01, C-index=0.57, 95% CI:0.51-0.62). Significant separation was observed when patients were stratified into good and poor prognosis groups by their dominant signature exposure (Figure S1; Training: P=0.0007, log-rank test; Test: P=0.0003, log-rank test).

Linking copy-number signatures with underlying mutational processes

Examination of cases with a predominant single signature showed that CN signatures 1, 7 and 8 were characterised by high genome complexity (Figure 3B), with CN signature 7 reminiscent of a tandem duplicator phenotype (Ng et al., 2012). However, the vast majority of cases exhibited multiple signature exposures indicative of complex phenotypes shaped by several mutational processes (Figure 3A). Because our approach breaks down this complexity into its basic constituents, we could begin to characterize the individual contributing mutational processes.

To link individual CN signatures to mutational processes, we used the weights identified by NMF to determine which pattern of global or local copy-number change defined each signature. For example, in CN signature 3, the highest weights were observed for the skewness and kurtosis variables of the breakpoint distance from the centromere feature (Figure 4). This suggested frequent breakpoints close to the end of the chromosomes, which are characteristic of peri-telomeric breaks from breakage-fusion-bridge (BFB) events (Murnane, 2012). To test this hypothesis, we correlated CN signature 3 exposures with other genomic features across samples that underwent deep WGS (Figure 5, Figure S2, Figure S3). CN signature 3 was negatively correlated with telomere length (r^2=-0.21, P=0.08, multiple testing corrected) and positively correlated with age-associated SNV signature 1 (r^2=0.35, P=0.0004), consistent with BFB events. In addition, CN signature 3 was positively correlated with amplification-associated fold-back inversion structural variants (r^2=0.34, P=0.07) that have been associated with inferior survival in HGSOC (Wang et al., 2017) and have been strongly implicated in BFB events (Zakov et al., 2013). Thus, we believe that BFB is the underlying mechanism for CN signature 3.

We systematically applied the same approach to the remaining signatures to identify other statistically significant genomic associations using a false discovery rate <0.1 (Figure 5, Figure 6, Figure S2, Figure S3). For two CN signatures, 1 and 6, we did not find evidence for specific underlying mechanisms. For three signatures, we identified significant associations with putative canonical signalling pathways, which can inform future mechanistic studies. CN signature 2 was associated with mutations in genes involved in histone deacetylation (P=0.007, one-sided t-test) and lengthening of telomeres (r^2=0.47, P=3e-05); histone deacetylase (HDAC) activity is closely linked to DNA damage responses (Cea et al., 2016), specifically via ATM regulation (Thurn et al., 2013), whilst inhibition of HDACs can induce profound DNA damage (Conti et al., 2010) with upregulation of error-prone non-homologous end-joining (Smith et al., 2014). CN signature 4 was enriched in cases with oncogenic MAPK signalling involving *NF1* loss and mutated *KRAS* (P=0.003, one-sided t-test)*,* which has previously been shown to induce chromosomal instability as a result of aberrant

G2 and mitotic checkpoint controls and missegregation (Knauf et al., 2006; Saavedra et al., 1999). CN signature 7 was enriched in patients with high tandem duplicator phenotype scores (r^2=0.42, P=1e-04) and mutations in the Toll-like receptor (TLR) pathway (P=0.06, one-sided t-test). TLR4 activity regulates expression of Ku70 (Wang et al., 2013), whilst persistent TLR signalling leads to mitotic defects and potent DNA damage responses (Herrtwich et al., 2016), suggesting possible links between TLR pathway mutations and CN abnormalities.

For the final two signatures, 5 and 8, we found strong evidence for the underlying mutational processes. CN signature 5 was found at significantly higher exposures (P=1.7e-16, one-sided t-test) in tumours with aberrant cell cycle control, including either amplification of *CCNE1, CCND1, CDK4* or *MYC*, or deletion/inactivation of *RB1*. In addition, signature 5 was enriched in tumours with activated PI3K/AKT signalling (P=1.9e-07, one-sided t-test) through mutation of *PIK3CA* or amplification of *EGFR, MET, FGFR3* and *ERBB2*, suggesting that the underlying mechanism for this signature is failure of cell cycle control with co-existent oncogenic signalling. Exposure to CN signature 5 was positively correlated with age at diagnosis (r^2=0.33, P=3e-13) and age-related SNV signature 1 (Alexandrov et al., 2013) (r^2=0.33, P=5e-04). CN signature 5 was characterised by large values in the copy-number change-point distribution, which suggests that the signature represents a single, late whole-genome duplication event (Zack et al., 2013). By contrast, CN signature 3 is likely to represent an ongoing process of localised copy-number changes accumulating over multiple cell divisions.

CN signatures 5 and 8 had mutually exclusive associations with mutated driver genes and other genomic features (Figure 5). Signature 8 had significantly higher exposures in cases with mutations in *BRCA1/2* or other homology-mediated repair (HR) pathway genes (P=1.4e-05, one-sided t-test) and was positively correlated with SNV signature 3 (associated with *BRCA1/2* mutations, r^2=0.53, P=4e-10). It was also negatively correlated both with age at diagnosis (r^2=-0.23, P=8e-07) and the age-related SNV signature 1 (r^2=-0.57, P=7e-12), suggesting early age of disease onset. Together, these results are consistent with HR deficiency being the underlying mechanism for signature 8, and confirm previous data showing mutual exclusivity of *CCNE1* amplification and defective HR (Etemadmoghadam et al., 2013), as well as the poor prognosis conferred by cyclin E amplification (Etemadmoghadam et al., 2009).

Discussion

Overall, the CN signatures provide a framework that is able to rederive the major defining elements of HGSOC genomes, including defective HR (TCGA, 2011), amplification of cyclin E (Etemadmoghadam et al., 2009), and amplification-associated fold-back inversions (Wang et al., 2017). We derived CN signatures using inexpensive shallow whole genome sequencing of DNA from core biopsies, providing a rapid path to clinical implementation. The CN signatures open new avenues for clinical trial design by stratifying patients based upon contributions from BFB, tandem duplicator phenotype, NF1/KRAS and PI3K/AKT signalling.

A major finding is that almost all patients with HGSOC demonstrated a mixture of signatures indicative of combinations of mutational processes. These results suggest that early *TP53* mutation, the ubiquitous initiating event in HGSOC, may permit multiple mutational processes to evolve simultaneously. Even in the context of an additional early driver event such as *BRCA2* mutation in germline carriers, a diverse and variable number of CN signatures is possible (Figure S4). Exposure to multiple signatures could alter the risk of developing therapeutic resistance, which challenges current therapeutic strategies that target a single process such as HRD.

Future studies on larger sample collections are needed to refine CN signature definitions and interpretation. Application to other tumour types is likely to extend the set of signatures beyond the robust core set identified here. Basal-like breast cancers and lung squamous cell carcinoma, which also have high rates of *TP53* mutation and genomic instability (Hoadley et al., 2014), are promising next targets for our approach. Other limitations are technical: we integrated data from three sources, using three different pre-processing pipelines, and the ploidy determined by different pipelines can have a significant effect on the derived signatures. For example, high-ploidy CN signature 5 was predominantly found in the sequenced samples that underwent careful manual curation to identify whole-genome duplication events. When extending to larger sample sets, a unified processing strategy with correct ploidy determination is likely to produce improved signature definitions.

In summary, our results represent a substantial advance in deconvoluting the profound genomic complexity of HGSOC, allowing robust molecular classification of a disease that lacks classic driver oncogene mutations or recurrent copy-number changes. By dissecting the mutational forces shaping highly aberrant copy-number, our study paves the way to understanding and categorising extreme genomic complexity.

Author contributions

Conceptualisation: GM, TEG, FM, IMcN, JDB; Study conduct: SD, RMG, ML, EB, AM, AW, SS, RE, GDH, AC, CG, MH, CF, HG, DM, AHo, GB, IMcN, JDB; Investigation: TEG, DE, AMP, LAL, AHa, CW, CN, LMi, LNS, MJL, LMo, AS, JP; Formal analysis: GM, TEG, DDS, ME, DS, BY, OH, FM; Methodology and software: GM, DDS, FM; Writing: GM, TEG, DDS, FM, IMcN, JDB

Acknowledgements

The BriTROC-1 study was funded by Ovarian Cancer Action (to IMcN and JDB, grant number 006). We would like to acknowledge funding and support from Cancer Research UK (grant numbers A15973, A15601, A18072, A17197 and A19274), the Universities of Cambridge and Glasgow, National Institute for Health Research Cambridge Biomedical Research Centre, National Cancer Research Network, the Experimental Cancer Medicine Centres at participating sites, the Beatson Endowment Fund and Hutchison Whampoa Limited. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. We thank the Biorepository, Bioinformatics, Histopathology and Genomics Core Facilities of the Cancer Research UK Cambridge Institute and the Pathology Core at the Cancer Research UK Beatson Institute for technical support. We would like to thank members of PCAWG Evolution and Heterogeneity Working Group for the consensus copy-number analysis, PCAWG Structural Variation Working Group for the consensus structural variants and PCAWG Technical Working Group for annotating driver mutations in the 95 ICGC samples.

References

Ahmed, A.A., Etemadmoghadam, D., Temple, J., Lynch, A.G., Riad, M., Sharma, R., Stewart, C., Fereday, S., Caldas, C., Defazio, A.*, et al.* (2010). Driver mutations in TP53 are ubiquitous in high grade serous carcinoma of the ovary. J Pathol *221*, 49-56.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Borresen-Dale, A.L.*, et al.* (2013). Signatures of mutational processes in human cancer. Nature *500*, 415-421.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological) *57*, 289-300.

Campbell, P.J., Getz, G., Stuart, J.M., Korbel, J.O., Stein, L.D., and ICGC/TCGA (2017). Pan-cancer analysis of whole genomes. In bioRxiv.

Carter, S.L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P.W., Onofrio, R.C., Winckler, W., Weir, B.A.*, et al.* (2012). Absolute quantification of somatic DNA alterations in human cancer. Nature biotechnology *30*, 413-421.

Cea, M., Cagnetta, A., Adamia, S., Acharya, C., Tai, Y.T., Fulciniti, M., Ohguchi, H., Munshi, A., Acharya, P., Bhasin, M.K.*, et al.* (2016). Evidence for a role of the histone deacetylase SIRT6 in DNA damage response of multiple myeloma cells. Blood *127*, 1138-1150.

Ciriello, G., Miller, M.L., Aksoy, B.A., Senbabaoglu, Y., Schultz, N., and Sander, C. (2013). Emerging landscape of oncogenic signatures across human cancers. Nature genetics *45*, 1127-1133.

Conti, C., Leo, E., Eichler, G.S., Sordet, O., Martin, M.M., Fan, A., Aladjem, M.I., and Pommier, Y. (2010). Inhibition of histone deacetylase in cancer cells slows down replication forks, activates dormant origins, and induces DNA damage. Cancer Res *70*, 4470-4480.

Davies, H., Glodzik, D., Morganella, S., Yates, L.R., Staaf, J., Zou, X., Ramakrishna, M., Martin, S., Boyault, S., Sieuwerts, A.M.*, et al.* (2017). HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. Nature medicine *23*, 517-525.

Etemadmoghadam, D., Au-Yeung, G., Wall, M., Mitchell, C., Kansara, M., Loehrer, E., Batzios, C., George, J., Ftouni, S., Weir, B.A.*, et al.* (2013). Resistance to CDK2 inhibitors is associated with selection of polyploid cells in CCNE1-amplified ovarian cancer. Clinical cancer research : an official journal of the American Association for Cancer Research *19*, 5960-5971.

Etemadmoghadam, D., deFazio, A., Beroukhim, R., Mermel, C., George, J., Getz, G., Tothill, R., Okamoto, A., Raeder, M.B., Harnett, P.*, et al.* (2009). Integrated genome-wide DNA copy number and expression analysis identifies distinct mechanisms of primary chemoresistance in ovarian carcinomas. Clin Cancer Res *15*, 1417-1427.

Farmery, J.H.S., Mike L; Lynch Andy G (2017). Telomerecat: A Ploidy-Agnostic Method For Estimating Telomere Length From Whole Genome Sequencing Data. In bioRxiv.

Fong, P.C., Yap, T.A., Boss, D.S., Carden, C.P., Mergui-Roelvink, M., Gourley, C., De Greve, J., Lubinski, J., Shanley, S., Messiou, C.*, et al.* (2010). Poly(ADP)-Ribose Polymerase Inhibition: Frequent Durable Responses in BRCA Carrier Ovarian Cancer Correlating With Platinum-Free Interval. J Clin Oncol *28*, 2512-2519.

Gaujoux, R., and Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. BMC Bioinformatics *11*, 367.

Gehring, J.S., Fischer, B., Lawrence, M., and Huber, W. (2015). SomaticSignatures: inferring mutational signatures from single-nucleotide variants. Bioinformatics *31*, 3673-3675.

Gelmon, K.A., Tischkowitz, M., Mackay, H., Swenerton, K., Robidoux, A., Tonkin, K., Hirte, H., Huntsman, D., Clemons, M., Gilks, B.*, et al.* (2011). Olaparib in patients with recurrent high-grade serous or poorly differentiated ovarian carcinoma or triple-negative breast cancer: a phase 2, multicentre, open-label, non-randomised study. Lancet Oncol *12*, 852-861.

Gonzalez-Perez, A., Perez-Llamas, C., Deu-Pons, J., Tamborero, D., Schroeder, M.P., Jene-Sanz, A., Santos, A., and Lopez-Bigas, N. (2013). IntOGen-mutations identifies cancer drivers across tumor types. Nat Methods *10*, 1081-1082.

Goranova, T., Ennis, D., Piskorz, A.M., Macintyre, G., Lewsley, L.A., Stobo, J., Wilson, C., Kay, D., Glasspool, R.M., Lockley, M.*, et al.* (2017). Safety and utility of image-guided research biopsies in relapsed high-grade serous ovarian carcinoma-experience of the BriTROC consortium. British journal of cancer *116*, 1294-1301.

Harrell, F.E. (2016). Hmisc: Harrell Miscellaneous. R package version 4.0-0.

Hartigan, J.A., and Hartigan, P.M. (1985). The Dip Test of Unimodality. Ann Statist *13*, 70-84.

Herrtwich, L., Nanda, I., Evangelou, K., Nikolova, T., Horn, V., Sagar, Erny, D., Stefanowski, J., Rogell, L., Klein, C.*, et al.* (2016). DNA Damage Signaling Instructs Polyploid Macrophage Fate in Granulomas. Cell *167*, 1264-1280.e1218.

Hoadley, K.A., Yau, C., Wolf, D.M., Cherniack, A.D., Tamborero, D., Ng, S., Leiserson, M.D., Niu, B., McLellan, M.D., Uzunangelov, V.*, et al.* (2014). Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. Cell *158*, 929-944.

Huebschmann, D., Gu, Z., and Schlesner, M. (2015). YAPSA: Yet Another Package for Signature Analysis. R package version 1.2.0.

Jones, D., Raine, K.M., Davies, H., Tarpey, P.S., Butler, A.P., Teague, J.W., Nik-Zainal, S., and Campbell, P.J. (2016). cgpCaVEManWrapper: Simple Execution of CaVEMan in Order to Detect Somatic Single Nucleotide Variants in NGS Data. Curr Protoc Bioinformatics *56*, 15 10 11-15 10 18.

Kim, S. (2015). ppcor: Partial and Semi-Partial (Part) Correlation.

Knauf, J.A., Ouyang, B., Knudsen, E.S., Fukasawa, K., Babcock, G., and Fagin, J.A. (2006). Oncogenic RAS induces accelerated transition through G2/M and promotes defects in the G2 DNA damage and mitotic spindle checkpoints. The Journal of biological chemistry *281*, 3800-3809.

Korbel, J.O., and Campbell, P.J. (2013). Criteria for inference of chromothripsis in cancer genomes. Cell *152*, 1226-1236.

Lee, M., Napier, C.E., Yang, S.F., Arthur, J.W., Reddel, R.R., and Pickett, H.A. (2017). Comparative analysis of whole genome sequencing-based telomere length measurement techniques. Methods (San Diego, Calif) *114*, 4-15.

Macintyre, G., Ylstra, B., and Brenton, J.D. (2016). Sequencing Structural Variants in Cancer for Precision Therapeutics. Trends Genet *32*, 530-542.

Menghi, F., Inaki, K., Woo, X., Kumar, P.A., Grzeda, K.R., Malhotra, A., Yadav, V., Kim, H., Marquez, E.J., Ucar, D.*, et al.* (2016). The tandem duplicator phenotype as a distinct genomic configuration in cancer. Proceedings of the National Academy of Sciences of the United States of America *113*, E2373-2382.

Murnane, J.P. (2012). Telomere dysfunction and chromosome instability. Mutation research *730*, 28-36.

Ng, C.K., Cooke, S.L., Howe, K., Newman, S., Xian, J., Temple, J., Batty, E.M., Pole, J.C., Langdon, S.P., Edwards, P.A.*, et al.* (2012). The role of tandem duplicator phenotype in tumour evolution in high-grade serous ovarian cancer. J Pathol *226*, 703-712.

Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorena, I., Alexandrov, L.B., Martin, S., Wedge, D.C.*, et al.* (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature *534*, 47-54.

Patch, A.-M., Christie, E.L., Etemadmoghadam, D., Garsed, D.W., George, J., Fereday, S., Nones, K., Cowin, P., Alsop, K., Bailey, P.J.*, et al.* (2015). Whole–genome characterization of chemoresistant ovarian cancer. Nature *521*, 489-494.

Piskorz, A.M., Ennis, D., Macintyre, G., Goranova, T.E., Eldridge, M., Segui-Gracia, N., Valganon, M., Hoyle, A., Orange, C., Moore, L.*, et al.* (2016). Methanol-based fixation is superior to buffered formalin for next-generation sequencing of DNA from clinical cancer samples. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO

27, 532-539.

Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B.S., and Swanton, C. (2016). DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol *17*, 31.

Saavedra, H.I., Fukasawa, K., Conn, C.W., and Stambrook, P.J. (1999). MAPK mediates RAS-induced chromosome instability. The Journal of biological chemistry *274*, 38083-38090.

Scheinin, I., Sie, D., Bengtsson, H., van de Wiel, M.A., Olshen, A.B., van Thuijl, H.F., van Essen, H.F., Eijk, P.P., Rustenburg, F., Meijer, G.A.*, et al.* (2014). DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. Genome Res *24*, 2022-2032.

Schumacher, S. (2015). pancan12_absolute.segtab.txt.

Secrier, M., Li, X., de Silva, N., Eldridge, M.D., Contino, G., Bornschein, J., MacRae, S., Grehan, N., O'Donovan, M., Miremadi, A.*, et al.* (2016). Mutational signatures in esophageal adenocarcinoma define etiologically distinct subgroups with therapeutic relevance. Nature genetics *48*, 1131-1141.

Smith, S., Fox, J., Mejia, M., Ruangpradit, W., Saberi, A., Kim, S., Choi, Y., Oh, S., Wang, Y., Choi, K.*, et al.* (2014). Histone deacetylase inhibitors selectively target homology dependent DNA repair defective cells and elevate non-homologous endjoining activity. PloS one *9*, e87203.

Swisher, E.M., Lin, K.K., Oza, A.M., Scott, C.L., Giordano, H., Sun, J., Konecny, G.E., Coleman, R.L., Tinker, A.V., O'Malley, D.M.*, et al.* (2017). Rucaparib in relapsed, platinum-sensitive high-grade ovarian carcinoma (ARIEL2 Part 1): an international, multicentre, open-label, phase 2 trial. The Lancet Oncology *18*, 75-87.

Tamborero, D., Rubio-Perez, C., Deu-Pons, J., Schroeder, M.P., Vivancos, A., Rovira, A., Tusquets, I., Albanell, J., Rodon, J., Tabernero, J.*, et al.* (2017). Cancer Genome Interpreter Annotates The Biological And Clinical Relevance Of Tumor Alterations. In bioRxiv.

TCGA (2011). Integrated genomic analyses of ovarian carcinoma. Nature *474*, 609-615.

Therneau, T.M., Grambsch, Patricia M. (2000). Modeling Survival Data: Extending the Cox Model (New York: Springer).

Thurn, K.T., Thomas, S., Raha, P., Qureshi, I., and Munster, P.N. (2013). Histone deacetylase regulation of ATM-mediated DNA damage signaling. Molecular cancer therapeutics *12*, 2078-2087.

Vaughan, S., Coward, J.I., Bast, R.C., Jr., Berchuck, A., Berek, J.S., Brenton, J.D., Coukos, G., Crum, C.C., Drapkin, R., Etemadmoghadam, D.*, et al.* (2011). Rethinking ovarian cancer: recommendations for improving outcomes. Nat Rev Cancer *11*, 719-725.

Verhaak, R.G., Tamayo, P., Yang, J.Y., Hubbard, D., Zhang, H., Creighton, C.J., Fereday, S., Lawrence, M., Carter, S.L., Mermel, C.H.*, et al.* (2013). Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. The Journal of clinical investigation *123*, 517-525.

Wang, Y.K., Bashashati, A., Anglesio, M.S., Cochrane, D.R., Grewal, D.S., Ha, G., McPherson, A., Horlings, H.M., Senz, J., Prentice, L.M.*, et al.* (2017). Genomic consequences of aberrant DNA repair mechanisms stratify ovarian cancer histotypes. Nature genetics *49*, 856-865.

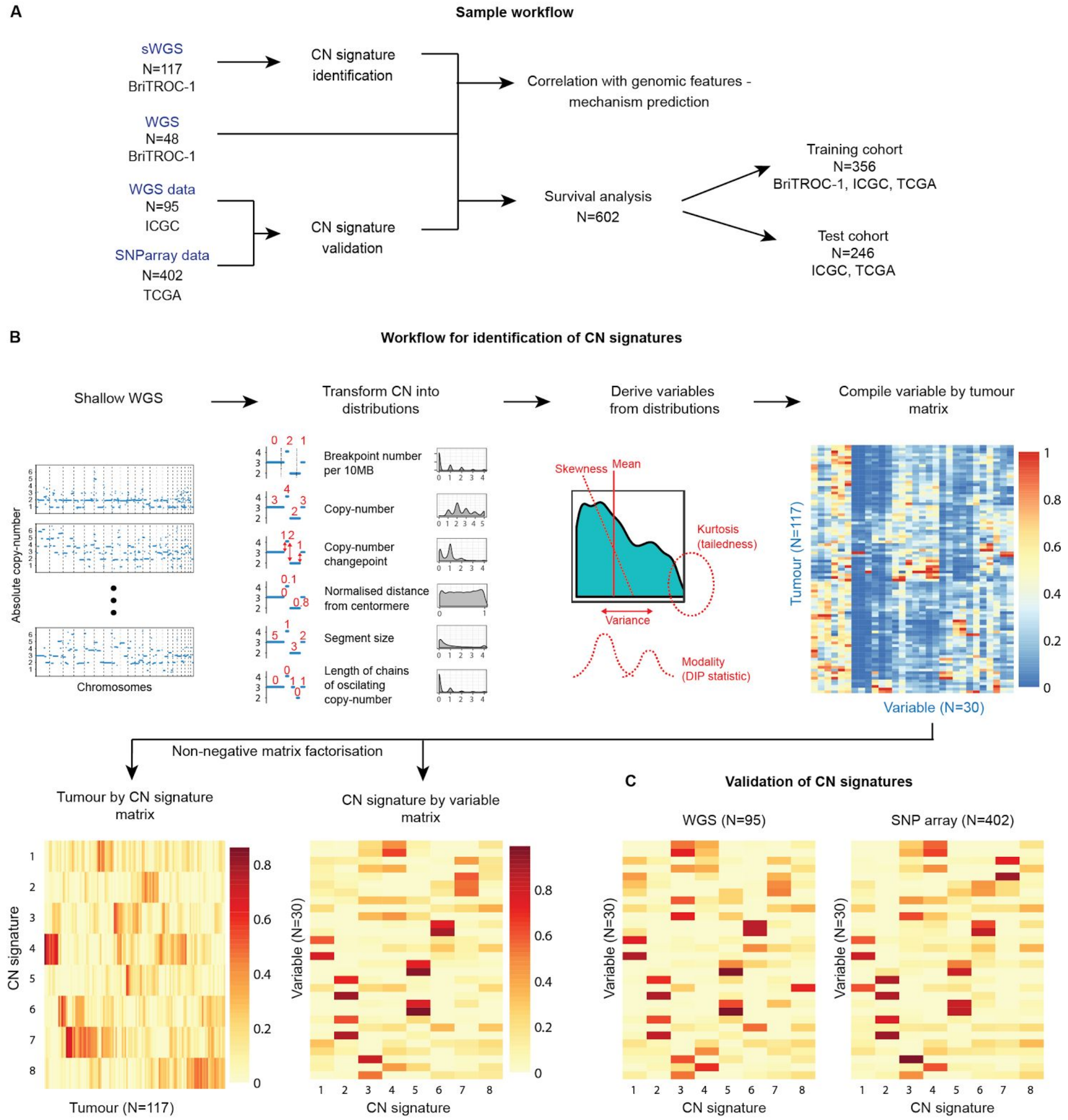
Wang, Z., Yan, J., Lin, H., Hua, F., Wang, X., Liu, H., Lv, X., Yu, J., Mi, S., Wang, J.*, et al.* (2013). Toll-like receptor 4 activity protects against hepatocellular tumorigenesis and progression by regulating expression of DNA repair protein Ku70 in mice. Hepatology (Baltimore, Md) *57*, 1869-1881.

Yu, G., and He, Q.Y. (2016). ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol Biosyst *12*, 477-479.

Zack, T.I., Schumacher, S.E., Carter, S.L., Cherniack, A.D., Saksena, G., Tabak, B., Lawrence, M.S., Zhsng, C.Z., Wala, J., Mermel, C.H.*, et al.* (2013). Pan-cancer patterns of somatic copy number alteration. Nature genetics *45*, 1134-1140.

Zakov, S., Kinsella, M., and Bafna, V. (2013). An algorithmic approach for breakage-fusion-bridge detection in tumor genomes. Proceedings of the National Academy of Sciences of the United States of America *110*, 5546-5551.

Figure 1 | Copy-number signature identification - sample and analysis workflows

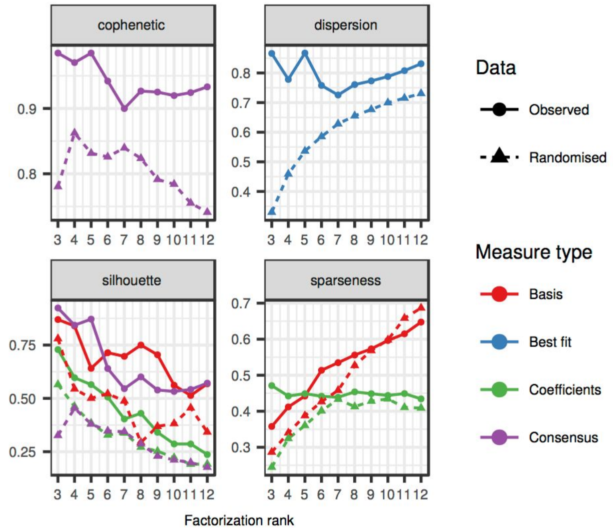


A schematic showing the HGSOC samples appearing in the manuscript along with the technology used to interrogate copy number. sWGS was performed on 300 samples from 142 patients BriTROC-1 patients. 117 high and intermediate quality samples were identified, maximum one sample per patient. CN signatures were initially derived using 91 high quality samples and subsequently applied to the remaining 26 samples of intermediate quality. 48 of those 117 samples were also sequenced by deep WGS and used for correlation analyses. (sWGS = shallow whole genome sequencing, WGS = whole genome sequencing, SNParray = single nucleotide polymorphism hybridization array.)

B A summary of the 5 steps taken to identify copy-number signatures using shallow whole genome sequencing. Step 1: compute absolute copy-numbers from sWGS data; Step 2: compute genome-wide distributions of six fundamental copy-number features; Step 3: summarize the shape of each distribution using the mean, variance, skewness, kurtosis, and modality; Step 4: Represent the data as a matrix with 30 (=6x5) variables per tumour. Step 5: Apply Non-negative matrix factorisation to this matrix to deconvolute it into signature exposures.

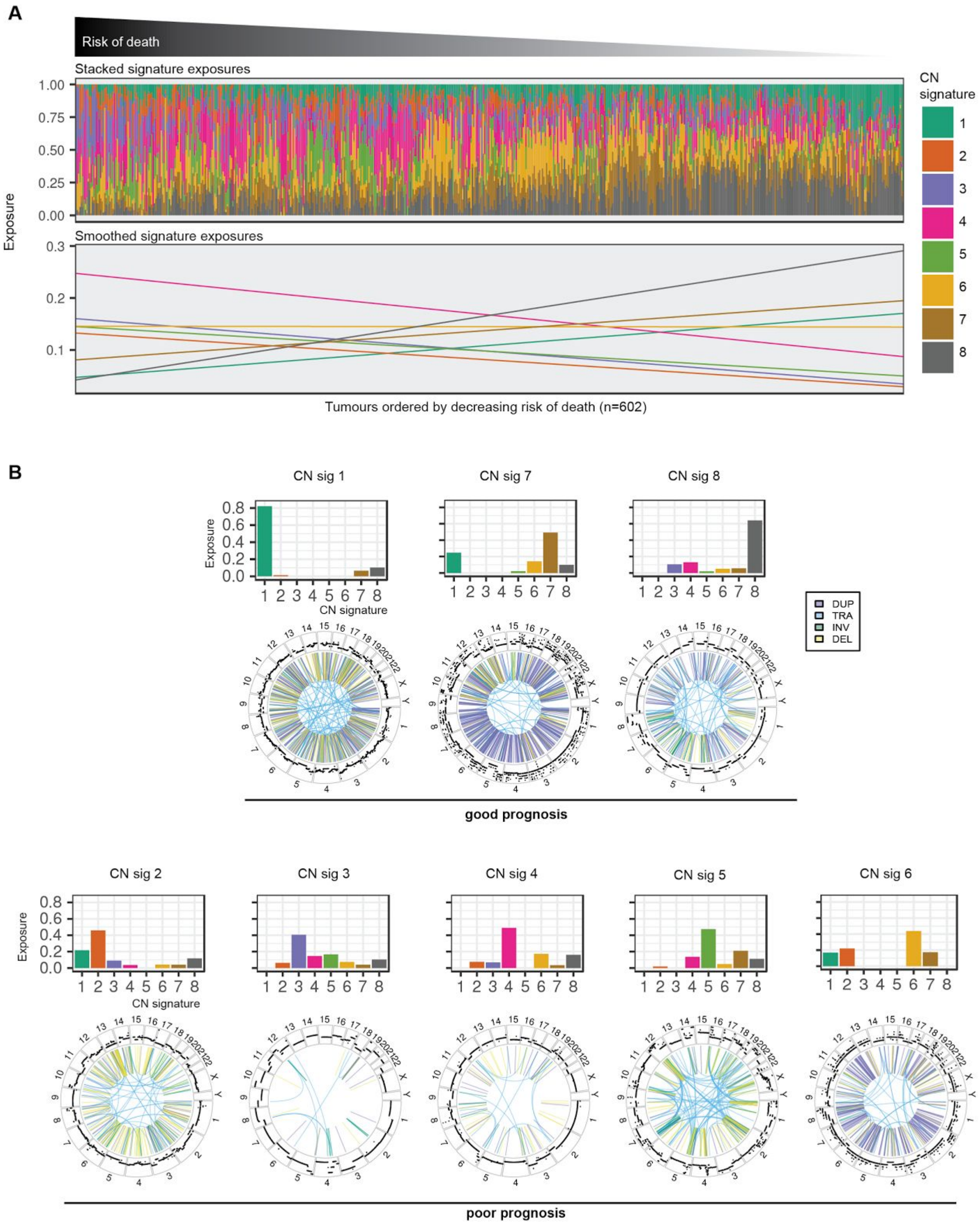
C Validation of the copy number signature identification approach across two independent cohorts of samples profiled using WGS and SNP array. The similarity between these matrices and the sWGS in a can be clearly observed. Correlation coefficients of the similarity can be found in Table S1.

Figure 2 | Measures for selecting optimal signature number



A comparison of signature number (x-axis) across four measures for determining optimal signature number. The circle and solid lines represent the results from the BriTROC-1 samples run whereas the triangles and dotted lines represent results from 1000 randomly permuted BriTROC-1 matrices (these can be considered a null measure). Here, basis refers to the signature-by-variable matrix, coefficients refers to patient-by-signature matrix, and consensus refers to the connectivity matrix of patients clustered by their dominant signature across 1000 runs. Best-fit is the run that showed the lowest objective score across the 1000 runs. A value of 8 defines the point of stability in the cophenetic, dispersion and silhouette coefficients, and is the maximum sparsity achievable above the null model.

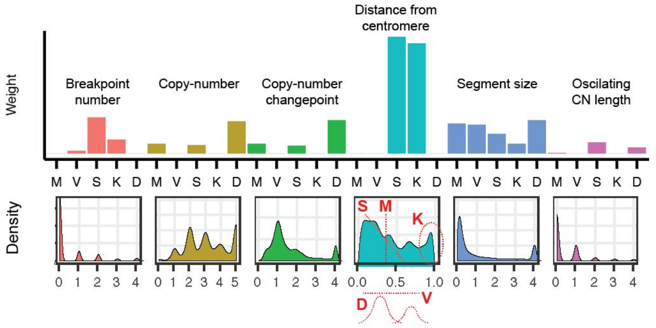
Figure 3 | Copy-number signatures predict survival across 602 patients



A Stacked CN signature exposures in every patient (top) and smoothed exposures independently for every CN signature (bottom). Patients were ranked from short survival (left) to long survival (right) using the risk estimated by a multivariate Cox proportional hazards model stratified by age and cohort, with CN signature exposures as covariates.

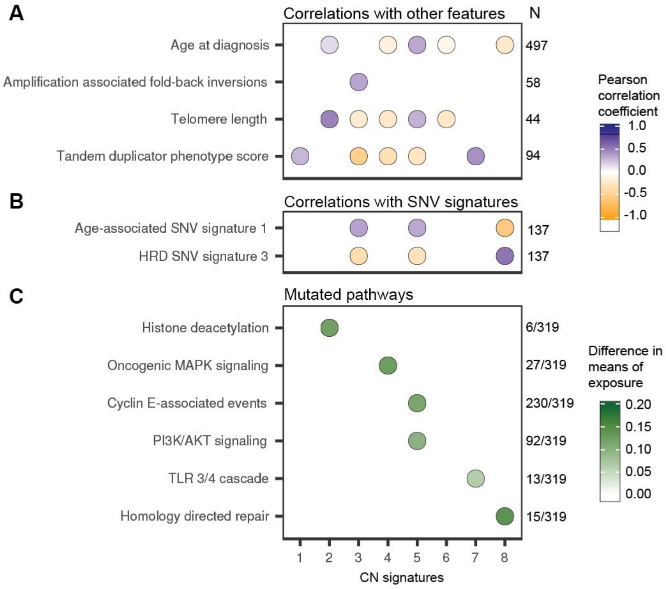
B Examples of genomes with dominant copy-number signatures. Bar plots show the proportion of each CN signature identified in the example case. The outer ring of each circos plot shows copy-number changes. The inner ring shows structural variants: duplications (DUP), translocations (TRA), inversions (INV), and deletions (DEL).

Figure 4 | A strategy to highlight mutational processes underlying CN signatures



The barplot shows the weights of the 30 variables defining copy-number signature 3, which represent the shapes of genome-wide distributions of six fundamental copy-number features (M: Mean, V: Variance, S: Skew, K: Kurtosis, D: DIP statistic). The turquoise distribution is annotated to show which part is highlighted by each shape measure. For example, in CN signature 3 shown here, the highest bars point to the skewness and kurtosis of the distribution of breakpoint distances from the centromere. This indicates that the copy-number changes of importance in genomes shaped by CN signature 3 are those that lie in the tail of the distribution.

Figure 5 | Association of copy-number signatures with mutated driver genes and other features.



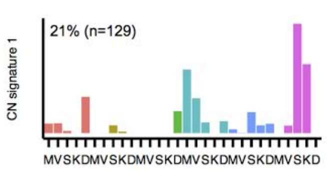
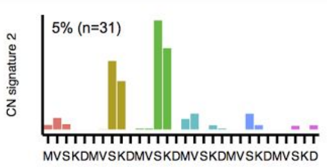
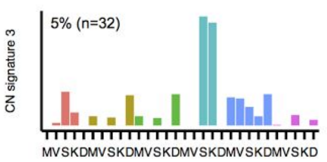
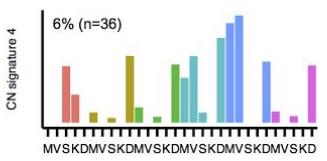
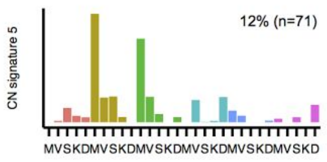
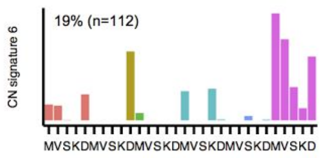
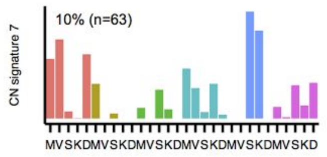
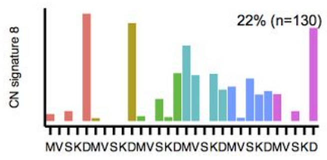
In both **A** and **B**, each dot corresponds to a significant correlation coefficient (P<0.1). The numbers to the right of the plots show the number of cases included in each analysis.

A Significant correlation coefficients between other features and CN signature exposures. Blue indicates positive correlation and orange negative correlation.

B Significant correlation coefficients between SNV signature and CN signature exposures. SNV signatures were taken from the COSMIC database. Blue indicates positive correlation and orange negative correlation.

C Reactome pathways showing significantly different mean signature exposures between mutant and non-mutant cases (P<0.1, one-sided t-test). Colour scale indicates extent of difference. The numbers to the right of the plot show the number of mutated cases for each pathway out of a total 319 cases.

Figure 6 | Copy-number signatures found in HGSOC

CN Signature and prevalence*	Important features	Associations	Proposed mechanism
	• Breaks tending towards chromosome ends • Small number of oscillating CN regions	• High exposure predicts good overall survival • Correlated with tandem duplicator phenotype score	Unknown
	• Mostly single copy-number changes • Occasional large copy-number changes	• High exposure predicts poor overall survival • Correlated with age at diagnosis • Correlated with telomere lengthening • Enriched in cases with driver mutations in histone deacetylation pathway: ARID4B, CHD4, HIST1H3B, HST1H3F, NCOR1, TBL1XR1	Impaired DNA damage response and error-prone non-homologous end-joining
	• Occasional breaks at telomeres	• High exposure predicts poor overall survival • Correlated with number of amplification associated fold-back inversions • Correlated with telomere shortening • Anti-correlated with tandem-duplicator phenotype score • Correlated with age-related SNV signature 1 • Anti-correlated with HRD associated SNV signature 3	Breakage-fusion-bridge
	• Breaks at centromeres and telomeres • Large, chromosome arm sized breaks	• High exposure predicts poor overall survival • Anti-correlated with age at diagnosis • Correlated with telomere shortening • Anti-correlated with tandem-duplicator phenotype score • Enriched in cases with driver mutations in MAPK signaling: NF1, KRAS, MACF1, MAP3K11, RASA1, BRAF, FN1, RASA2	Chromosomal instability, aberrant G2 and mitotic checkpoint controls, and missegregation
	• High copy-number states (ploidy) • Large copy-number change points	• High exposure predicts poor overall survival • Correlated with telomere lengthening • Anti-correlated with tandem-duplicator phenotype score • Correlated with age-related SNV signature 1 • Anti-correlated with HRD associated SNV signature 3 • Enriched in cases with driver mutations in Cyclin-E pathway and PI3K/AKT signaling: MYC, CCNE1, CDK4, RB1, MET, EGFR, FGFR1, FGFR2, ERBB2, FGFR4, PIK3CA, ERBB3, FGFR3, HGF, IRS2, PDGFRA, PTPN11	Failure of cell cycle control with co-existent oncogenic signaling leading to whole-genome duplication
	• Mostly 2 and 3 copy-number states • Mostly single-copy changes • Oscillating copy-number states	• High exposure predicts poor overall survival • Anti-correlated with age at diagnosis • Correlated with telomere shortening	Unknown
	• Large number of breaks • Occasional large segments	• High exposure predicts good overall survival • Correlated with tandem duplicator phenotype score • Enriched in cases with mutations in the TLR3/4 cascade: MAP2K4, PPP2R1A, CASP8, DNM2, IKBKB, IRF7, PTPN11	Mitotic defects, aberrant DNA repair
	• Clusters of breaks between large segments • Diploid with single loss and gain • Mostly oscillating copy-number	• High exposure predicts good overall survival • Anti-correlated with age at diagnosis • Anti-correlated with age-related SNV signature 1 • Correlated with HRD associated SNV signature 3 • Enriched in cases with driver mutations in homology directed repair: BRCA1, BRCA2, HERC2, PALB2, ABL1, ATR, FAM175A	Homologous recombination deficiency

*Percentage of cases with a dominant signature across entire cohort
M=Mean, V=Variance, S=Skewness, K=Kurtosis, D=DIP statistic

Features

🟥 Breakpoint number per 10MB	🟩 Copy-number changepoint	🟦 Segment size
🟨 Copy number	🟦 Normalised distance from centromere	🟪 Length of chains of oscillating copy-number

This figure summarises the CN signatures, their weights, important features and associations highlighting potential underlying mechanisms.

Methods

Patients and samples

The BriTROC-1 study has been described previously (Goranova et al., 2017). The study is sponsored by NHS Greater Glasgow and Clyde and ethics/IRB approval was given by Cambridge Central Research Ethics Committee (Reference 12/EE/0349). The study enrolled patients with recurrent ovarian high-grade serous or grade 3 endometrioid carcinoma who had relapsed following at least one line of platinum-based chemotherapy and whose disease was amenable either to image-guided biopsy or secondary debulking surgery. All patients provided written informed consent. Access to archival diagnostic formalin-fixed tumour was also required.

Tagged-amplicon sequencing

DNA was extracted from relapsed biopsies and archival diagnostic tissue, and mutation screening of *TP53*, *PTEN*, *EGFR*, *PIK3CA*, *KRAS* and *BRAF* was performed using tagged-amplicon sequencing as previously described (Goranova et al., 2017).

Shallow whole genome sequencing (sWGS)

Libraries for sWGS were prepared from 100ng DNA using modified TruSeq Nano DNA LT Sample Prep Kit (Illumina) protocol (Piskorz et al., 2016). Quality and quantity of the libraries were assessed with DNA-7500 kit on 2100 Bioanalyzer (Agilent Technologies) and with Kapa Library Quantification kit (Kapa Biosystems) using to the manufacturer's protocols. Sixteen to twenty barcoded libraries were pooled together in equimolar amounts and each pool was sequenced on HiSeq4000 in SE-50bp mode.

Deep whole genome sequencing

Deep whole-genome sequencing was performed on 56 tumour and matched normal samples, of which 48 passed quality control. Libraries were constructed with ~350-bp insert length using the TruSeq Nano DNA Library prep kit (Illumina) and sequenced on an Illumina HiSeq X Ten System in paired-end 150-bp reads mode. The average depth was 60× (range 40-101×) in tumours and 40× (range 24-73×) in matched blood samples.

Variant calling

on mapping quality, base quality, position in read, and strand bias as described (Secrier et al., 2016). In addition, the blacklisted SNVs from the Sanger Cancer Genomics Project pipeline derived from a panel of unmatched normal samples were used for filtering (Jones et al., 2016).

Data download

ICGC: Consensus SNVs and INDELs (October 2016 release), consensus structural variants (v 1.6), consensus copy-number calls (January 2017 release), donor clinical (August 2016 v7-2) and donor histology information (August 2016 v7) for 95 ovarian cancer samples were downloaded from the PCAWG data portal. ACEseq copy-number calls were used for analysis.

TCGA: ABSOLUTE (Carter et al., 2012) copy-number profiles from Zack et al (Zack et al., 2013) for 402 ovarian cancer TCGA samples were downloaded from Synapse (Schumacher, 2015). SNVs for these samples were downloaded from the Broad Institute TCGA Genome Data Analysis Center (Broad Institute TCGA Genome Data Analysis Center: Firehose stddata__2016_01_28 run. doi:10.7908/C11G0KM9, Broad Institute of MIT and Harvard). Donor clinical data were downloaded from the TCGA data portal.

Absolute copy-number calling from sWGS

After inspection of the TP53 mutation status and relative copy-number profiles of the 300 BriROC-1 sequenced samples, 44 were excluded from downstream analysis for the following reasons: low purity (21), mislabelled (7), pathology re-review revealed sample was not HGSOC (3), no detectable TP53 mutation (13). Of the remaining samples, 57 showed an over segmentation artefact (likely due to fixation). A more strict segmentation was subsequently applied to these samples to yield a usable copy-number profile. Relative copy-number profiles were generated using QDNAseq (Scheinin et al., 2014) and these were combined with TP53 mutant allele frequency identified using tagged amplicon sequencing in a probabilistic graphical modelling approach adapted from ABSOLUTE (Carter et al., 2012) to generate absolute copy-number profiles. Using Expectation-Maximisation, the model generated a posterior over a range of TP53 copy-number states, using the TP53 mutant allele frequency to estimate purity for each state. The TP53 copy-number state which provided the highest likelihood of generating a clonal absolute copy-number profile was used to determine the final absolute copy-number profile. Following absolute copy-number fitting, the samples were rated using a 1-3 star system. 1-star samples (n=54) showed a noisy copy-number profile and were considered likely to have incorrect segments and missing calls. These were excluded from further analysis. 2-star samples (n=52) showed a reasonable copy-number profile with only a small number of miscalled segments. These samples were used (with caution) for some subsequent analyses. 3-star samples (n=147) showed a high-quality copy-number profile that was used in all downstream analyses. The maximum star rating observed per patient was 15 patients 1-star, 26 2-star, and 91 3-star.

Copy-number signature identification

Preprocessing: 91 3-star BriTROC-1 absolute copy-number profiles were summarised using six different feature density distributions (Outlined in Figure 1B): 1. segment size - the length of each genome segment in 10MB units; 2. Breakpoint number per 10MB - the number of genome breaks appearing in 10MB sliding windows across the genome; 3. change-point copy-number - the absolute difference in CN between adjacent segments across the genome; 4. segment copy-number - the observed absolute copy-number state of each segment; 5. normalised distance of breakpoints from centromere - the distance of a break from the centromere divided by the length of the chromosome arm; 6. length of segments with oscillating copy-number - a traversal of the genome counting the number of contiguous CN segments alternating between two copy-number states. Each of these distributions was then reduced to a set of five variables encoding the shape of the distribution: the first four moments (mean, variance, skewness and kurtosis) and a fifth variable, Hartigan's dip test (Hartigan and Hartigan, 1985), which is a measure of the modality of the distribution. As skew can take a negative value, its absolute value was used. All variables were then scaled to lie between (0,1). This approach resulted in a set of 30 variables representing the copy-number profile of each sample.

Signature identification using NMF. The NMF Package in R (Gaujoux and Seoighe, 2010) was used to decompose the patient-by-variable matrix into a patient-by-signature and signature-by-variable matrix. A signature search interval of 3-12 was used, running the NMF 1000 times with different random seeds for each signature number. As provided by the NMF Package in R (Gaujoux and Seoighe, 2010), the cophenetic, dispersion, silhouette, and sparseness coefficients were computed for the signature-by-variable matrix (basis), patient-by-signature matrix (coefficients) and connectivity matrix (consensus, representing patients clustered by their dominant signature across the 1000 runs). 1000 random shuffles of the input matrix were performed to get a null estimate of each of the scores (Figure 2). We sought the minimum signature number that yielded stability in the cophenetic, dispersion and silhouette coefficients, and that yielded the maximum sparsity which could be achieved without exceeding that which was observed in the randomly permuted matrices. 8 signatures was deemed optimal under these constraints and was therefore chosen for the remaining analysis.

Assigning signature exposures to samples: For the remaining 26 2-star patient samples, and the 82 secondary patient samples (from patients with 2- or 3-star profiles from additional tumours), the LCD function in the YAPSA package in Bioconductor (Huebschmann et al., 2015) was used to assign signature exposures.

Copy-number signature validation

The variable summary procedure described above was applied to copy-number profiles from two independent datasets: 95 whole-genome sequenced (approximately 40×) HGSOC samples processed as part of ICGC Pan-Cancer Analysis of Whole Genomes Project (Campbell et al., 2017), and SNParray profiling of HGSOC cases as part of TCGA (Zack et al., 2013). The number of signatures was fixed at 8 for matrix decomposition with NMF. A Pearson correlation coefficient was obtained by correlating the signature by feature matrix from each cohort with one derived from the high-quality sWGS copy-number profiles.

Identification of copy-number signature patterns

The density distributions appearing in Figure S5 were used to assist with determining important patterns associated with each signature. These were generated using a weighted kernel density estimator in R where the copy-number features were weighted by their signature exposures for 117 BriTROC-1 cases. We used these, in combination with the variable weights for each signature (Figure 6), to determine which parts of each distribution were important for defining each signature.

Survival analysis

Overall survival in BriTROC-1 patients was calculated from the date of enrolment to the date of death or the last documented clinical assessment, with data cutoff at 1 December 2016. Given that the BriTROC-1 study only enrolled patients with relapsed disease, it was necessary to left truncate the overall survival time. In addition, cases where the patient was not deceased were right censored. Survival data for the ICGC and TCGA cohorts were right censored as required (left truncation was not necessary). The combined samples were split into training (100% BriTROC-1, 50% ICGC and 50% TCGA = 356) and test (50% ICGC and 50% TCGA = 246) cohorts. All of the BriTROC-1 samples were used in the training set to avoid issues calculating prediction performance on left-truncated data. A Cox proportional hazards model was fitted on the training set, with the signature exposures as covariates, stratified by cohort (BriTROC-1, OV-AU, OV-US, TCGA) and age (<51; 52-59; 60-69; >70), using the survival package in Bioconductor (Therneau, 2000). As the signature exposures summed to 1, signature 2, which demonstrated the lowest variance across patients was excluded. After fitting, the model was used to predict risk in the test set and performance was assessed using the concordance index calculation in the survcomp package in Bioconductor[42]. An alternative survival analysis approach was also applied, where the maximum signature exposure in each patient was used to assign a patient to either a good prognosis group (signature 1, 7, or 8) or a poor prognosis group (signature 2,3,4,5,6). A second Cox proportional hazards model was fitted on the training set using the groups as covariates to determine a log-rank score. The survfit function was used to generate a Kaplan-Meier plot for the grouping. The survdiff function was used test the performance of this approach in the test set and the survfit function used to generate Kaplan-Meier curves (Figure S1).

Association of copy-number signature exposures with other features

Of the 48 deep WGS BriTROC-1 samples, 44 had matched 2- and 3-star sWGS copy-number profiles. As the signature exposures from the sWGS were used for analysis, associations were made only with these 44 samples.

Mutational signatures. For 44 deep WGS BriTROC-1 samples and 95 ICGC samples, motif matrices were extracted using the SomaticSignatures R package (Gehring et al., 2015) and the weights of known COSMIC signatures were determined using the deconstructSigs R package

(Rosenthal et al., 2016). Signatures showing a median exposure of 0 across the samples were removed. The rcorr function in the Hmisc R package (Harrell, 2016) was used to calculate the correlation matrix between the remaining SNV and CN signature exposures.

Telomere length. Telomere lengths of 44 deep WGS tumour samples from the BriTROC-1 cohort was estimated using the Telomerecat algorithm (Farmery, 2017). Partial correlation was calculated between telomere length and copy-number signature exposures greater than 10%, with age and tumour purity as covariates, using the ppcor package in R (Kim, 2015).

Tandem duplicator phenotypes. Tandem duplicator phenotype (TDP) scores were calculated for 95 ICGC samples using the method described in Menghi et al (Menghi et al., 2016).

Amplification associated fold-back inversion fraction. For the 95 ICGC samples, the fraction of amplification associated fold-back inversion events per sample was calculated as the proportion of head-to-head inversions within a 100kb window amplified regions (copy number ≥5).

The significance of all observed correlations was estimated from a t-distribution where the null hypothesis was that the true correlation was 0. All reported p-values have been adjusted for multiple testing with Benjamini & Hochberg (BH) method (Benjamini and Hochberg, 1995). Comparison plots can be found in Figure S2.

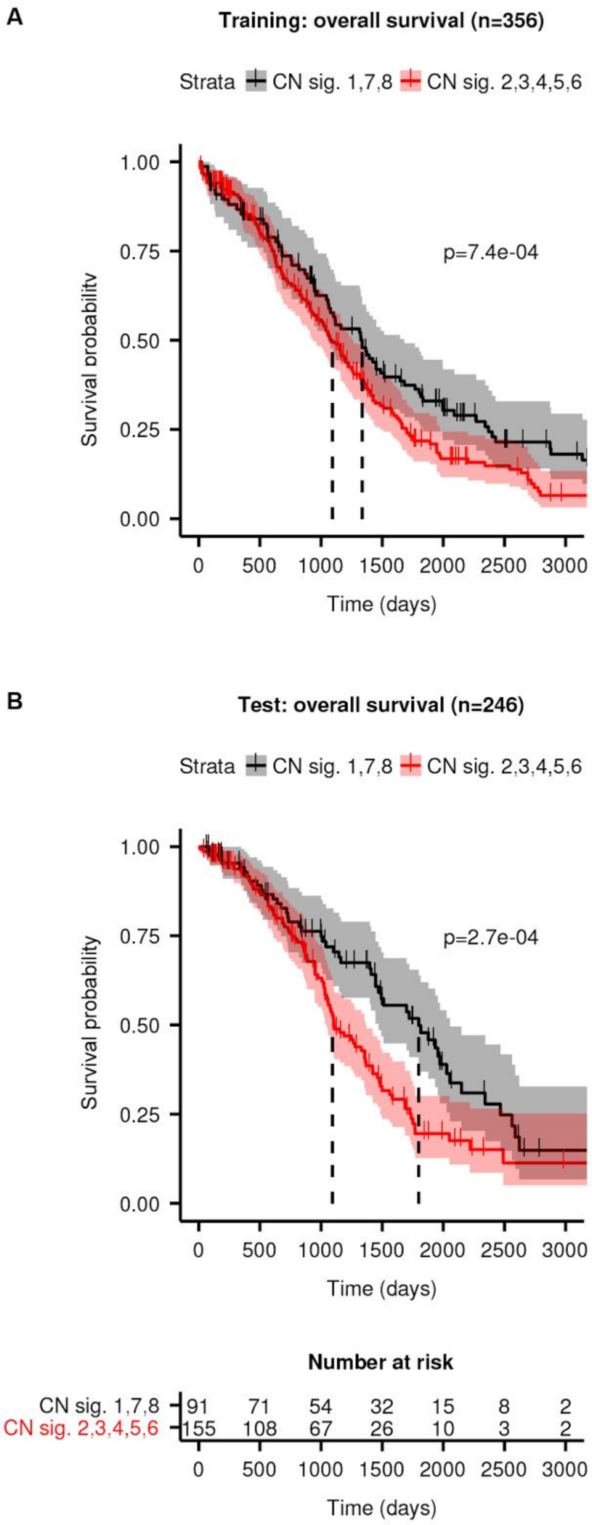
Mutated pathway enrichment analysis

A combined set of 319 samples (39 deep WGS BriTROC-1, 94 ICGC and 186 TCGA) showing at least one driver mutation were used for mutated pathway enrichment analysis. SNVs, INDELs, amplifications (CN>5) or deletions (CN<0.4) affecting 459 ovarian cancer driver genes from IntOGen (Gonzalez-Perez et al., 2013) were considered bona fide driver mutations if they had TIER1 or TIER2 status based on predictions by Cancer Genome Interpreter (Tamborero et al., 2017) (Table S2 and S3). 137 of the 459 genes were mutated in a least one case. These genes were used to test for enriched pathways in the Reactome database using the ReactomePA (Yu and He, 2016) R package with a p-value cutoff of 1 and q-value cutoff of 0.05. Pathways with at least 5 genes were retained. For each pathway, patients were split into two groups: those with mutated genes in the pathways, and those with wild-type genes in the pathways. For each signature, a one-sided t-test was carried out to determine if the signature exposure was significantly higher in mutated cases versus wild-type cases. After multiple testing correction using the Benjamini & Hochberg method (thresholding the p-value <0.1 and the mean difference in exposures >0.06), 113 pathways were significantly enriched (Table S4). Manual inspection revealed significant redundancy in the list and 5 representative pathways were selected as a final output (Figure S3, Table S5).

Data and Software Availability

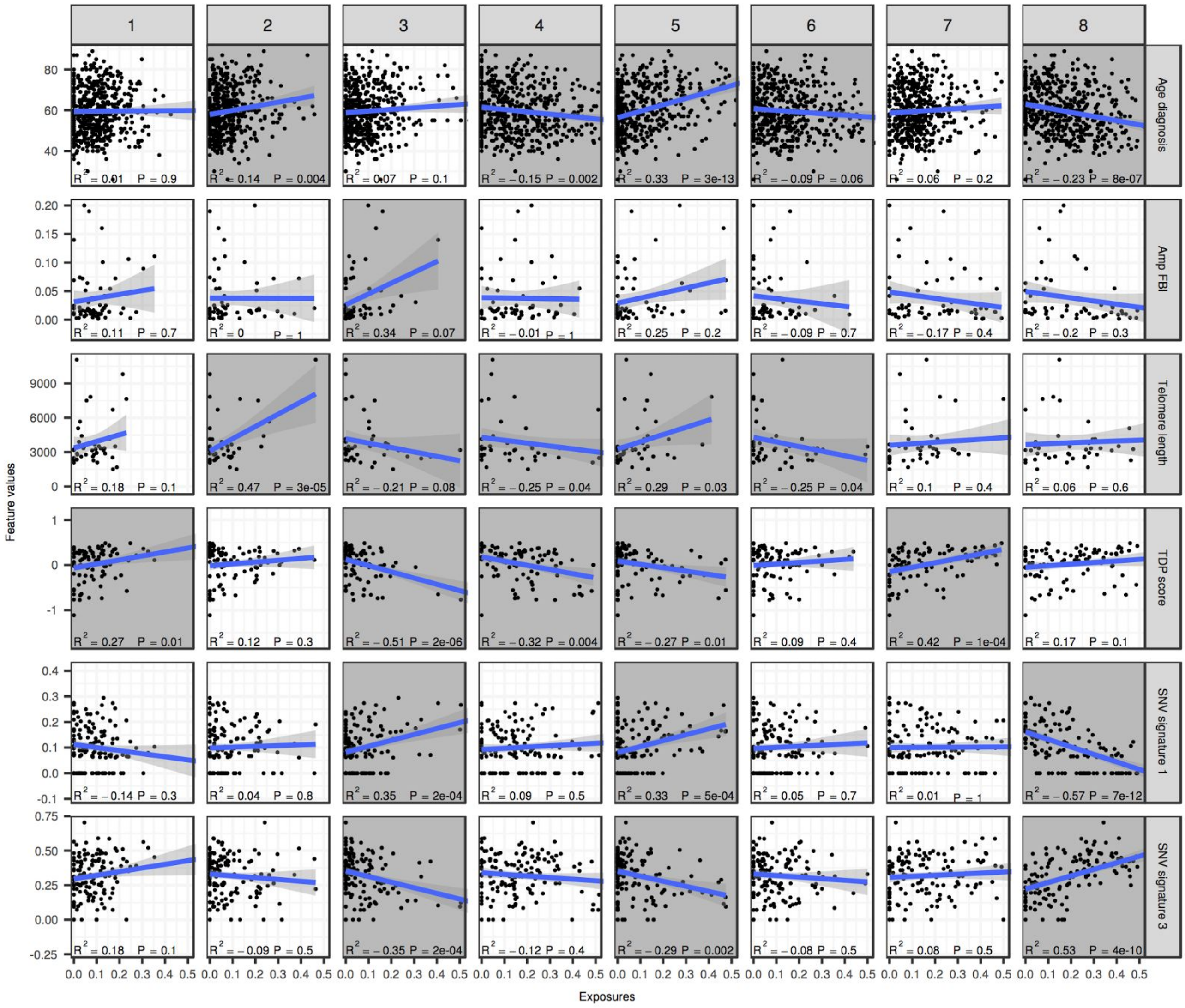
Sequence data that support the findings of this study have been deposited in the European Genome-phenome Archive with the accession code EGAS00001002557. All code required to reproduce the analysis outlined in this manuscript can be found in the following repository: https://bitbucket.org/britroc/cnsignatures

Figure S1 | Survival analysis of patients stratified by dominant signature exposure



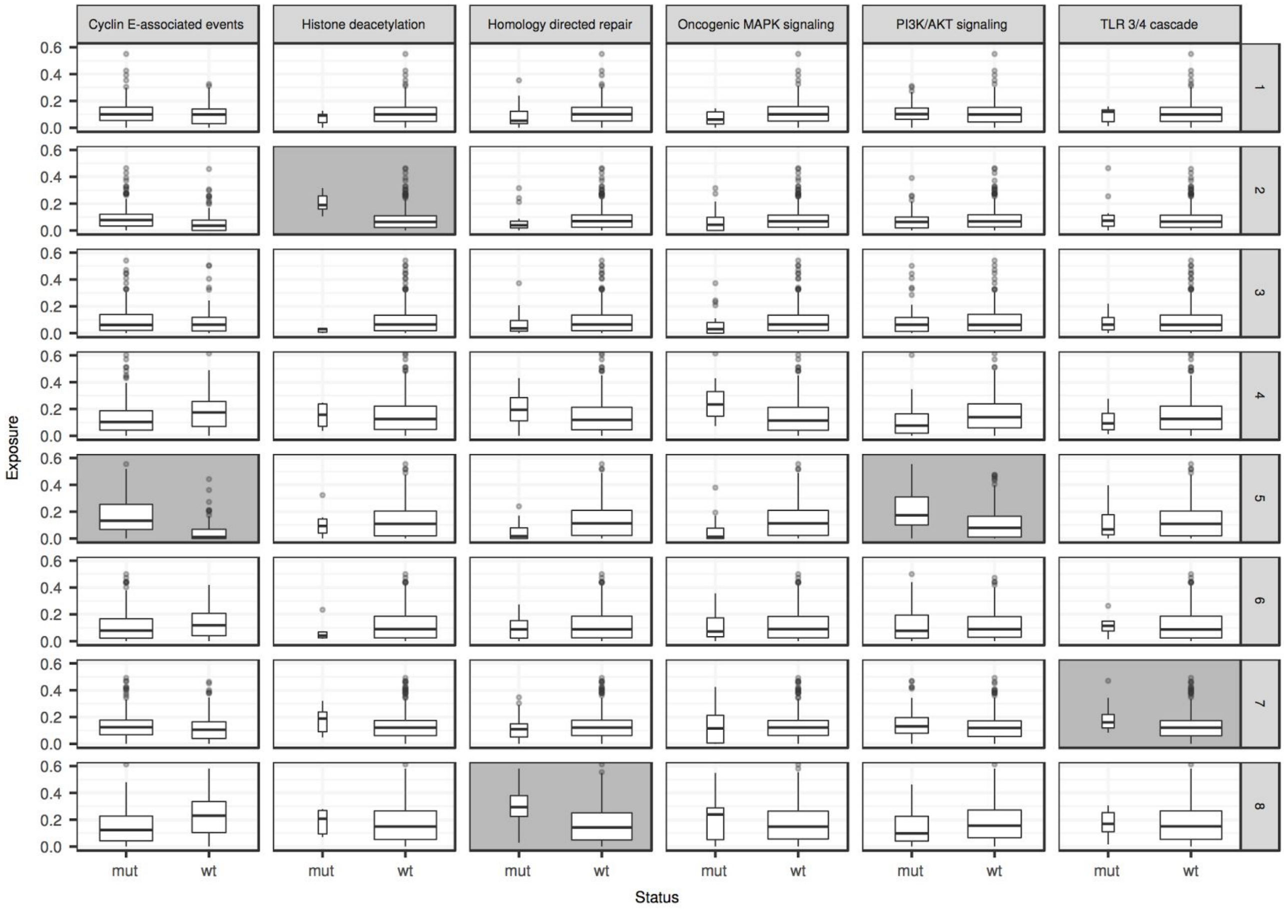
Kaplan-Meier curves of overall survival probabilities for patients in **A** training cohort and **B** test cohort. Each cohort was divided into two groups based on the patients' maximum observed signature exposure (Red: maximum observed signature exposure for signatures 2,3,4,5, or 6; Black: maximum observed signature exposure for signatures 1,7 or 8). Black dotted lines indicate the median survival for each group. The p-values reported are from a log-rank test. Shaded areas represent 95% confidence intervals. Below **B** is a table specifying the number of remaining cases at risk every 500 days. These is no at risk table for **A** as the BriTROC-1 cases have left truncated survival.

Figure S2 | Correlation plots of signature exposures with other features and SNV signatures



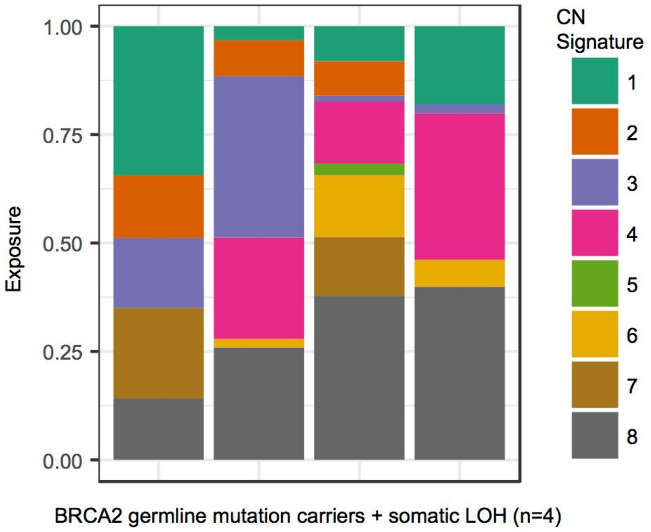
Feature values (right) versus signature exposures (top). Blue lines represent a linear model fit and shading around the lines represent the 95% confidence interval. Shaded panels represent results which are significantly correlated (adjusted P<0.1).

Figure S3 | Differences in exposures between cases with mutated pathways versus wild-type



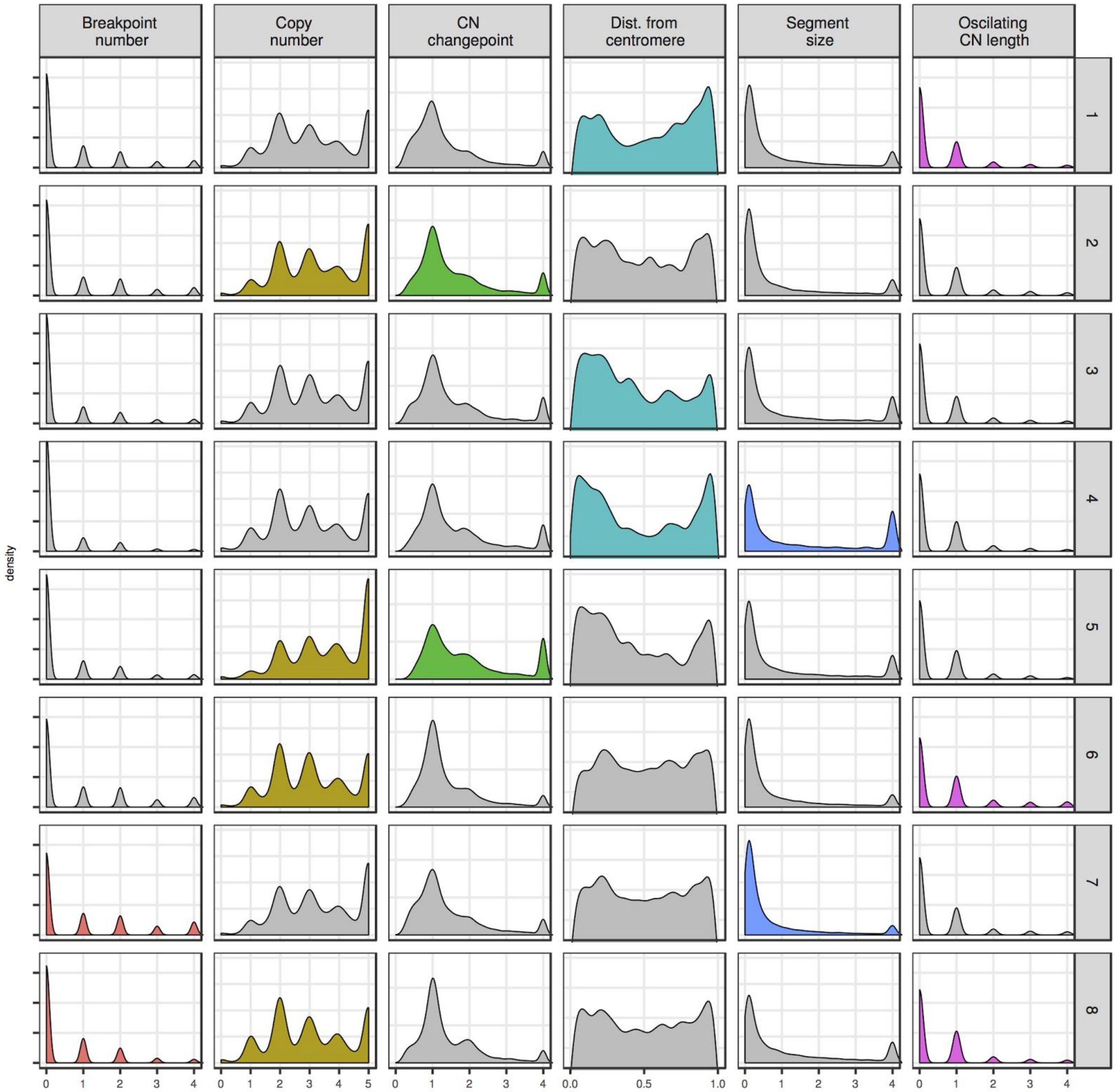
Boxplots representing the signature exposures of cases with mutations (mut) in a given pathway (top) versus those with wild-type alleles (wt). The box widths are proportional to the number of cases (exact numbers can be found in Figure 5). Shaded panels indicate significant differences (adjusted P<0.1, values found in Table S4).

Figure S4 | CN signature exposures of 4 BriTROC-1 patients with germline *BRCA2* mutations and somatic loss of heterozygosity.



This figure shows stacked bar plots of copy-number signature exposures for 4 BriTROC-1 cases with confirmed pathogenic germline BRCA2 mutations plus somatic loss of heterozygosity (LOH).

Figure S5 | Overview of copy-number feature distributions



Separate density distributions are plotted for each copy-number feature across all of the signatures. The distributions that have highly weighted variables (see Figure 6) for each of the feature distributions are coloured.