

# Genomic evolution and chemoresistance in germ-cell tumours

Amaro Taylor–Weiner<sup>1,2\*</sup>, Travis Zack<sup>1,3\*</sup>, Elizabeth O’Donnell<sup>4,5</sup>, Jennifer L. Guerriero<sup>4</sup>, Brandon Bernard<sup>4</sup>, Anita Reddy<sup>6</sup>, G. Celine Han<sup>2,4</sup>, Saud AlDubayan<sup>7,8</sup>, Ali Amin–Mansour<sup>2</sup>, Steven E. Schumacher<sup>2,9</sup>, Kevin Litchfield<sup>10,11</sup>, Clare Turnbull<sup>10,11</sup>, Stacey Gabriel<sup>2</sup>, Rameen Beroukhim<sup>2,4</sup>, Gad Getz<sup>2,12</sup>, Scott L. Carter<sup>2,13,14,15</sup>, Michelle S. Hirsch<sup>16</sup>, Anthony Letai<sup>4</sup>, Christopher Sweeney<sup>4§</sup> & Eliezer M. Van Allen<sup>2,4,13§</sup>

**Germ-cell tumours (GCTs) are derived from germ cells and occur most frequently in the testes<sup>1,2</sup>. GCTs are histologically heterogeneous and distinctly curable with chemotherapy<sup>3</sup>. Gains of chromosome arm 12p and aneuploidy are nearly universal in GCTs<sup>4–6</sup>, but specific somatic genomic features driving tumour initiation, chemosensitivity and progression are incompletely characterized. Here, using clinical whole-exome and transcriptome sequencing of precursor, primary (testicular and mediastinal) and chemoresistant metastatic human GCTs, we show that the primary somatic feature of GCTs is highly recurrent chromosome arm level amplifications and reciprocal deletions (reciprocal loss of heterozygosity), variations that are significantly enriched in GCTs compared to 19 other cancer types. These tumours also acquire *KRAS* mutations during the development from precursor to primary disease, and primary testicular GCTs (TGCTs) are uniformly wild type for *TP53*. In addition, by functional measurement of apoptotic signalling (BH3 profiling) of fresh tumour and adjacent tissue<sup>7</sup>, we find that primary TGCTs have high mitochondrial priming that facilitates chemotherapy-induced apoptosis. Finally, by phylogenetic analysis of serial TGCTs that emerge with chemotherapy resistance, we show how TGCTs gain additional reciprocal loss of heterozygosity and that this is associated with loss of pluripotency markers (*NANOG* and *POU5F1*)<sup>8,9</sup> in chemoresistant teratomas or transformed carcinomas. Our results demonstrate the distinct genomic features underlying the origins of this disease and associated with the chemosensitivity phenotype, as well as the rare progression to chemoresistance. These results identify the convergence of cancer genomics, mitochondrial priming and GCT evolution, and may provide insights into chemosensitivity and resistance in other cancers.**

GCTs arise from germ cells, which can be found in the gonads (testicles and ovaries), mediastinum, pituitary and retroperitoneum, and progress from precursor lesions (germ-cell neoplasia *in situ*; GCNIS)<sup>1,2,10</sup>. Histologically, GCTs are classified as seminoma or non-seminoma. The latter may consist of one histology or be mixed non-seminoma. Clinically, most GCTs are exceptionally chemosensitive; combination cisplatin- and etoposide-based chemotherapy cures more than 80% of GCT patients with metastatic disease<sup>3</sup>. However, approximately 10% of patients with metastatic GCTs will die as a result of chemoresistant disease.

Leveraging preclinical models, several mechanisms have been proposed for GCT oncogenesis and chemosensitivity, including DNA

repair deficiencies, embryonic stem (ES) cell-like cell cycle properties, or high apoptotic propensity<sup>11–13</sup>. Germline genome wide association studies have identified several risk loci for GCT, most notably in *KITLG*<sup>14,15</sup>. For GCTs, karyotyping demonstrated marked aneuploidy<sup>6</sup>, targeted sequencing identified rare *KRAS* hotspot mutations<sup>16</sup>, and whole-exome sequencing (WES) confirmed nearly universal arm-level chromosome arm 12p gain<sup>4</sup>, low mutation rates, and activating *KIT* mutations specific to seminomas<sup>5</sup>. However, genomic features in GCTs that contribute to the origins, chemosensitivity phenotype, and progression remain incompletely characterized.

Comparison of serial treatment-responsive and refractory tumours has informed genomic progression in other cancers<sup>17,18</sup>. We proposed that genomic evolutionary analysis of clinical GCTs from patients who respond and become resistant to chemotherapy would identify features underlying disease progression and chemotherapy response phenotype. We performed clinically integrated molecular analysis of 59 tumours samples (with matched germ line) from 47 patients with TGCTs and two patients with primary mediastinal GCTs (PMGCTs) to identify genomic features associated with disease origin and progression. The aggregate cohort features are summarized in Fig. 1a and Supplementary Tables 1, 2.

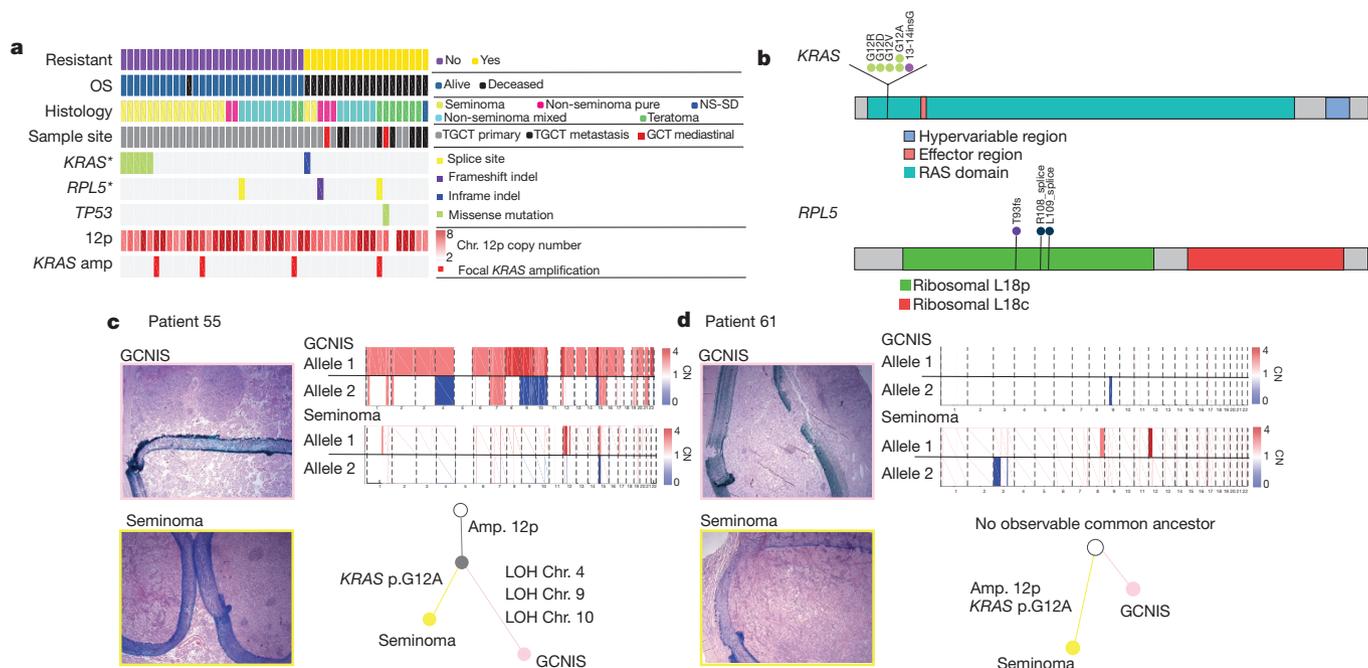
The median mutational load in our cohort was 0.9 mutations Mb<sup>-1</sup> (Supplementary Table 1). Mutational significance analysis<sup>19</sup> identified *KRAS* as the most statistically significant altered gene in this cohort (Fig. 1a, b, Supplementary Table 3). The only other significantly mutated gene was *RPL5*, a ribosomal protein previously implicated in other cancer types<sup>19</sup> (Fig. 1b). *RPL5* may regulate the MDM2–TP53 axis<sup>20</sup> but was mutated infrequently and is of unknown biological relevance in GCTs. *KIT* mutations were not observed in our cohort despite sufficient genomic power to detect mutations (*KIT* hotspot loci mean coverage: 316×; range: 89–800×), although meta-analysis of our cohort with a separate TGCT WES set<sup>5</sup> confirms *KIT* as significantly mutated (Extended Data Fig. 1a). Expanded investigation for clinically or biologically relevant alterations identified singleton mutations in DNA repair or sex determination pathway gene sets (see Supplementary Table 4 for all somatic mutations), although these were rare and of unclear significance.

Because *KRAS* emerged as the primary significantly mutated gene, we next sought to explore the point of *KRAS* mutation emergence in the context of TGCT progression. We performed phylogenetic analysis of WES from patient-matched GCNIS tumours (pre-invasive) and

<sup>1</sup>Division of Medical Sciences, Harvard University, Boston, Massachusetts 02115, USA. <sup>2</sup>Cancer Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. <sup>3</sup>Health Sciences and Technology, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>4</sup>Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. <sup>5</sup>Department of Medical Oncology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. <sup>6</sup>Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>7</sup>Division of Genetics and Genomics, Department of Medicine, Boston Children’s Hospital, Massachusetts 02115, USA. <sup>8</sup>Department of Medicine, King Saud bin Abdulaziz University for Health Sciences, Saudi Arabia. <sup>9</sup>Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. <sup>10</sup>Division of Genetics and Epidemiology, The Institute of Cancer Research, Fulham Road, London SW3 6JB, UK. <sup>11</sup>William Harvey Research Institute, Queen Mary University London, Charterhouse Square, London EC1M 6BQ, UK. <sup>12</sup>Cancer Center and Department of Pathology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. <sup>13</sup>Center for Cancer Precision Medicine, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. <sup>14</sup>Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. <sup>15</sup>Harvard T.H. Chan School of Public Health, Boston, Massachusetts 02115, USA. <sup>16</sup>Department of Pathology, Brigham and Women’s Hospital, Boston, Massachusetts 02115, USA.

\*These authors contributed equally to this work.

§These authors jointly supervised this work.



**Figure 1 | Mutational landscape and evolution from precursor lesions.**

**a**, Clinical and genomic overview. Each column represents a patient sample. The top four rows represent chemoresistant disease status, overall survival (OS) vital status, predominant histological classification of the primary sample (when available, or selected secondary sample otherwise), and anatomic site, respectively. The next three rows indicate the patient's mutation status for the two significantly ( $*q < 0.1$ ) mutated genes, and *TP53*. Next median copy number of chromosome arm 12p is shown, followed by the focal amplification status of *KRAS* (defined as *KRAS* copy number greater than  $2 \times$  median 12p copy number). NS-SD refers to non-seminoma with sarcoma differentiation. **b**, Mutations in *KRAS* and

*RPL5*. Each circle represents a mutation observed in a single patient, and labels indicate amino acid change. **c**, **d**, Phylogenetic relationship between GCNIS and primary TGCT. Histological slide images from DFCL\_55 and DFCL\_61 are shown on the left. Images in the pink borders are GCNIS and below in yellow are tumour images. Dark blue ink indicates the border of the lesion. Next to the histological images are phylogenetic trees determined by shared allelic copy number (CN) mutations. The non-shaded circle at the top of the tree indicates the patient's germ line. Below that the filled grey is the inferred shared ancestor clone and then each sample branches ends in the sequenced sample. Above are allelic copy number data supporting the tree.

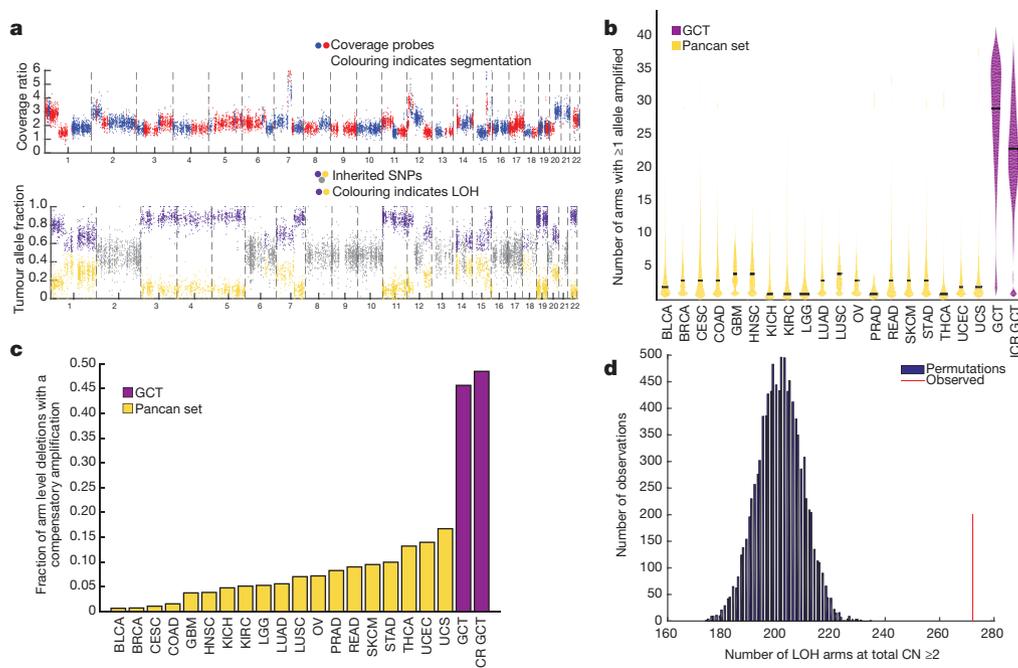
primary TGCT tumours (Methods). In one representative case, both samples contained chromosome arm 12p gain; however, the only somatic putative driver mutation distinguishing the two samples was a *KRAS*<sup>G12A</sup> mutation in the TGCT (0 out of 377 reads in GCNIS, 65 out of 340 reads in TGCT) (Fig. 1c, Extended Data Fig. 2a). A similar pattern of *KRAS* mutation exclusive to the TGCT and not in the patient-matched GCNIS was also seen in a second patient (0 out of 210 reads in GCNIS, 35 out of 184 reads in TGCT) (Fig. 1d, Extended Data Fig. 2b). Although not required for progression, activating *KRAS* mutations may occur after arm level gain of chromosome 12p, thereby indicating that these are two separate processes involved in TGCT evolution.

While GCTs had low point mutation rates, nearly all tumours contained arm level gain of chromosome arm 12p, as previously described<sup>5</sup>. Given previous reports of aneuploidy in GCTs<sup>6,21</sup>, we performed allelic and absolute copy number analysis through allele deconstruction<sup>22,23</sup> to determine whether there were specific copy number patterns in GCTs not been previously identifiable or reported (Methods). In a representative case (DFCL\_7; Fig. 2a), coverage-based copy number analysis identified 12p gain, but only allelic copy number analysis revealed frequent arm and chromosome level gains of one parental allele with simultaneous loss of the other parental allele, leading to loss of heterozygosity (LOH). These reciprocal LOH (RLOH) events often maintained the germline number of DNA copies (copy-neutral LOH), but also frequently have further amplification of the remaining parental allele. For example, in DFCL\_7 chromosome 4 has undergone copy-neutral LOH, whereas chromosome 3q contains a LOH event in which the total copy number is three. Across the GCT cohort, the mean number of chromosomal arm level amplifications was  $28.3 \pm 7.5$  (mean  $\pm$  s.d.), which is significantly increased compared to 6,509 tumours representing 19 other cancer types ( $P < 0.0001$ ; Mann-Whitney) (Fig. 2b, Extended Data Fig. 3 and Supplementary Tables 5, 6).

Approximately 45% of arm level deletions contained a compensatory reciprocal amplification in GCTs, again sharply increased compared to other tumour types ( $P < 0.0001$ ; Mann-Whitney) (Fig. 2c).

To validate this RLOH pattern, we performed the same analysis on a separate primary TGCT fresh-frozen WES cohort (ICR GCT)<sup>5</sup> and observed similar frequencies of reciprocal copy number alterations (Fig. 2b, c, Extended Data Fig. 1b–d). Given the prevalence of arm level amplifications observed in GCTs, we performed a permutation test involving deleted chromosomal arms to examine whether these deletions would be compensatory by chance<sup>23</sup> (Methods). We observed significantly increased reciprocal events in GCTs than expected by chance (empirical  $P < 0.0001$ ) (Fig. 2d). Finally, to examine whether this copy number feature was an intrinsic germ-cell property and not related to microenvironmental features, we performed absolute copy number analysis on three tumours originating in the testes with non-germ-cell histologies and did not identify such chromosomal abnormalities (Extended Data Fig. 4). Overall, GCTs have a unique pattern of highly recurrent reciprocal copy number alterations in both discovery and validation cohorts.

Despite extensive copy number events in GCTs, all TGCTs were *TP53* wild type, and we only observed one *TP53* mutation in a PMGCT. Transcriptome profiling of a patient subset ( $n = 28$ : 21 primary TGCTs, five metastatic samples from TGCTs, and two PMGCT; Methods) demonstrated *TP53* expressed across histologies (Fig. 3a, Supplementary Table 7). The overall constellation of genomic changes in clinical TGCT tumours—wild-type *TP53*, chromosome arm 12p gain, and additional recurrent copy number changes—is similar to adaptive mutations acquired by human ES cells during prolonged passaging *in vitro*<sup>24,25</sup>. Like GCTs, ES cells are acutely sensitive to DNA damage induced by chemotherapy, and it has been shown that this sensitivity is a result of intact p53 status and high mitochondrial priming<sup>26</sup>.



**Figure 2 | RLOH in GCTs.** **a**, Representative allelic somatic copy number data. Top, coverage-derived copy number events in a tumour. Bottom, the allele fraction of germ-line variants observed in the tumour. Regions in which there is LOH are coloured in purple and orange for the retained and lost allele, respectively. **b**, Arm level amplification rates in GCTs and 19 other tumour types (see <https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/tcga-study-abbreviations> for abbreviations). Shown are the distributions of counts of arm level amplifications, after controlling for whole-genome doubling, in samples from each tumour type. The distributions are estimated using a Gaussian kernel density<sup>31</sup> and the median count is indicated by the black bar. GCTs (purple, with the independent replication series denoted as ICR GCT) have significantly more arm level amplifications than other tumour types (gold) ( $P < 0.0001$ ;

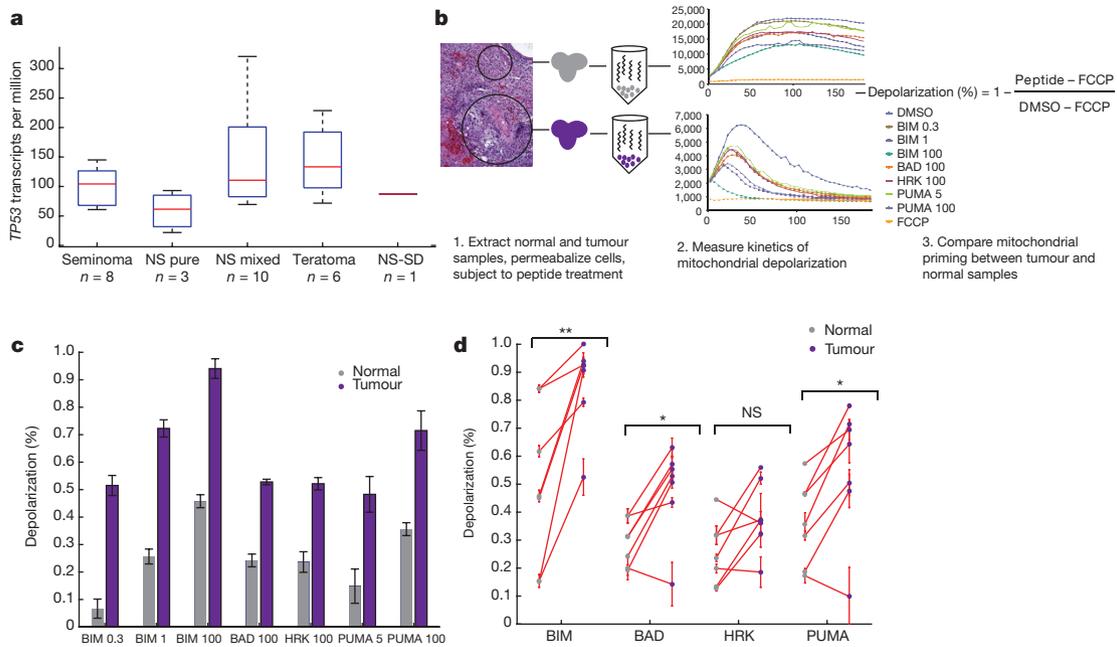
Mann–Whitney). **c**, Rate of reciprocal deletion on arms with amplification in GCTs and 19 other tumour types. For each sample, we counted the number of arm level amplifications with and without a reciprocal deletion. Shown are the mean fractions of amplified arms with a reciprocal arm level deletion, after controlling for whole-genome doubling. GCTs (purple) have significantly more reciprocal deletions than other tumour types (gold) ( $P < 0.0001$ ; Mann–Whitney). **d**, Permutation test of arm level deletions. To generate a null distribution of reciprocal deletions, we shuffled arm level deletions within samples<sup>23</sup> and for each permutation we counted the number of observed reciprocal arm level deletions. Shown is a histogram of the counts of reciprocal deletions observed in permutations. The red line indicates the number observed in our dataset corresponding to an empirical  $P < 0.0001$ .

Also, one of the earliest identified susceptibility loci in TGCTs is in *BAK1*, a *BCL2* antagonist and pro-apoptotic gene<sup>14</sup>, and chemosensitivity in other cancer types has been associated with increased mitochondrial priming<sup>7</sup>. We therefore proposed that the basis of chemosensitivity in TGCTs with a wild-type *TP53* and RLOH abundant genomic background was a result of a fundamental apoptotic propensity caused by increased mitochondrial priming. Mitochondrial priming refers to the intrinsic potential of cellular apoptotic propensity due to the balance of pro-apoptotic and anti-apoptotic *BCL2* family proteins at the mitochondria, a feature that can be assessed by BH3 profiling<sup>27</sup>. To test this hypothesis, we obtained fresh tumour samples from seven TGCTs (Fig. 3b) and performed dynamic BH3 profiling on the tumours and adjacent normal tissue<sup>7</sup> (Methods). In one representative case, we observed an increase in BIM BH3-induced mitochondrial depolarization between the tumour and adjacent normal (Fig. 3c) ( $P < 0.0001$ ; paired *t*-test). Across the seven samples, BIM BH3-induced mitochondrial priming remained significant (Fig. 3d) ( $P < 0.0001$ ; paired *t*-test). Thus, intact *TP53*, RLOH, and high mitochondrial priming may form the foundation of chemosensitivity in TGCTs.

Although most TGCTs are cured with chemotherapy, some recur and approximately 10% of patients with metastatic disease will die from chemoresistant TGCT. To track the genomic evolution of chemoresistant TGCTs, phylogenetic analysis of 13 WES tumour samples (pre- and post-chemotherapy) from 5 patients was performed (Fig. 4, Extended Data Fig. 5, Supplementary Table 8, Methods). Chromosome arm 12p was a truncal event shared in all samples. In each patient, the resistant tumour that arose after either first- or second-line chemotherapy accumulated additional copy number events, including reciprocal deletions and events involving new chromosomes (Fig. 4). In one

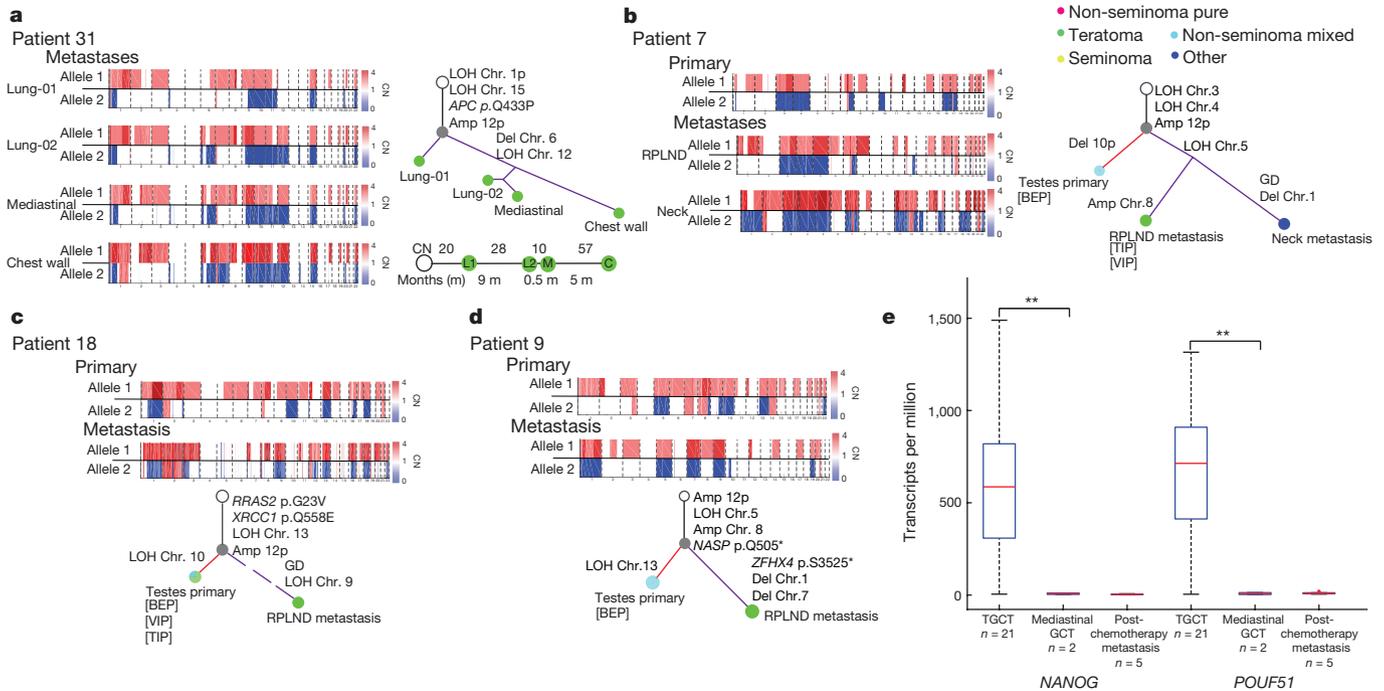
patient with serial tumours resected over a 13-month period without intervening chemotherapy, the copy-number-derived phylogeny was proportional to the sequence and elapsed time of sample acquisition (Fig. 4a). In two cases in which the final post-chemotherapy sample had teratoma and/or poorly differentiated carcinoma components consistent with differentiation (Fig. 4b, c), complete genome doubling was observed. Notably, the pluripotency and apoptosis regulators *NANOG* and *POU5F1* (also known as *OCT3/4*) expressed in GCT<sup>8,9,13</sup> were not expressed in the tumours obtained from metastatic deposits resistant to chemotherapy (Fig. 4e;  $P < 0.001$ , Mann–Whitney). They were also not expressed in the two incurable PMGCTs (Fig. 4e, Supplementary Table 7). Taken together, chemotherapy-resistant germ-cell tumours are associated with continued progression of RLOH copy number events and loss of pluripotency markers in this clinical cohort, although it is uncertain whether the loss of pluripotency markers is a driver of chemoresistance<sup>11,28</sup>, a marker of methylation changes in a further differentiated histology post-chemotherapy<sup>29</sup>, or a feature of both processes.

Overall, while GCTs represent a heterogeneous set of histological subtypes, they share a distinct burden and pattern of reciprocal copy number alterations. This observation is consistent with the notion that GCTs emerge, in part, through dysregulation in the mitosis/meiosis switch from arrested gonocytes, since this process may result in abnormal chromosomal segregation<sup>30</sup> and persist during the evolution of GCTs. Expanded molecular analysis may inform the initiators of RLOH and its role in oncogenesis. Subsets of these tumours acquire *KRAS*-activating mutations, although the necessity for *KRAS* mutation is less clearly defined, and these tumours lack recurrent nonsynonymous mutations in established cancer genes<sup>19</sup>. Instead, chemosensitivity in TGCTs may be the result of high mitochondrial priming properties, and



**Figure 3 | Mitochondrial priming in germ cell tumours.** **a**, Gene expression of *TP53* in GCTs. Values are in transcripts per million (TPM). **b**, BH3 profiling pipeline schematic. Samples were collected surgically and then processed immediately as previously described<sup>7</sup>. Mitochondrial priming was measured in both the tumour and normal samples after exposure to each peptide at varying concentrations. The priming caused by each peptide is compared to the negative and positive controls (dimethylsulfoxide (DMSO) and the uncoupler carbonyl cyanide-*p*-trifluoromethoxyphenylhydrazone (FCCP), respectively) and between the tumour and normal sample. BAD, BIM, HRK and PUMA denote BH3-only BCL2 proteins, and the values (0.3, 1, 100) represent micromolar

concentrations. **c**, Example comparison of peptide-induced depolarization in a tumour and normal sample. Mean depolarization across three replicates for the tumour (purple) and normal (grey) samples are shown. The tumour shows significantly more mitochondrial priming than the normal sample ( $P = 2.17 \times 10^{-5}$ ; paired *t*-test). **d**, BH3 peptide profiling in tumour–normal pairs. Shown is the mean depolarization across replicates for seven tumour and normal samples subjected to each peptide at the maximum concentration. Matched tumour (purple) and normal (grey) samples are connected in red. Error bars in **c** and **d** denote s.d. \* $P < 0.05$ ; \*\* $P < 0.01$ ; NS, not significant ( $P > 0.05$ ); paired *t*-test.



**Figure 4 | Phylogenetic analysis and pluripotency of primary and metastatic GCTs.** **a–d**, For each patient sample (DFCI\_31 (**a**), DFCI\_7 (**b**), DFCI\_18 (**c**) and DFCI\_9 (**d**)), histology proportion is indicated by pie charts within each phylogenetic tree. Treatments received after a given sample was acquired are listed below that respective sample (BEP, bleomycin, etoposide, cisplatin; TIP, paclitaxel, ifosfamide, cisplatin; VIP, vinblastine, ifosfamide, cisplatin). GD, genome doubling; RPLND, retroperitoneal lymph node dissection. Phylogenetic trees were constructed using allelic copy number deconstructions (Methods). Branches leading to primary samples are red,

and branches leading to metastases are purple. Dotted branches indicate deconstructions impacted by formalin-fixed, paraffin-embedded (FFPE) sample degradation, limiting length estimation. For patient DFCI\_31 (**a**) the timeline below the tree shows time (months) and number of copy number events called by allelic deconstruction between samples (C, chest wall; L1, lung-01; L2, lung-02; M, mediastinal). **e**, Gene expression quantification of *NANOG* and *POU5F1* in primary and metastatic GCTs. Values are in transcripts per million. The red point denotes a *TP53*-mutant mediastinal primary GCT (DFCI\_19). \*\* $P < 0.01$ , Mann–Whitney test.

molecular characterization paired with mitochondrial priming studies of treatment-resistant tumours may inform features exclusive to this rare patient subset with significant unmet medical need for new therapies. This strategy may also inform strategies to improve the efficacy of chemotherapy in other malignancies.

**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 30 May; accepted 2 November 2016.**

- Oosterhuis, J. W. & Looijenga, L. H. Testicular germ-cell tumours in a broader perspective. *Nat. Rev. Cancer* **5**, 210–222 (2005).
- Sonne, S. B. *et al.* Analysis of gene expression profiles of microdissected cell populations indicates that testicular carcinoma in situ is an arrested gonocyte. *Cancer Res.* **69**, 5241–5250 (2009).
- Hanna, N. H. & Einhorn, L. H. Testicular cancer—discoveries and updates. *N. Engl. J. Med.* **371**, 2005–2016 (2014).
- Atkin, N. B. & Baker, M. C. Specific chromosome change, i(12p), in testicular tumours? *Lancet* **2**, 1349 (1982).
- Litchfield, K. *et al.* Whole-exome sequencing reveals the mutational spectrum of testicular germ cell tumours. *Nat. Commun.* **6**, 5973 (2015).
- Oosterhuis, J. W. *et al.* Ploidy of primary germ cell tumors of the testis. Pathogenetic and clinical relevance. *Lab. Invest.* **60**, 14–21 (1989).
- Montero, J. *et al.* Drug-induced death signaling strategy rapidly predicts cancer response to chemotherapy. *Cell* **160**, 977–989 (2015).
- Looijenga, L. H. *et al.* POU5F1 (OCT3/4) identifies cells with pluripotent potential in human germ cell tumors. *Cancer Res.* **63**, 2244–2250 (2003).
- Hart, A. H. *et al.* The pluripotency homeobox gene *NANOG* is expressed in human germ cell tumors. *Cancer* **104**, 2092–2098 (2005).
- Moch, H., Cubilla, A. L., Humphrey, P. A., Reuter, V. E. & Ulbright, T. M. The 2016 WHO classification of tumours of the urinary system and male genital organs—part A: renal, penile, and testicular tumours. *Eur. Urol.* **70**, 93–105 (2016).
- Gutekunst, M. *et al.* Cisplatin hypersensitivity of testicular germ cell tumors is determined by high constitutive Noxa levels mediated by Oct-4. *Cancer Res.* **73**, 1460–1469 (2013).
- Jacobsen, C. & Honecker, F. Cisplatin resistance in germ cell tumours: models and mechanisms. *Andrology* **3**, 111–121 (2015).
- Rijlaarsdam, M. A. & Looijenga, L. H. An oncofetal and developmental perspective on testicular germ cell cancer. *Semin. Cancer Biol.* **29**, 59–74 (2014).
- Rapley, E. A. *et al.* A genome-wide association study of testicular germ cell tumor. *Nat. Genet.* **41**, 807–810 (2009).
- Kanetsky, P. A. *et al.* Common variation in *KITLG* and at 5q31.3 predisposes to testicular germ cell cancer. *Nat. Genet.* **41**, 811–815 (2009).
- Feldman, D. R. *et al.* Presence of somatic mutations within PIK3CA, AKT, RAS, and FGFR3 but not BRAF in cisplatin-resistant germ cell tumors. *Clinical Cancer Res.* **20**, 3712–3720 (2014).
- Stachler, M. D. *et al.* Paired exome analysis of Barrett's esophagus and adenocarcinoma. *Nat. Genet.* **47**, 1047–1055 (2015).
- Brastianos, P. K. *et al.* Genomic characterization of brain metastases reveals branched evolution and potential therapeutic targets. *Cancer Discov.* **5**, 1164–1177 (2015).
- Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
- Goudarzi, K. M. & Lindström, M. S. Role of ribosomal protein mutations in tumor development (Review). *Int. J. Oncol.* **48**, 1313–1324 (2016).
- Wang, L. *et al.* Novel somatic and germline mutations in intracranial germ cell tumours. *Nature* **511**, 241–245 (2014).
- Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).
- Zack, T. I. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).
- Baker, D. E. *et al.* Adaptation to culture of human embryonic stem cells and oncogenesis *in vivo*. *Nat. Biotechnol.* **25**, 207–215 (2007).
- Närvä, E. *et al.* High-resolution DNA analysis of human embryonic stem cell lines reveals culture-induced copy number changes and loss of heterozygosity. *Nat. Biotechnol.* **28**, 371–377 (2010).
- Liu, J. C. *et al.* High mitochondrial priming sensitizes hESCs to DNA-damage-induced apoptosis. *Cell Stem Cell* **13**, 483–491 (2013).
- Ni Chonghaile, T. *et al.* Pretreatment mitochondrial priming correlates with clinical response to cytotoxic chemotherapy. *Science* **334**, 1129–1133 (2011).
- Abada, P. B. & Howell, S. B. Cisplatin induces resistance by triggering differentiation of testicular embryonal carcinoma cells. *PLoS One* **9**, e87444 (2014).
- Wermann, H. *et al.* Global DNA methylation in fetal human germ cells and germ cell tumours: association with differentiation and cisplatin resistance. *J. Pathol.* **221**, 433–442 (2010).
- Jørgensen, A. *et al.* Dysregulation of the mitosis-meiosis switch in testicular carcinoma in situ. *J. Pathol.* **229**, 588–598 (2013).
- Hoffman, H. violin.m — Simple violin plot using Matlab default kernel estimation. INRES (Univ. of Bonn, 2015).

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

**Acknowledgements** We thank the patients for contributing to this study, and H.Taylor-Weiner for feedback on ES cells. This work was supported by NIH U54 HG003067, NIH 1K08 CA188615 (E.M.V.), Damon Runyon Clinical Investigator Award (E.M.V.), Shawmut Design and Construction Pan Mass Challenge Team (C.S.), and Giovino Jimmy Fund Golf Tournament (C.S.).

**Author Contributions** A.T.-W., T.Z., B.B., G.C.H., S.A., A.A.-M. and E.M.V. performed genomic analysis of discovery cohort. A.T.-W., T.Z., B.B., E.O., M.H., C.S. and E.M.V. performed clinical integration and analysis. J.L.G. and A.L. performed BH3 profiling experiments. S.S., S.L.C., R.B. and G.G. contributed methodology and analysis review. A.R. and E.M.V. performed biological review of genomic findings. S.G. performed sequencing assays. A.T.-W., T.Z., K.L., C.T. and E.M.V. performed genomic analysis of validation cohort. M.H. performed pathology and histological evaluation of clinical samples. A.T.-W., T.Z., B.B., C.S. and E.M.V. prepared manuscript and figures.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://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 E.M.V. ([eliezerm\\_vanallen@dfci.harvard.edu](mailto:eliezerm_vanallen@dfci.harvard.edu)).

**Reviewer Information** *Nature* thanks K. Nathanson and the other anonymous reviewer(s) for their contribution to the peer review of this work.

## METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

**Patient enrollment.** Patients with GCTs with available tumour and germline samples provided informed consent to an Institutional Review Board protocol that allows research molecular characterization of tumour and germline samples (Dana-Farber Cancer Institute 10-025 and 11-104). All available cases that met these criteria were included in this study. The clinical characteristics for these patients are available in Supplementary Tables 1 and 2.

**Pathology review and sample processing.** All samples were reviewed by a genitourinary oncology pathologist (M.H.). After fixation and mounting, 5–10 10- $\mu\text{m}$  slices from formalin-fixed, paraffin-embedded (FFPE) tumour blocks were obtained, and tumour-enriched tissue was macrodissected. Paraffin was removed from FFPE sections and cores using CitriSolv (Fisher Scientific), followed by ethanol washes and tissue lysis overnight at 56 °C. Samples were then incubated at 90 °C to remove DNA crosslinks, and DNA (and when possible, RNA) extraction was performed using Qiagen AllPrep DNA/RNA Mini Kit (51306). Germline DNA was obtained from adjacent normal tissue or peripheral blood mononuclear cells. **WES.** WES from FFPE tumour samples and matched germline DNA was performed as previously described<sup>32</sup>.

**Library construction.** DNA libraries for massively parallel sequencing were generated as previously described<sup>32</sup> with the following modifications: the initial genomic DNA input into the shearing step was reduced from 3  $\mu\text{g}$  to 10–100 ng in 50  $\mu\text{l}$  of solution. For adaptor ligation, Illumina paired-end adapters were replaced with palindromic forked adapters (purchased from Integrated DNA Technologies) with unique 8-base index molecular barcode sequences included in the adaptor sequence to facilitate downstream pooling. With the exception of the palindromic forked adapters, all reagents used for end repair, A-base addition, adaptor ligation, and library enrichment PCR were purchased from KAPA Biosciences in 96-reaction kits. In addition, during the post-enrichment solid phase reversible immobilization bead cleanup, elution volume was reduced to 20  $\mu\text{l}$  to maximize library concentration, and a vortexing step was added to maximize the amount of template eluted from the beads. Libraries with concentrations above 40 ng  $\mu\text{l}^{-1}$ , as measured by a PicoGreen assay automated on an Agilent Bravo instrument, were considered acceptable for hybrid selection and sequencing.

**Solution-phase hybrid selection.** The exon capture procedure was performed as previously described<sup>32</sup>: before hybridization, any libraries with concentrations >60 ng  $\mu\text{l}^{-1}$  (as determined by PicoGreen) were brought to 60 ng  $\mu\text{l}^{-1}$ , and 8.3  $\mu\text{l}$  of library was combined with blocking agent, bait, and hybridization buffer. Libraries with concentrations between 50 and 60 ng  $\mu\text{l}^{-1}$  were normalized to 50 ng  $\mu\text{l}^{-1}$ , and 10.3  $\mu\text{l}$  of library was combined with blocking agent, bait, and hybridization buffer. Libraries with concentrations between 40 and 50 ng  $\mu\text{l}^{-1}$  were normalized to 40 ng  $\mu\text{l}^{-1}$ , and 12.3  $\mu\text{l}$  of library was combined with blocking agent, bait, and hybridization buffer. Finally, the hybridization reaction was reduced to 17 h; with no changes to the downstream capture protocol.

**Preparation of libraries for cluster amplification and sequencing.** After post-capture enrichment, libraries were quantified using PicoGreen, normalized to equal concentration using a Perkin Elmer MiniJanus instrument, and pooled by equal volume on the Agilent Bravo platform. Library pools were then quantified using quantitative PCR (KAPA Biosystems) with probes specific to the ends of the adapters; this assay was automated using Agilent's Bravo liquid handling platform. On the basis of quantitative PCR (qPCR) quantification, libraries were brought to 2 nM and denatured using 0.2 N NaOH on the Perkin-Elmer MiniJanus. After denaturation, libraries were diluted to 20 pM using hybridization buffer purchased from Illumina.

**Cluster amplification and sequencing.** Cluster amplification of denatured templates was performed according to the manufacturer's protocol (Illumina), HiSeq v3 cluster chemistry and flowcells, as well as Illumina's Multiplexing Sequencing Primer Kit, were used. DNAs were added to flowcells and sequenced using the HiSeq 2000 v3 Sequencing-by-Synthesis method, then analysed using RTA v1.12.4.2 or later. Each pool of whole exome libraries was subjected to paired 76 bp runs. An 8-base index-sequencing read was performed to read molecular indices, across the number of lanes needed to meet coverage for all libraries in the pool.

**Alignment/assembly.** Exome sequence data processing was performed using established analytical pipelines at the Broad Institute. A BAM file was produced with the Picard pipeline (<http://picard.sourceforge.net/>), which aligns the tumour and normal sequences to the hg19 human genome build using Illumina sequencing reads. The BAM was uploaded into the Firehose pipeline (<http://www.broadinstitute.org/cancer/cga/Firehose>), which manages input and output files to be executed.

**Sequencing quality control.** Quality control modules within Firehose were applied to all sequencing data for comparison of the origin for tumour and normal genotypes and to assess fingerprinting concordance. Cross-contamination of samples was estimated using ContEst<sup>33</sup>. Samples with greater than 5% contamination or tumour purity less than 15% were excluded. All BAM files for this cohort are shared in dbGaP (phs000923.v1.p1). BAM files from the ref. 5 cohort were obtained through the European Genome-Phenome Archive (EGAS00001001084).

**Mutation calling and significance analysis.** MuTect<sup>34</sup> was applied to identify somatic single-nucleotide variants. Strelka<sup>35</sup> was applied to identify small insertions or deletions. Artefacts introduced by DNA oxidation during sequencing were computationally removed using a filter-based method<sup>36</sup>. Mutations with allelic fractions of less than 0.05 excluded. Annotation of identified variants was done using Oncotator<sup>37</sup>. Mutational significance analysis was performed using MutSigCV<sup>38</sup>. Manual review of mutations in putative significantly mutated genes was subsequently performed using Integrated Genomics Viewer<sup>39</sup>.

To compare mutations between distinct samples from the same patient, we used a previously described method<sup>17</sup> designed to recover evidence for mutations called in one sample in all other samples derived from the same individual. In brief, the 'force-calling' method uses the strong prior of the mutation being present in at least one sample in the patient to more sensitively detect and recover mutations that might otherwise be missed.

**Copy number and phylogenetic analysis.** Copy-ratio profiles were inferred using ReCAPSEG (<http://gatkforums.broadinstitute.org/gatk/categories/recapseg-documentation>). Read depth at capture probes in tumour samples was normalized using a panel of normal samples to model noise and other biases. The resulting normalized copy ratios are then segmented using the circular binary segmentation algorithm<sup>40</sup>. These data were then transformed into allelic copy number data via integration of allele fraction data from informative germline SNPs. Allelic copy number data allow for inference of contribution of each homologous allele to observed copy number shifts as well as identification of copy neutral events<sup>17,18</sup>. Finally allelic copy number data was integrated with data from point mutations and short deletions and insertions as input to ABSOLUTE<sup>22</sup>.

The ABSOLUTE algorithm was used as previously described to generate purity and ploidy solutions for each tumour samples (Supplementary Table 6). These solutions provided estimates of total allelic copy number and cancer cell fraction for mutations and copy number events in the tumour samples. ABSOLUTE solutions were selected using manual curation. Using methods previously described<sup>23</sup>, we performed allelic deconstruction using ABSOLUTE copy number data. In brief, deconstructions are performed using a Bayes approach in which likelihood of event progressions are defined probabilistically. We defined copy number events as arm level if the event spanned at least 80% of that arm and affected at least one allele. Events were described as reciprocal if homologous chromosomes were amplified and deleted. We used the allelic copy number deconstructions to draw phylogenetic trees. A copy number event was considered shared between two related samples if each of those samples contained an event whose start points fell within two probes of one another, whose end points fell within two probes of one another, and whose event amplitude was identical. Arm lengths are proportional to number of events delineating the samples on the ends of the branch.

We performed a permutation test of arm level deletions to determine the empirical significance of the observed rate of reciprocal deletions in our data set as previously described<sup>22</sup>. In brief, arm level deletions were shuffled between arms while keeping total deletions per sample constant. For each permutation the number of reciprocally deleted chromosome arms were counted. In 10,000 iterations we did not observe the number of reciprocal events in our data set.

**Whole transcriptome sequencing and analysis: cDNA library construction.** Total RNA was assessed for quality using the Caliper LabChip GX2. The percentage of fragments with a size greater than 200 nucleotides (DV200) was calculated using software. An aliquot of 200 ng of RNA was used as the input for first strand cDNA synthesis using Illumina's TruSeq RNA Access Library Prep Kit. Synthesis of the second strand of cDNA was followed by indexed adaptor ligation. Subsequent PCR amplification enriched for adapted fragments. The amplified libraries were quantified using an automated PicoGreen assay. Approximately 200 ng of each cDNA library, not including controls, was combined into 4-plex pools. Capture probes that target the exome were added, and hybridized for recommended time. After hybridization, streptavidin magnetic beads were used to capture the library-bound probes from the previous step. Two wash steps effectively remove any non-specifically bound products. These same hybridization, capture and wash steps are repeated to assure high specificity. A second round of amplification enriches the captured libraries. After enrichment the libraries were quantified with qPCR using the KAPA Library Quantification Kit for Illumina Sequencing Platforms and then pooled equimolarly. The entire process is in 96-well format and all pipetting is done by either Agilent Bravo or Hamilton Starlet.

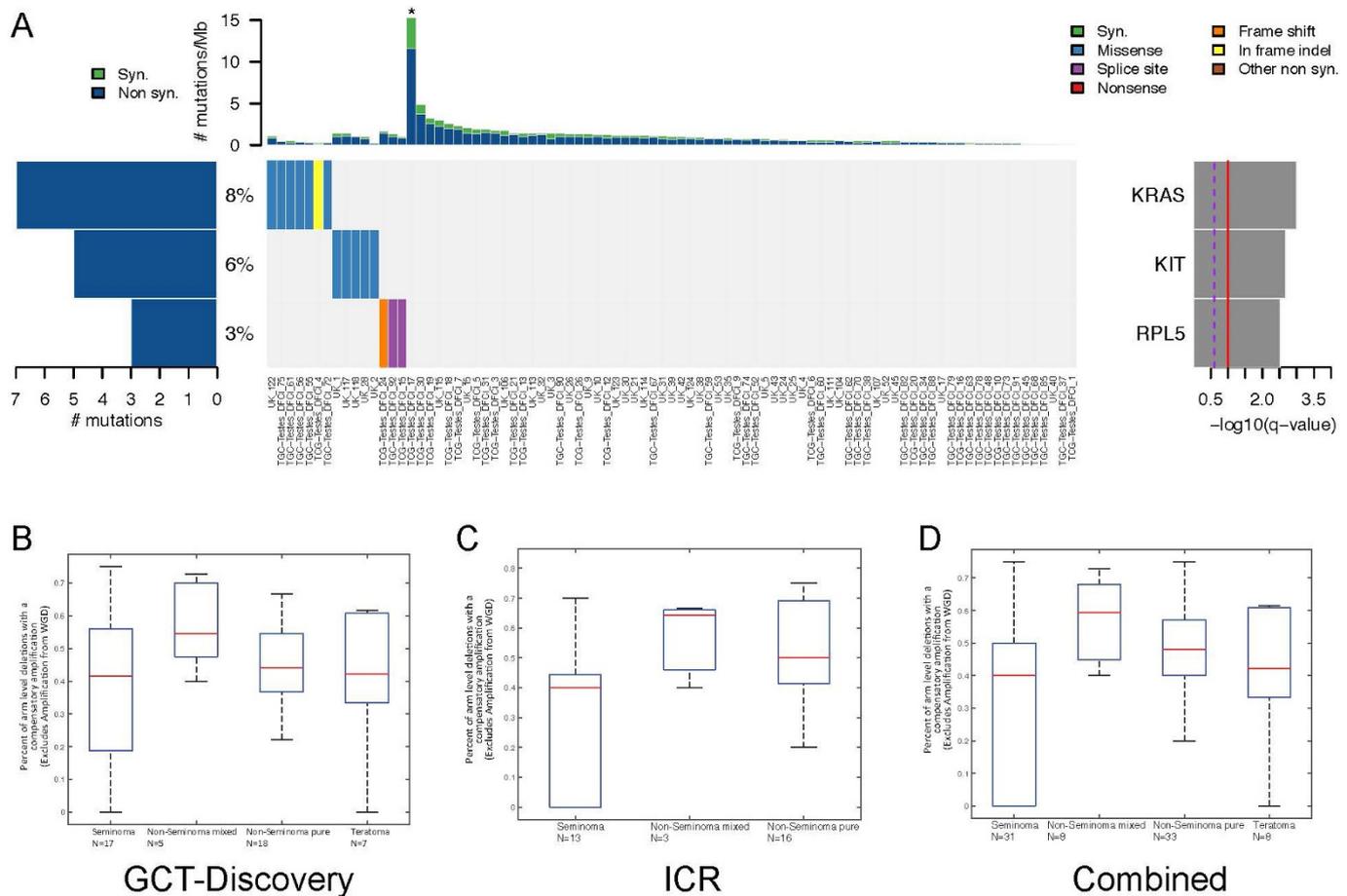
**Illumina sequencing.** Pooled libraries were normalized to 2 nM and denatured using 0.1 N NaOH before sequencing. Flowcell cluster amplification and sequencing were performed according to the manufacturer's protocols using HiSeq 2500. Each run was a 76-bp paired-end with an 8-base index barcode read. Data was analysed using the Broad Picard Pipeline which includes de-multiplexing and data aggregation.

**Analysis.** All BAMs were deduplicated using Samtools<sup>41</sup> before gene expression quantification. Gene expression quantification was performed using RSEM<sup>42</sup> to generate transcripts per million. Comparison of gene expression was performed for *POU5F1* and *NANOG* with Mann–Whitney test.

**BH3 profiling.** Normal and cancer testicular tissue was obtained from the Brigham and Women's Hospital Pathology Department (M.H.). Normal and tumour tissue was immediately processed. Tissue was chopped with a razor blade for 2 min until the tissue was a fine paste. It was then digested in trypsin with 10 mg ml<sup>-1</sup> collagenase IV and 10 U ml<sup>-1</sup> DNase I. The tissue was then incubated in digestion buffer at 37°C for 21 min, vortexing hard every 7 min. The tissue was resuspended in full media (DMEM supplemented with 10% FBS, 2 mM L-glutamine, 100 U ml<sup>-1</sup> penicillin, and 100 µg ml<sup>-1</sup> streptomycin) and passed through a 70 µM filter to generate a single-cell suspension. The cells were spun down, washed once with PBS, and red blood cells were lysed using RBC lysis buffer. Cells were washed again with PBS, counted, and then used for BH3 profiling analysis<sup>27,43</sup>.

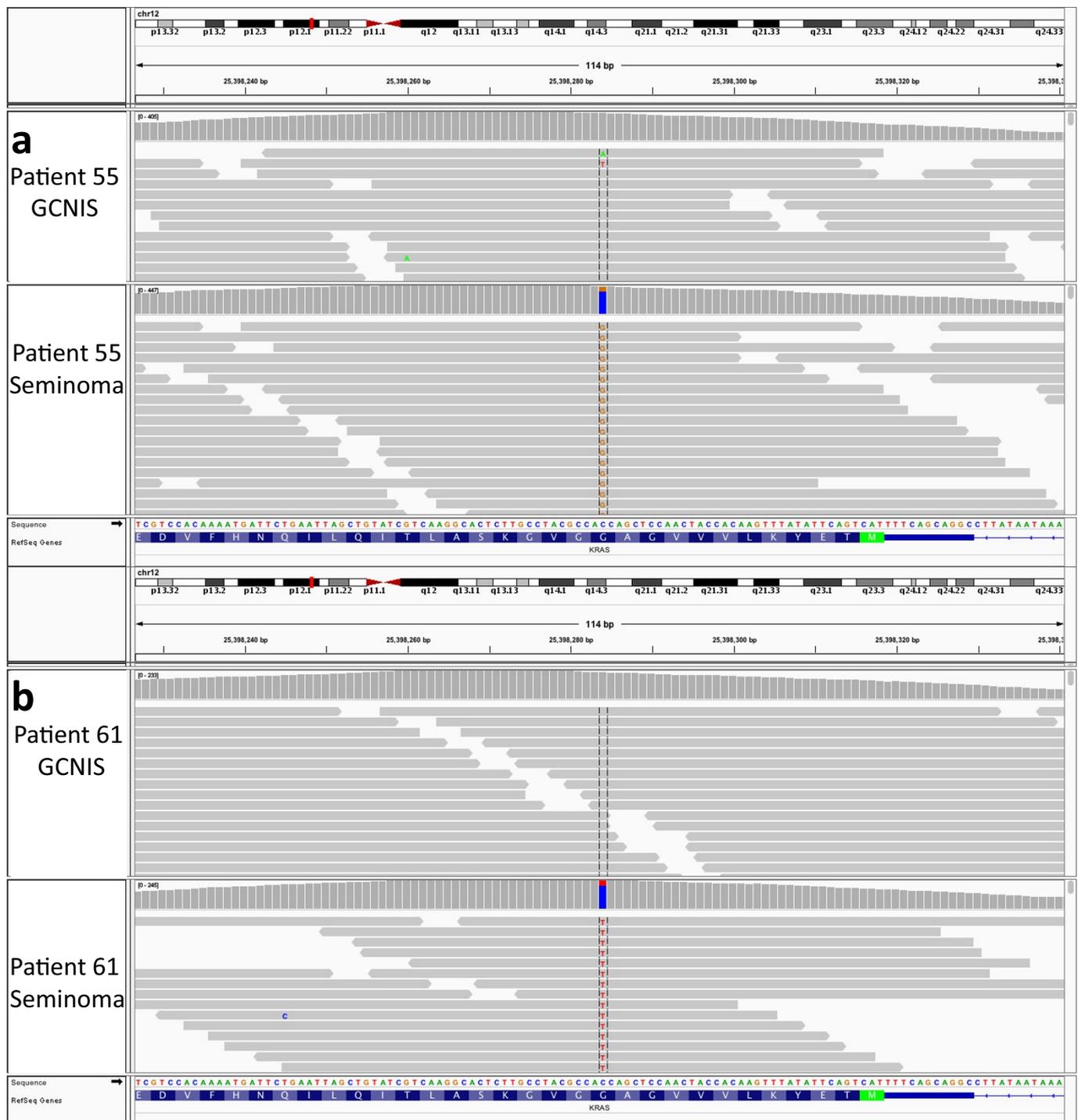
**Data availability statement.** BAM files are deposited in dbGaP (phs000923.v1.p1).

32. Van Allen, E. M. *et al.* Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. *Nat. Med.* **20**, 682–688 (2014).
33. Cibulskis, K. *et al.* ContEst: estimating cross-contamination of human samples in next-generation sequencing data. *Bioinformatics* **27**, 2601–2602 (2011).
34. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
35. Saunders, C. T. *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).
36. Costello, M. *et al.* Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res.* **41**, e67 (2013).
37. Ramos, A. H. *et al.* Oncotator: cancer variant annotation tool. *Hum. Mutat.* **36**, E2423–E2429 (2015).
38. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
39. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
40. Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**, 557–572 (2004).
41. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
42. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
43. Touzeau, C. *et al.* BH3 profiling identifies heterogeneous dependency on Bcl-2 family members in multiple myeloma and predicts sensitivity to BH3 mimetics. *Leukemia* **30**, 761–764 (2016).



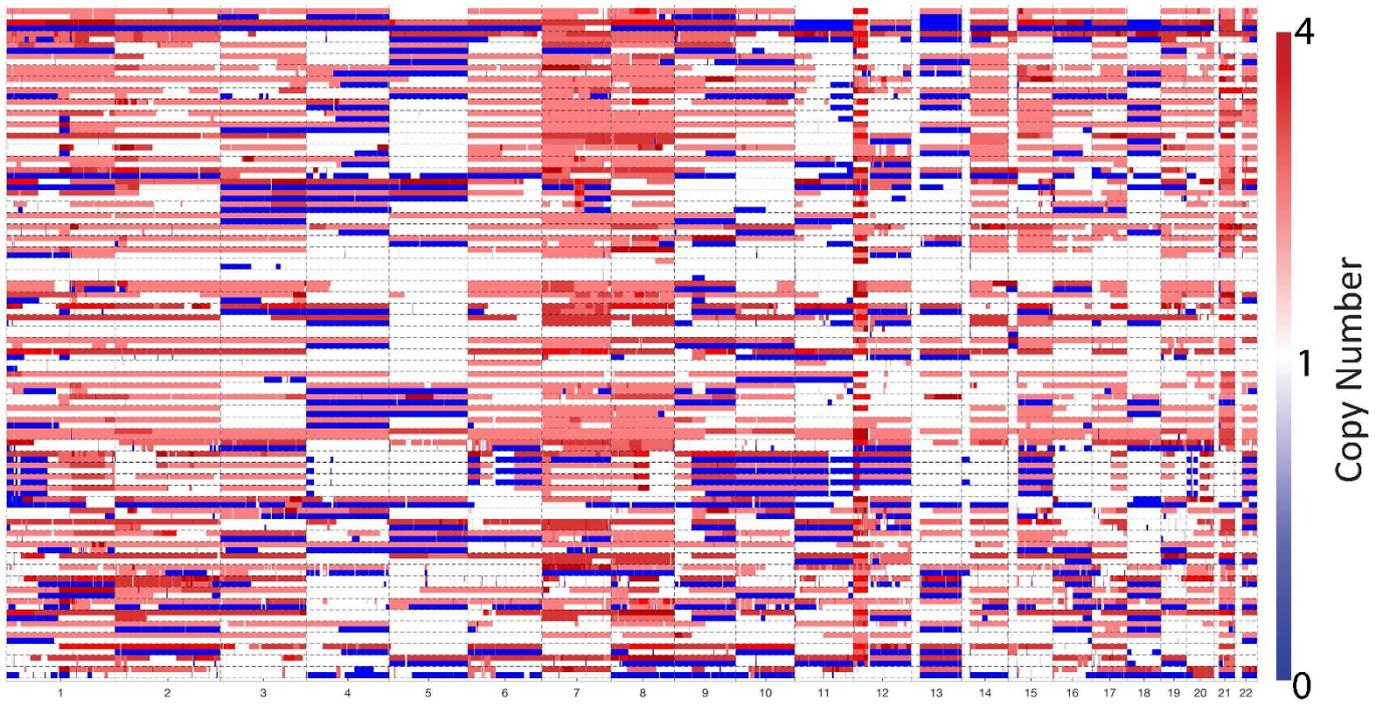
**Extended Data Figure 1 | Mutational significance and copy number meta-analysis.** **a**, Mutational significance meta-analysis of discovery and ICR cohorts identify *KRAS*, *KIT* and *RPL5* as significantly mutated in TGCT, with a spectrum of mutation rates. In this plot, each column

represents a patient WES. Asterisk denotes the hypermutated PMGCT (DFCI\_17). **b**, RLOH distribution by histology in discovery cohort. **c**, RLOH distribution by histology in ICR cohort. **d**, RLOH distribution in the meta-analysis, consistent with both subsets.



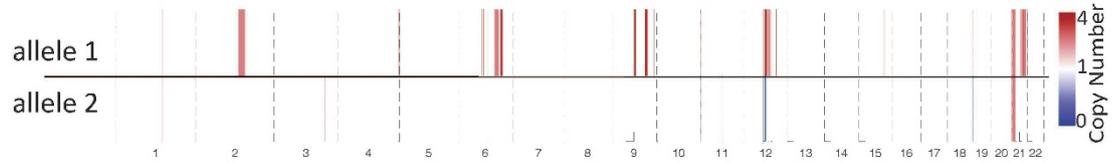
**Extended Data Figure 2 | Genomic reads for KRAS loci in two patient cases. a,** Integrative genomics viewer snapshot of KRAS p.G12A mutation in DFCI\_55 GCNIS and seminoma. The mutation is present in the primary tumour but absent from the GCNIS. **b,** Integrative genomics

viewer snapshot of KRAS p.G12A mutation in DFCI\_61 GCNIS and seminoma. The mutation is present in the primary tumour but absent from the GCNIS.

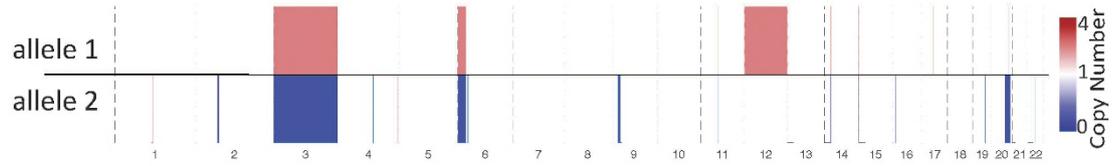


**Extended Data Figure 3 | Allelic copy number heat map of the discovery cohort.** Each tumour sample is a row, and chromosomes are listed as columns. Blue regions note deletions, and red regions denote amplifications.

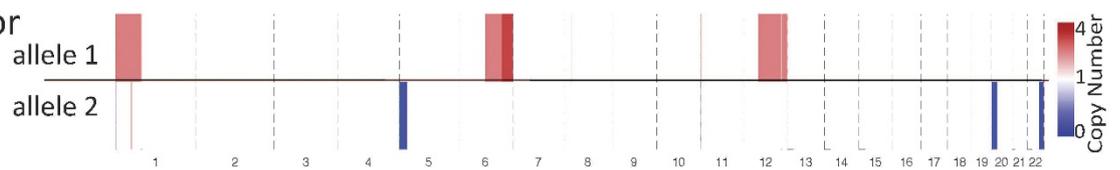
Liposarcoma  
DFCI 65



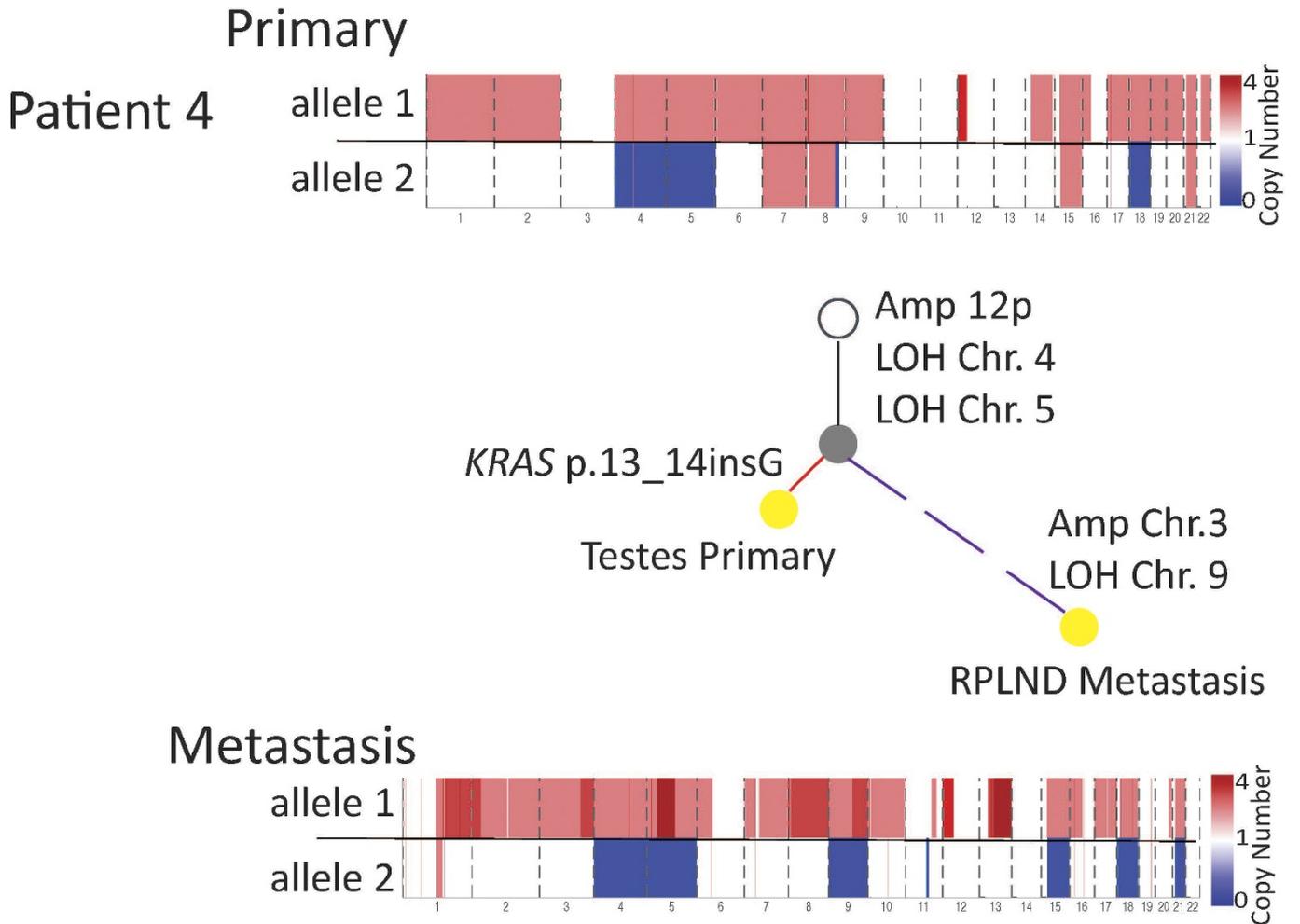
DLBCL  
DFCI 42



Sertoli cell tumor  
DFCI 39



**Extended Data Figure 4 | Testes tumours of different cell types.** Allelic copy number data from testes tumours of different cell types are shown. These three tumours do not contain the same level of arm level chromosomal events as GCTs.



**Extended Data Figure 5 | Phylogenetic analysis of DFCI\_4.** Histology proportion is indicated by pie charts within each phylogenetic tree. Phylogenetic trees were constructed using allelic copy number deconstructions. Branch lengths are proportional to the number of

deconstructed copy number events. Branches leading to primary samples are red, and branches leading to metastases are purple. The dotted branch indicates deconstructions which may be impacted by FFPE sample degradation, limiting discrete branch length estimation.