Novel somatic and germline mutations in intracranial germ cell tumors

Linghua Wang[1], **Shigeru Yamaguchi**[2], **Matthew D. Burstein**[3,4], **Keita Terashima**[2,5], **Kyle Chang**[1], **Ho-Keung Ng**[6], **Hideo Nakamura**[7], **Zongxiao He**[8], **Harshavardhan Doddapaneni**[1], **Lora Lewis**[1], **Mark Wang**[1], **Tomonari Suzuki**[9], **Ryo Nishikawa**[9], **Atsushi Natsume**[10], **Shunsuke Terasaka**[11], **Robert Dauser**[12], **William Whitehead**[12], **Adesina Adekunle**[13], **Jiayi Sun**[4], **Yi Qiao**[14], **Gábor Marth**[14], **Donna M. Muzny**[1], **Richard A. Gibbs**[1], **Suzanne M. Leal**[8], **David A. Wheeler**[1], and **Ching C. Lau**[2,4,15]

[1]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA

[2]Texas Children's Cancer and Hematology Centers, Baylor College of Medicine, Houston, TX 77030, USA

[3]Structural and Computational Biology and Molecular Biophysics Program, Baylor College of Medicine, Houston, TX, USA

[4]Medical Scientist Training Program, Baylor College of Medicine, Houston, TX, USA

[5]National Center for Child Health and Development, Tokyo, Japan

[6]Dept of Pathology, Chinese University of Hong Kong, Hong Kong, China

[7]Dept of Neurosurgery, Kumamoto University, Kumamoto, Japan

[8]Center for Statistical Genetics, Dept of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA

[9]Dept of Neurosurgery, Saitama Medical University, Saitama, Japan

[10]Dept of Neurosurgery, Nagoya University, Nagoya, Japan

[11]Dept of Neurosurgery, Hokkaido University, Hokkaido Prefecture, Japan

[12]Dept of Neurosurgery, Baylor College of Medicine, Houston, TX 77030, USA

[13]Dept of Pathology, Baylor College of Medicine, Houston, TX 77030, USA

[14]Department of Biology, Boston College, Chestnut Hill, MA 02467 USA

[15]Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, TX 77030, USA

Abstract

Intracranial germ cell tumors (IGCTs) are a group of rare heterogeneous brain tumors which are clinically and histologically similar to the more common gonadal GCTs. IGCTs show great variation in their geographic and gender distribution, histological composition and treatment outcomes. The incidence of IGCTs is historically 5–8 fold greater in Japan and other East Asian countries than in Western countries[1] with peak incidence near the time of puberty[2]. About half of the tumors are located in the pineal region. The male-to-female incidence ratio is approximately 3–4:1 overall but even higher for tumors located in the pineal region[3]. Due to the scarcity of tumor specimens available for research, little is currently known about this rare disease. Here we report the analysis of 62 cases by next generation sequencing, SNP array and expression array. We find the KIT/RAS signaling pathway frequently mutated in over 50% of IGCTs including novel recurrent somatic mutations in *KIT*, its downstream mediators *KRAS* and *NRAS*, and its negative regulator *CBL*. Novel somatic alterations in the AKT/mTOR pathway included copy number gain of the *AKT1* locus at *14q32.33* in 19% of patients, with corresponding upregulation of AKT1 expression. We identified loss-of-function mutations in *BCORL1*, a transcriptional corepressor and tumor suppressor. We report significant enrichment of novel and rare germline variants in *JMJD1C*, a histone demethylase and coactivator of the androgen receptor, among Japanese IGCT patients. This study establishes a molecular foundation for understanding the biology of IGCTs and suggests potentially promising therapeutic strategies focusing on the inhibition of KIT/RAS activation and the AKT1/mTOR pathway.

IGCT are divided into two main groups, pure germinoma and nongerminomatous germ cell tumors (NGGCTs). Germinoma is the most common subtype[1], accounting for over two thirds of all IGCTs[1,4]. NGGCTs include teratoma, embryonal carcinoma, yolk sac tumor and choriocarcinoma. About 10% of germinomas and most NGGCTs remain refractory to multimodality therapy[4]. Previously the biology of these tumors is largely unknown except for gain-of-function mutations of *KIT* reported in ~25% of pure germinomas[5,6] and a few chromosomal abnormalities revealed by comparative genomic hybridization (CGH)[7–9]. Through an international multicenter collaboration, we have conducted an in-depth analysis of the genetic abnormalities of IGCTs. For the discovery study, whole-exome sequencing (WES) of 28 cases yielded an average of 139x coverage with 95.4% of targeted bases covered by ≥20x (Extended Data Figure 1). We validated a mean of 6 non-silent mutations per sample (Supplementary Table 1), corresponding to approximately 0.50 non-silent mutations per megabase (Mb, Extended Data Figures 2a–b). Although there was no significant difference in average mutation rate between pure germinomas and NGGCTs, the mutation rates varied dramatically among NGGCTs (Extended Data Figure 2c). For the validation study, we performed targeted deep sequencing, average depth of coverage, ~1000x, for an additional 34 IGCT cases using a custom-designed AmpliSeq array (**Online Methods**). The identified mutations were listed in Supplementary Table 2.

The recurrent genetic alterations and clinical data are summarized in Figures 1 and 2. Except for *KIT*, none of these genes were previously linked to IGCTs. Overall, 53% of the tumors harbored somatic mutations in at least one of the genes involved in KIT/RAS or AKT/mTOR pathways (Figures 3a and 3b). 93% of somatic mutations identified in these pathways were predicted to be deleterious (Supplementary Table 3).

Oncogenic *KIT* mutations are common in testicular seminomas[10] and gastrointestinal stromal tumors (GISTs)[11], resulting in ligand-independent kinase activation[12]. *KIT* was mutated in 16 IGCT tumors (Figure 2a) but not in any NGGCT cases. Mutations in *KIT* were clustered primarily in exon 17, followed by exon 11 in a pattern similar to that of testicular seminomas[10], but quite different from GISTs[11] in which mutations cluster in Exon 11, followed by the extracellular domain (Figure 2a). *KIT* was over-expressed in the majority of pure germinomas (Extended Data Figure 3) but rarely in NGGCTs. So far, eight tyrosine kinase inhibitors (TKIs) targeting *KIT* have been approved (Supplementary Table 4). Since the long-term side effects of radiation are well recognized, the application of TKIs could potentially benefit IGCT patients through dose reduction or elimination of radiation therapy.

KRAS or *NRAS* were mutated in 19% of IGCTs cases (Extended Data Figure 4a). The *KRAS/NRAS* and *KIT* mutations were mutually exclusive genetic events in IGCTs (Figure 3a, Extended Data Figure 4b, P=0.018). So far, no approved targeted therapies are available for cancers with *KRAS* mutations. A recent preclinical study of the *MEK* inhibitor, Selumetinib, indicated great efficacy against *KRAS*-mutated non-small-cell lung cancer cells[13]. Another study showed effectiveness of *ERK* inhibitors in *MEK* inhibitor-resistant cells[14].

CBL, encoding a RING finger ubiquitin E3 ligase, was the third most frequently mutated gene in IGCTs (Figure 2b). The protein CBL has been shown to function as a negative regulator of receptor protein tyrosine kinases (RPTKs) including *KIT* by targeting activated receptors for polyubiquitination and subsequent degradation[15]. *CBL* mutation is associated with KIT over-expression, aberrant expression of phosphorylated *STAT5*, and poor prognosis in myeloid malignancies[16]. In this study, 3 somatic mutations were identified in the RING finger domain, which might abolish *CBL*-directed polyubiquitination and downregulation of RPTKs[15] and 2 recurrent mutations were found in the small linker domain, which are supposed to be oncogenic[15] (Figure 2b). *KIT* was over-expressed in tumor G23 that was wildtype for *KIT* but had a somatic mutation in *CBL* (Extended Data Figure 3), consistent with its role as a negative regulator of *KIT*. A recent report showed that Dasatinib is the most effective TKI against myeloid leukemia cell lines that harbor homozygous *CBL* mutations[17], and therefore might be an effective agent for the treatment of *CBL*-mutated IGCTs. We further analyzed the copy-number status of *CBL* and found that 13 out of 28 cases had either clonal or subclonal 11qLOH spanning the *CBL* locus (Supplementary Table 5 and Extended Data Figure 5a). Tumors G4 and G11 with *CBL* somatic mutation lost the wild-type allele through clonal and subclonal 11qLOH, respectively (Extended Data Figure 5b) further reinforcing the notion that *CBL* might play an important role in the pathogenesis of IGCTs.

We observed focal amplification of 14q32.33 in 5 tumors (18%) in the discovery set (Extended Data Figure 6a). The amplified region included 2 Mbp spanning the *AKT1* locus. Among 34 genes within this region, *AKT1* was the only known oncogene. We validated *AKT1* copy number gains by qPCR for all 5 cases and 7 additional cases in the validation set. Copy number gains at 14q32.33 were associated with elevated mRNA expression of *AKT1* (Figure 3c) but not for other genes in this region (Supplementary Table 6 and Extended Data Figure 6b). We also observed strong cytoplasmic and nuclear immunostaining of AKT1 in cases with focal *AKT1* copy number gains (Figure 3d). In this study, 75% cases with *AKT1* copy number gains were identified in tumors with wildtype *KIT, KRAS* and *NRAS* (Figure 3a), suggesting an important role of AKT/mTOR pathway in IGCTs. The serine-threonine protein kinase AKT is a key intermediate of signaling pathways that regulate cellular processes. AKT activation by mutation is frequently seen in human cancers but overexpression is rarely reported. However, high-level *AKT1* expression was reported in non-small cell lung cancer where it was correlated with poor response to cisplatin[18]. Use of *AKT1* inhibitors might also be a promising strategy for treating IGCT patients with *AKT1* over-expression.

Finally, we identified recurrent somatic mutations in *BCORL1*, *MTOR, TP53, SPTA1, KDM2A* and *LAMA4* (Figure 1). *BCORL1*, a transcriptional corepressor located on the X-chromosome (*Xq25-q26.1*), was thought to be a tumor suppressor gene in AML[19]. Functional studies have shown that *BCORL1* can interact with class II histone deacetylases, the CtBP corepressor, through the CtBP binding motif-PLDLS and affect the repression of E-cadherin[20]. In this study, *BCORL1* was mutated in 6 male patients (Figure 2b). Interestingly, in 5 of them, mutations were out-of-frame insertions/deletions, potentially resulting in a truncated protein lacking the conserved LXXLL motif, which mediates the hormone-induced interaction between nuclear receptor (NR) and co-activators[21].

Given the low mutation rate and early onset of this disease, we sought to identify potential IGCT-predisposing germline mutations. Therefore we screened germline sequence data for genes enriched in novel functional variants in the IGCT discovery cohort compared to normal populations (Extended Data Figure 7 and Supplementary Table 7). The top four genes most enriched in functional germline variants were *CDK5RAP2*, *JMJD1C*, *USP35* and *PCDH15*. Of these, *JMJD1C* functions as a chromatin modifier gene, and was the only one with a reported role in normal germinal tissue development in both humans[22] and mouse models[23]. A recent report showed that it is required in the development and maintenance of germ cells in mice[24]. Thus we included this gene for mutation screening in the validation cohort. In the combined cohorts of 62 cases, we identified novel and rare germline variants in *JMJD1C* in 10 patients including a rare dinucleotide polymorphism (AA to GC, S880P) in 3 genetically unrelated individuals (Figures 2b and Supplementary Table 8). Among the 10 patients, 9 carriers were from Japan and 1 was from Hong-Kong. *JMJD1C* germline variants were significantly enriched in the Japanese populations in control cohorts and further enriched (about 5 fold) in Japanese IGCT patients (Figure 4). The odds ratio is 4.8, indicating a strong association between *JMJD1C* variants and the risk of developing IGCT. The rare variant association tests[25,26] also revealed significant association of *JMJD1C* genotype with this disease (Supplementary Table 9). Like *BCORL1*, there are two conserved

LXXLL NR-interacting motifs at the C-terminal of *JMJD1C*, which mediate the interaction with nuclear receptors. Indeed, *JMJD1C* is known to interact with the thyroid hormone receptor in the thyroid[27] and was reported to interact with the androgen receptor (AR) in human[28]. Expression microarrays revealed high *JMJD1C* and *AR* expression levels in all 37 ICGT patients tested (Extended Data Figure 8).

DNA copy-number and clonality analysis (**Online Methods**) of IGCT tumors using SNP array data, revealed a heterogeneous and complex tumor genome. Nearly half of the tumors were tetraploid and 71% of the analyzed cases showed subclonal structure (Extended Data Figures 9 and 10). We observed almost exclusively chromosomal arm-level gains and losses. Chromosome X, *21q, 12p, 1q* and *14q* were frequently amplified; whereas 11q, 10q, 17p and 13q were frequently deleted. Only copy number gains of X, 12p and loss of 13q have been previously described[29]. Over 90% of the pure germinomas exhibited chromosomal imbalance and acquired uniparental disomy of multiple chromosomes whereas 40% of the NGGCTs were chromosomally stable. The preponderance of arm-level and chromosome-level imbalance suggests the possibility that aberrant meiotic division may play a role in the pathogenesis of pure germinomas.

Although pure germinomas respond well to surgery and radiotherapy, the long-term quality of life for about 1/3 of postoperative patients is severely impaired due to radiotherapy-associated complications. In addition, nearly 10% of patients develop recurrence even after multimodality therapy[1,4]. Except for mature teratomas, most NGGCTs remain refractory to available treatments and have a rather poor prognosis[1,4]. In this study we identified novel and frequent somatic alterations in the KIT/RAS and AKT/mTOR signaling pathways and suggest potentially promising targets for the development of novel therapeutic strategies focusing on inhibition of these pathways. Although fewer therapeutic targets were found in NGGCTs, frequent *AKT1* amplification and recurrent mTOR mutations may open the door to possible use of AKT1/mTOR inhibitors. It is intriguing that both histone-modifying genes discovered in this study are implicated in the interaction with nuclear receptor proteins. It suggests the possibility that *JMJD1C* and *BCORL1* might be associated with the male preponderance and age of peak incidence of IGCTs through interaction with *AR* triggered by elevated levels of androgen at puberty. Further functional studies using mouse models are needed to elucidate the possible role of *JMJD1C* in the pathogenesis of this disease. This study comprehensively described the genomic heterogeneity and complexity underlying IGCTs and provides a molecular foundation for understanding of the genetic alterations in IGCT.

ONLINE METHODS

Patients

The collection of samples was performed at multiple medical institutes in US, Japan and Hong Kong including Texas Children's Hospital (USA, *n* = 5), Saitama Medical University Hospital (Japan, *n* = 28), Kumamoto University Hospital (Japan, *n* = 9), Nagoya University Hospital (Japan, *n* = 11), Hokkaido University Hospital (Japan, *n* = 2), and Prince of Wales Hospital (Hong Kong, *n* = 7). Sample collection was approved by the local Institutional Review Board or institutional ethics committee. Written consent was obtained from all

patients in accordance with the Declaration of Helsinki. All tumor samples were given the histological diagnosis by the neuropathologists in the local hospitals according to World Health Organization criteria. In the discovery cohort, 23 out of 28 tumors were obtained before any adjuvant therapy, and 5 (G6, NG3, NG9, NG11 and NG12) tumors were obtained from residual tissue after or during first-line chemotherapy. All tumor specimens were obtained before radiation therapy except M8 which was a recurrent tumor 6 years after radiation therapy. In the validation cohort, 26 out of 34 tumors were obtained before any treatment, and 7 tumors (NG15, NG17, NG18, NG22, NG24, NG25 and NG26) were obtained after or during first-line chemotherapy, and one tumor (G25) was obtained at recurrence after radiation therapy. Of the 62 cases, 29 were diagnosed as pure germinoma, 25 were diagnosed as NGGCT, and 8 were mixed GCTs with germinoma component.

DNA extraction

In the discovery series, all analyzed tumor DNA were extracted from fresh frozen tissue using QIAamp® DNA mini kit (QIAGEN), and peripheral blood DNAs were isolated using Wizard® Genomic DNA purification kit (Promega). In the validation series, some cases of tumor DNA were extracted from formalin-fixed paraffin-embedded (FFPE) materials using Recover All™ Total Nucleic Acid Isolation Kit (Life Technologies) following the manufacturer's protocol.

Illumina library construction

Illumina libraries were constructed according to the manufacturer's protocol with modifications as described in HGSC website (https://hgsc.bcm.edu/sites/default/files/documents/Illumina_Barcoded_Paired-End_Capture_Library_Preparation.pdf). Libraries were prepared using Beckman robotic workstations (Biomek NXp and FXp models. Briefly, 1 ug of genomic DNA in 100ul volume was sheared into fragments of approximately 300–400 base pairs in a Covaris plate with E210 system (Covaris, Inc. Woburn, MA) followed by end-repair, A-tailing and ligation of the Illumina multiplexing PE adaptors. Pre-capture Ligation Mediated-PCR (LM-PCR) was performed for 7 cycles of amplification using the 2X SOLiD Library High Fidelity Amplification Mix (a custom product manufactured by Invitrogen). Universal primer IMUX-P1.0 and a pre-capture barcoded primer IBC were used in the PCR amplification. In total, a set of 12 such barcoded primers were used on these samples. Purification was performed with Agencourt AMPure XP beads after enzymatic reactions. Following the final XP bead purification, quantification and size distribution of the pre-capture LM-PCR product was determined using the LabChip GX electrophoresis system (PerkinElmer).

Exome capture

Four pre-capture libraries were pooled together (approximately 250 ng/sample, 1 ug per pool) and hybridized in solution to the HGSC VCRome 2.1 design1 (42Mb, NimbleGen) according to the manufacturer's protocol NimbleGen SeqCap EZ Exome Library SR User's Guide (Version 2.2) with minor revisions. Human COT1 DNA and full-length Illumina adaptor-specific blocking oligonucleotides were added into the hybridization to block repetitive genomic sequences and the adaptor sequences. Post-capture LM-PCR amplification was performed using the 2X SOLiD Library High Fidelity Amplification Mix

with 14 cycles of amplification. After the final AMPure XP bead purification, quantity and size of the capture library was analyzed using the Agilent Bioanalyzer 2100 DNA Chip 7500. The efficiency of the capture was evaluated by performing a qPCR-based quality check on the four standard NimbleGen internal controls. Successful enrichment of the capture libraries was estimated to range from a 6 to 9 of ΔCt value over the non-enriched samples. Aliquots of enriched libraries (10 nM) were submitted for sequencing.

Illumina sequencing

Library templates were prepared for sequencing using Illumina's cBot cluster generation system with TruSeq PE Cluster Generation Kits (Part no. PE-401-3001). Briefly, these libraries were denatured with sodium hydroxide and diluted to 3–6 pM in hybridization buffer in order to achieve a load density of ~800K clusters/mm2. Each library pool was loaded in a single lane of a HiSeq flow cell, and each lane was spiked with 1% phiX control library for run quality control. The sample libraries then underwent bridge amplification to form clonal clusters, followed by hybridization with the sequencing primer. Sequencing runs were performed in paired-end mode using the Illumina HiSeq 2000 platform. Using the TruSeq SBS Kits (Part no. FC-401-3001), sequencing-by-synthesis reactions were extended for 101 cycles from each end, with an additional 7 cycles for the index read. Real Time Analysis (RTA) software was used to process the image analysis and base calling. Sequencing runs generated approximately 300–400 million successful reads on each lane of a flow cell, yielding 9–10 Gb per sample. With these sequencing yields, samples achieved an average of 95% of the targeted exome bases covered to a depth of 20X or greater.

WES data processing and quality control

Exome sequence data processing and analysis were performed using the standard pipelines established at Human Genome Sequencing Center (HGSC) of Baylor College of Medicine[31]. Read sequences were mapped to the human reference genome (GRCh37) by BWA[32]. All BAM files were processed to identify duplicates using the Picard and then recalibrated and realigned by GATK[33]. Quality control modules were used to compare genotypes derived from Affymetrix arrays and sequencing data to ensure concordance. Genotypes from SNP arrays were also used to monitor for low levels of cross-contamination between samples from different individuals. The Ion Torrent sequencing data was analyzed using Torrent Suite Software v3.0.

WES data Mutation calling and annotation

Atlas-SNP2 was used to identify somatic single-nucleotide variants in targeted exons[34]. Pindel was also applied to call small-to-medium size of insertions and deletions[35]. A minimum of 4 high-quality supporting reads and a minimum mutant allele fraction of 0.08 was required for mutation calling. Somatic mutations and germline variants were annotated using information from publicly available databases, including dbSNP build 135 and variants from the 1000 Genome project. The pipeline integrates ANNOVAR[36] to determine whether the observed amino acid changes have synonymous, non-synonymous, nonsense, or splice-site changing properties on the encoded proteins and SIFT[37], PROVEAN (Protein Variation Effect Analyzer)[38] and Polyphen-2[39] algorithms to predict the functional impact of somatic point mutations. Variants were further annotated with entries from the Catalogue

of Somatic Mutation version 52 (COSMIC, http://www.sanger.ac.uk/genetics/CGP/cosmic/) to determine if the mutations were detected at previously reported hotspots.

Mutation validation

All somatic non-silent mutations identified and a subset of functionally interesting germline variants were selected for validation using the Ion Torrent Personal Genome Machine (PGM, Life Technologies Corporation). PCR primers were designed using standard Primer3 tool and the best primer pairs were selected using our on-house algorithm. PCR amplicons from the tumors and their matched blood samples were barcoded, pooled, sheared by enzymatic digestion, adaptor ligated, size selected, amplified and sequenced by Ion PGM system. The sequencing data was processed using Torrent Suite Software v3.0. The average read depth per base was 2533x and 1923x for the tumor and blood pools, respectively. A minimum of 50 high-quality supporting reads and a minimum mutant allele fraction of 0.05 was required to define a validated mutation. Furthermore, we chose 103 recurrent non-silent germline variants (allelic fraction ≥0.3, observed in both tumor and normal) to validate by Sanger sequencing (primer pairs provided on request).

Gene selection for mutation prevalence screening

Genes were selected based on the following criteria: 1) genes with somatic mutations occurring in two or more patients; 2) Genes with somatic mutation identified only in one patient but involved in KIT/RAS or AKT/mTOR pathways which were significantly mutated in IGCT patients (Supplementary Table 11), interacts with *KIT*, tumor suppressor genes which lost the wild-type allele or mutated at COSMIC reported codons; 3) Genes without somatic mutation but involve in KIT/RAS or AKT/mTOR pathways and with recurrent DNA copy-number changes or have important contribution, with recurrent novel germline variants or reported by previous GWAS studies. Totally, 23 potentially interesting genes were selected based on the results of the discovery study.

Designing of AmpliSeq custom array

The coding exons of 23 selected genes were extracted using UCSC table browser (hg19) and submitted to Ion AmpliSeq Designer using pipeline version 1.2 using settings for standard DNA and an amplicon range 125–225bp. The resulting custom design consists of a total of 916 amplicons. The average amplicon size was 195bp. On average, 98.5% of the coding sequences of targeted genes were covered and the total length of covered bases is 182,447 bp (Supplementary Table 10). The primers were binned into two pools (Pool #1: 461 amplicons & Pool #2: 455 amplicons) to avoid amplification of undesired targets.

Ion Torrent library construction

Ion Ampliseq library kit 2.0 (Cat#4480441, Life Technologies) consisting of Ampliseq PCR and library preparation reagents was used to prepare template DNA for sequencing. Ampliseq reactions were performed separately for Pool 1 and pool 2 for each sample. Each Ampliseq reaction was set up using 10 ng of DNA as input. Thermo cycling conditions included, initial denaturation for 2 min. at 99°C followed by 16 annealing and extension cycles of 15 s at 99°C and 4 min. at 60°C. Libraries were prepared using Beckman robotic

workstations. Following the Ampliseq reaction, 2ul of FuPa reagent was added to remove PCR adaptor regions and repair fragment ends. Ion Xpress™ Barcode Adapters were then ligated to each pool. The Post-ligation products were purified using Agencourt AMPure XP beads. Thermocycling conditions were initial denaturation for 2 min. at 98°C followed by 7 annealing and extension cycles of 15 s at 98°C and 1 Min. at 60°C. Agencourt XP® beads were used to purify DNA after each reaction step. PCR products were purified using the above SPRI beads followed by quantification and size distribution using the LabChip GX electrophoresis system (PerkinElmer). Four to eight samples (8–16 libraries) were sequenced per run on Ion Torrent PGM instrument.

The library templates were prepared for sequencing using the Life Technologies Ion OneTouch v2 DL protocols and reagents. Briefly, library fragments were clonally amplified onto Ion Sphere Particles (ISPs) through emulsion PCR and then enriched for template-positive ISPs. More specifically, PGM emulsion PCR reactions utilized the Ion OneTouch 200 Template Kit v2 DL (Life Technologies, Part no. 4480285), and as specified in the accompanying protocol, emulsions and amplification were generated using the Ion OneTouch System (Life Technologies, Part no. 4467889). Following recovery, enrichment was completed by selectively binding the ISPs containing amplified library fragments to streptavidin coated magnetic beads, removing empty ISPs through washing steps, and denaturing the library strands to allow for collection of the template-positive ISPs. For all reactions, these steps were accomplished using the Life Technologies ES module of the Ion OneTouch System, and template-positive ISPs were quantified using the Guava EasyCyte 5 (Millipore Technologies), obtaining >90% enrichment efficiency for all reactions. Approximately 20 million template-positive ISPs per run were deposited onto the Ion 318C chips (Life Technologies, Part no. 4469497) by a series of centrifugation steps that incorporated alternating the chip directionality. Sequencing was performed with the Ion PGM 200 Sequencing Kit (Life Technologies, 4474004) using the 440 flow ("200bp") run format.

Ion PGM sequencing data processing and mutation calling

The PGM sequencing data was processed using Ion Torrent Suite Software v3.0. Reads were aligned to the genome using TMAP against human reference genome build 37 (NCBI) with default parameters. Mutations were called using BAM files from the tumor and matched normal samples. Atlas-SNP[34] was run for SNP calling. The variants were further filtered to remove those supported by less than 5 sequencing reads or presented in less than 8% of aligned reads. For indels, the variant allele must be supported by at least 10 sequencing reads. In addition, it is requested that at least one variant had to be Q30 or better and had to lie in the central portion of the read. Besides, reads harboring the variant must have been observed in both forward and reverse orientations.

Filtering of germline variants

Germline variants were filtered step by step to identify the potentially interesting candidates. The detail workflow was described in **Supplementary Figure 12**. In brief, we first selected the non-silent variants including missense, nonsense, frameshift, and splice-site variants. Second, we selected the high-confident variants that meet the following criteria: 1) variant

allele fraction in tumor and normal equal to or greater than 0.20; 2) variant calling were supported by at least 4 sequencing reads for both tumor and normal samples. Third, we selected novel variants that haven't been reported in dbSNP database (dbSNP135). Then, we selected genes with COSMIC evidence, i.e. genes for which mutations have been reported in COSMIC database in at least 100 samples. After that, for each gene in the above list, we calculated the fold of enrichment of the germline variants in Japanese IGCT patients by comparing its frequency to that of Japanese patients in the control cohort and performed Fisher's exact test to calculate the p values (Supplementary Table 7). The potentially interesting genes were then selected based on the IGCT frequency bias (>=4) and significant p values (<0.05).

Genome-wide SNP array

DNA copy number analyses were performed using the high resolution Illumina HumanOmni2.5–8 (Omni2.5) BeadChip Kit (Illumina). In brief, 200ng genomic DNA was first denatured by NaOH. After neutralization of the sample, isothermal whole genome amplification was conducted to uniformly increase the DNA amount. The amplified DNA was enzymatically fragmented and hybridized to BeadChip for 16–24 hours at 48°C. After washing off unhybridized and non-specifically hybridized DNA fragments, allele-specific single-base extension reaction was performed to incorporate labeled nucleotides into the bead-bound primers. Following multi-layer staining to amply signals from the labeled extended primers and final washing and coating, beadchips were imaged using the Illumina iScan system. SNP calls were collected using the Illumina GenomeStudio Version 2011.1 Genotyping Module 1.9.4. For improved CNV analysis, B allele frequencies (BAF) were calculated and log2 R ratios (LRR) were extracted after re-clustering the raw data by applying the GenomeStudio clustering algorithms.

DNA copy number analysis

The BAF, LRR and X/Y intensities were exported using the Final Report tool of the Illumina GenomeStudio Version 2011.1 software and imported into the Circular Binary Segmentation (CBS)[40] and Gene Alteration Print (GAP)[30] pipelines. The allele-specific DNA copy number, ploidy and purity were analyzed by GAP utilizing default parameters (**Supplementary Figure 1**). Recognition of GAP patterns was performed *de novo*.

Clonality analysis

The subclone deconvolution consisted of several steps. First, the whole genomic BAF data of the tumor sample was filtered to exclude those SNPs that were identified as homozygous in the paired-normal sample to generate somatic LOH event profile (**Supplementary Figure 2a**), and from it a mirrored BAF (mBAF, **Supplementary Figure 2b**) profile was calculated by the following rules[41]:

$$mBAF = BAF, if\ BAF \geq 0.5; 1 - BAF, if\ BAF < 0.5$$

The mBAF profile was then subjected to segmentation with CBS algorithm[40] (**Supplementary Figure 2c–d**). Next, the fraction (*f*) of cells that were harboring the loss of

heterozygosity (LOH) events for each LOH segment, with a segmental mBAF mean (u) and absolute ploidy (n), was calculated by the following formula (**Supplementary Figure 2e**). Next, cell fractions

$$f = \frac{2u - 1}{n(1 - u) + (2u - 1)}$$

were clustered to further reduce noise by assigning each cluster a centroid value, which was calculated as the weighted average of its member segments, the length of the segments being the weight. A subclone profile was then constructed (**Supplementary Figure 2f**) according to a specific biological model that the most prevalent mutations emerged earliest during the tumor evolutionary process, and less prevalent mutations occurred later in a linear heritage line. Thus if two LOH clusters, A and B, were identified at cell fraction 70% and 30% respectively, the model would conclude two tumorous subclones: a 40% ancestral subclone with only LOH events in cluster A and a 30% descendant subclone with both events in cluster A and B. The remaining 30% of total 100% fraction would then be deemed as normal tissue cells mixture in the tumor sample.

## Immunohistochemistry for *AKT1*

Four μm thick formalin-fixed, paraffin embedded sections were deparaffinized in xylene and rehydrated in graded ethanol. Antigen retrieval was performed by heating the sections for 20 minutes in microwave oven in 10mM sodium citrate buffer (pH 6.0). Endogenous peroxidase activity was blocked by treating the sections in 3% hydrogen peroxide for 5 minutes. Non-specific antigen binding was blocked by incubating the sections with 3% bovine serum albumin for 5 minutes. Sections were then incubated overnight with primary polyclonal goat anti-AKT1 (1:75 dilution; D-17, sc-7126; Santa Cruz Biotechnology, Inc.) Bound antibody was detected by incubating the sections with rabbit anti-goat HRP (1:200 dilution; Dako) for 1 hour and then developed by EnVision+ System-HRP (DAB) (Dako) according to standard protocol. Sections were subsequently counterstained with Mayer's haematoxylin. Photomicrographs were obtained using 100X and 400X magnifications for low and high power images, respectively.

Extended Data

a

b
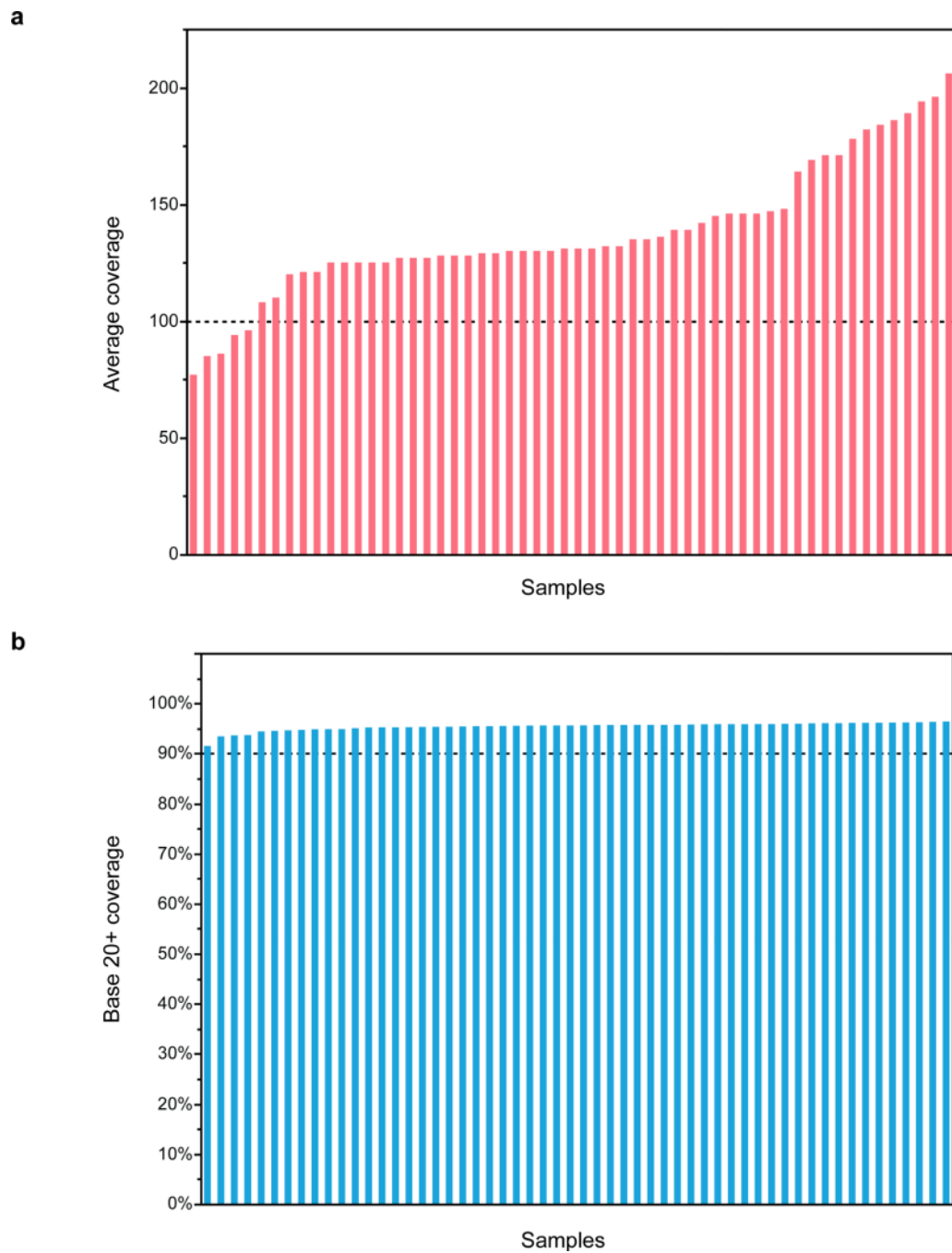
Extended Data Figure 1. The performance of whole-exome sequencing

a, The average read coverage across all samples in the discovery study. **b**, The base 20+ coverage. Base 20+ coverage, the percentage of targeted bases that were covered by at least 20 sequencing reads.
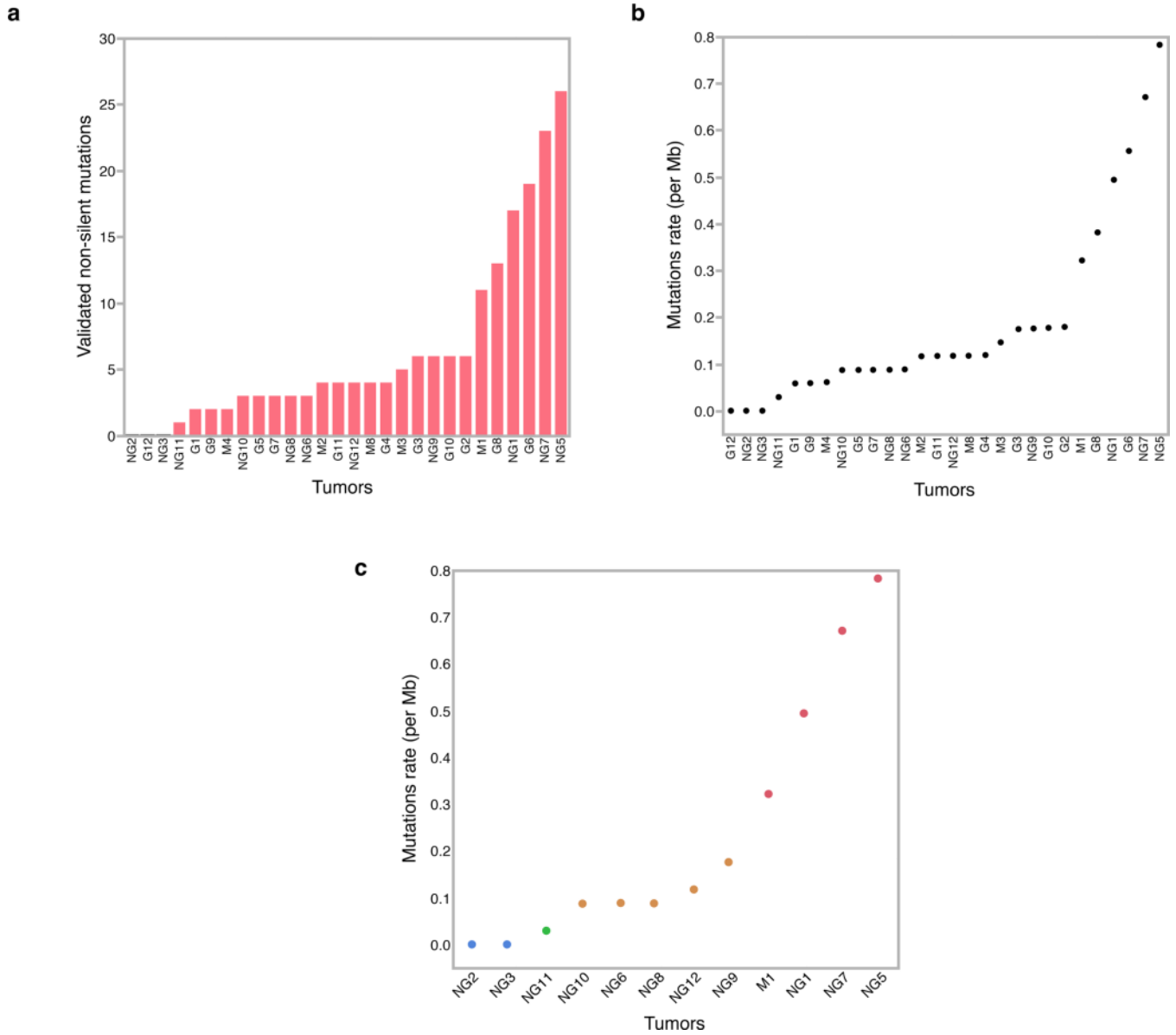
Extended Data Figure 2. The number and rate of validated somatic mutations
a, The number of validated somatic nonsynonymous mutations across all tumors in the discovery study. **b**, The rate of somatic non-silent mutations across all tumors in the discovery study. **c**, The mutation rate of NGGCTs. Only validated non-silent somatic mutations were counted. Blue filled circles, mature teratomas; Green filled circle, tumor defined as NGGCT but without detail information for subtype; Brown filled circles, immature teratomas; Red filled circles, yolk sac tumors. It is of note that M1 tumor is a mixture of germinoma and yolk sac tumor thus was included here. Mature teratomas, which are considered as histologically benign tumors, have the lowest mutation rate, followed by immature teratomas (0.1/Mb). Yolk sac tumors have the highest mutation rates (0.6/Mb).
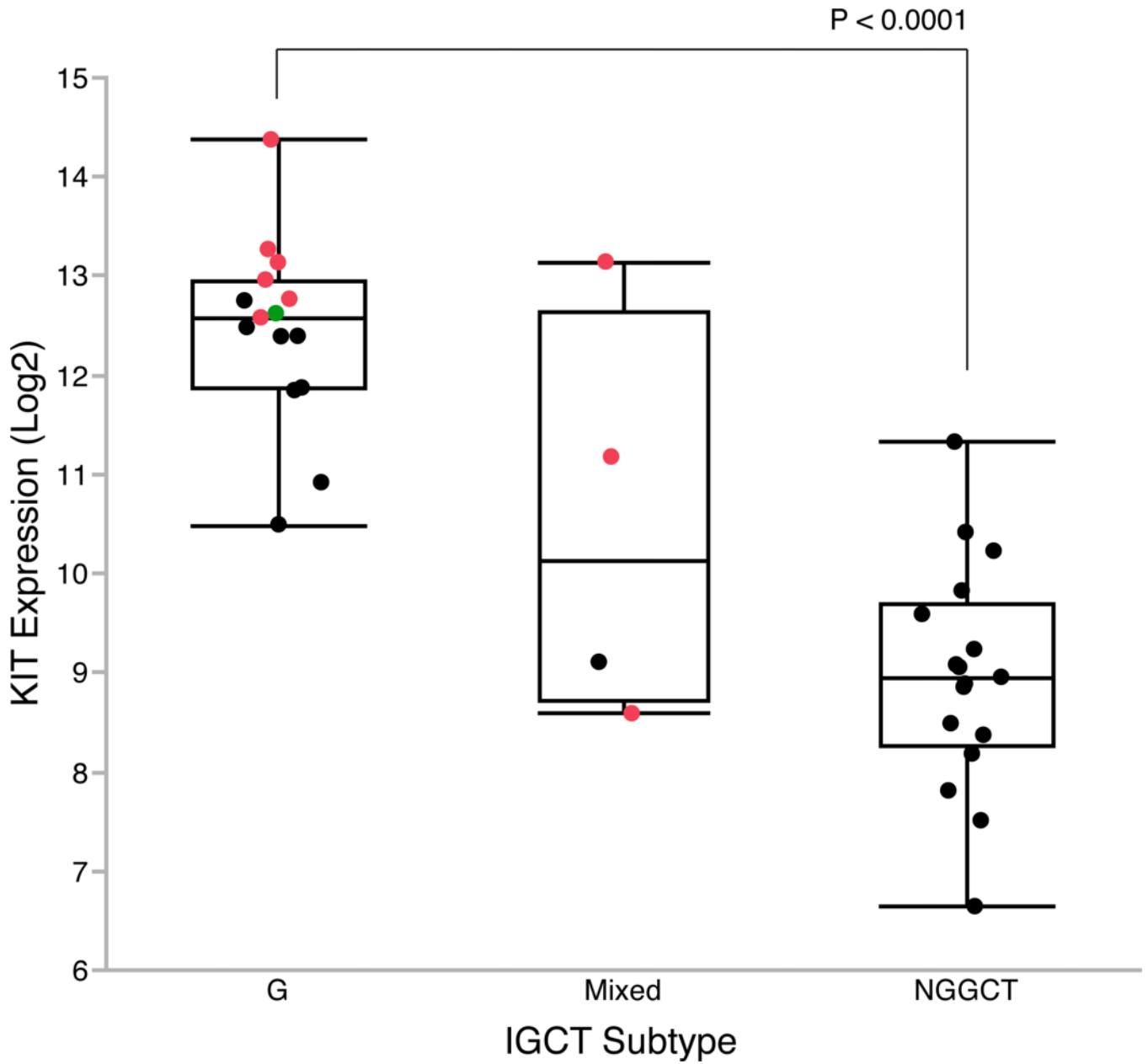
Extended Data Figure 3. The mRNA expression levels of *KIT*

The mRNA expression levels were determined by Affymetrix U133Plus2 human gene expression array for 37 out of the total 62 IGCT tumors with available RNA. Red dots, tumors with somatic mutation of *KIT*; green dot, tumor with somatic *CBL* mutation and was wildtype for *KIT*; G, pure germinomas; Mixed, mixed germ cell tumors with germinoma component; NGGCT, nongerminomatous germ cell tumors. P-value was calculated using one-way ANOVA analysis.
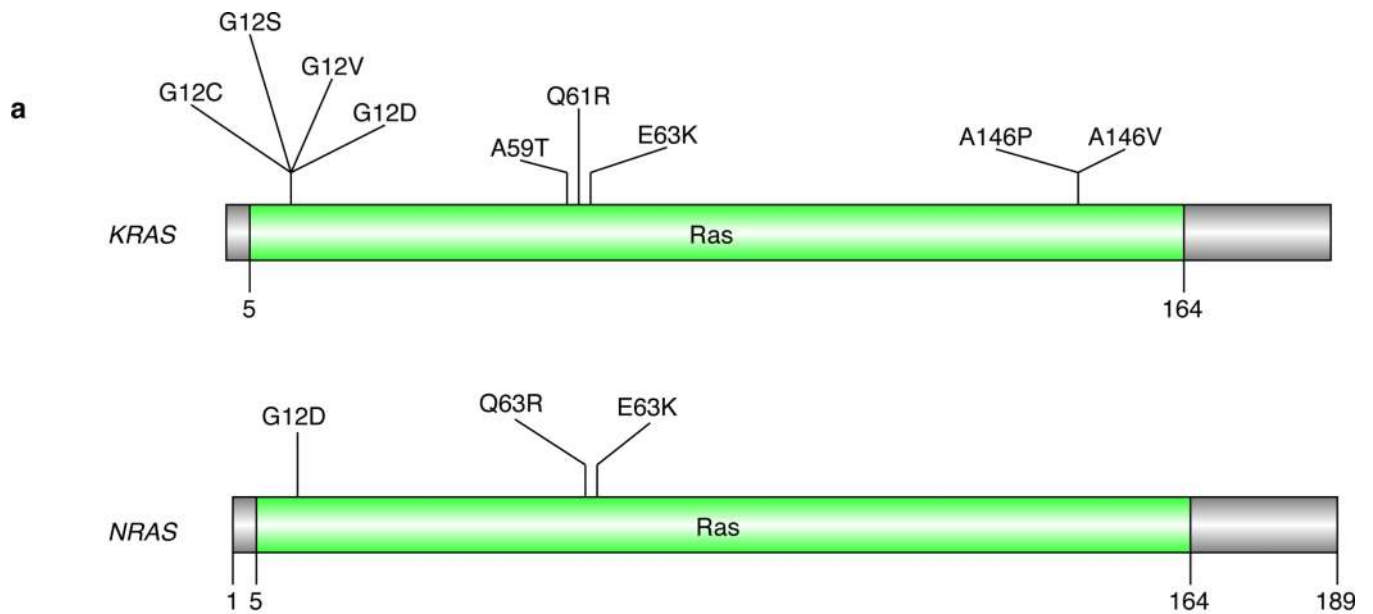
Extended Data Figure 4. The somatic mutations identified in *KRAS* and *NRAS*
a, The distribution of somatic mutations identified in *KRAS* and *NRAS*. The positions and amino acid changes were indicated for each mutation. **b**, *KRAS/NRAS* mutations are mutually exclusive with mutations in *KIT*. Two-by-Two table showing that mutations in *KIT* and *KRAS/NRAS* were mutually exclusive. Left-tailed Fisher's exact test was applied to calculate the p value. P=0.018.
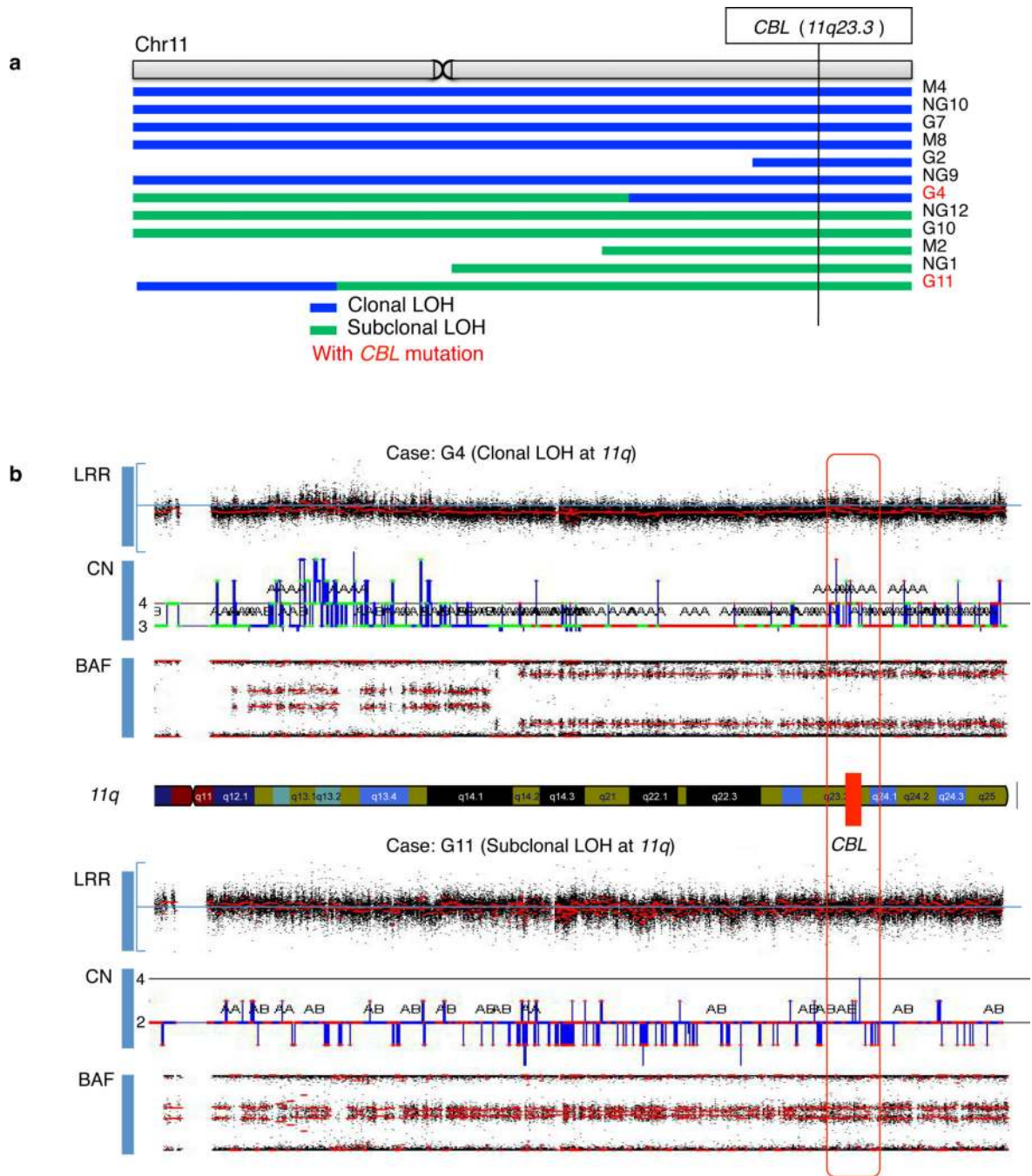
Extended Data Figure 5. The clonal and subclonal loss of heterozygosity events on chromosome 11q (11qLOH)
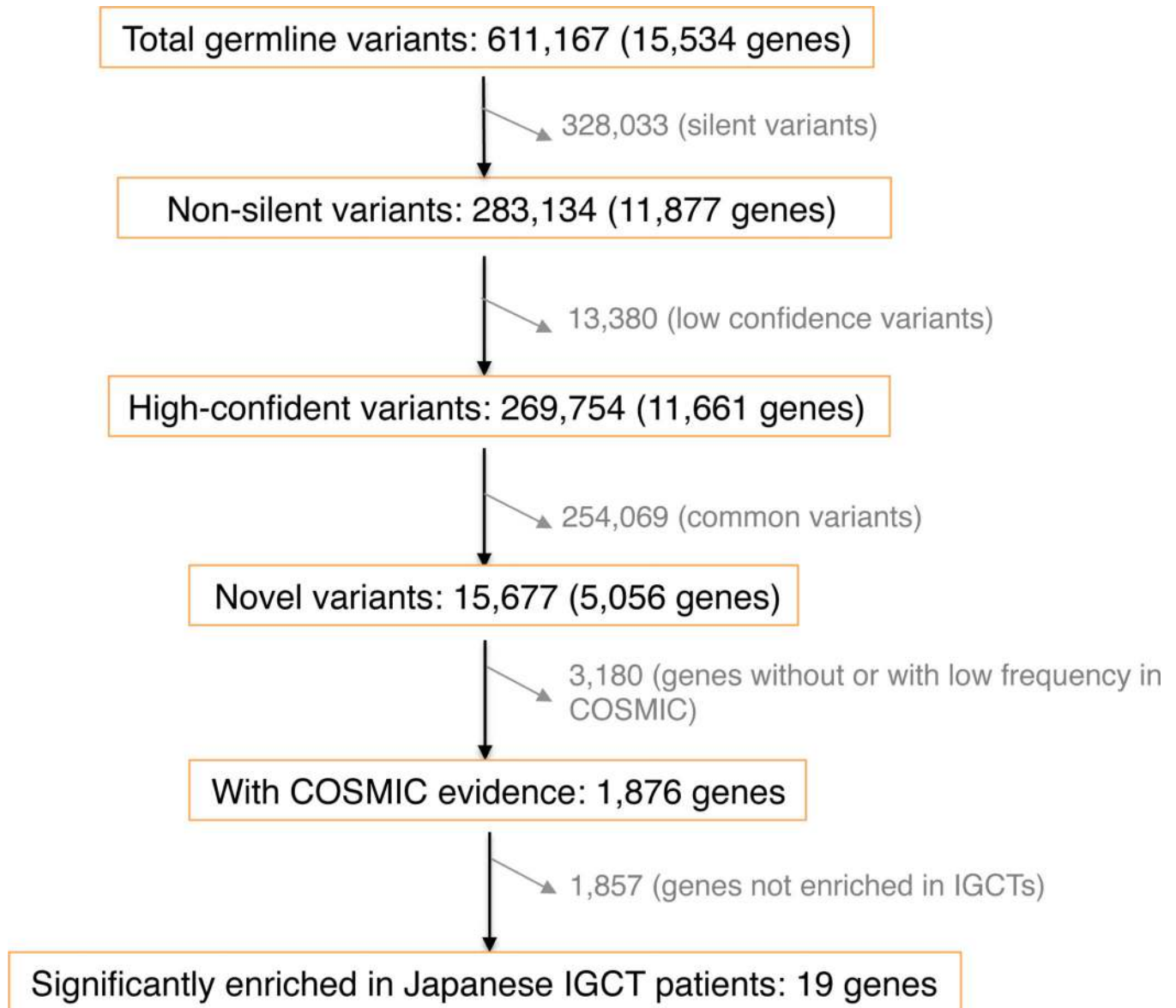
a, Topographic maps showing regions of clonal or subclonal 11qLOH spanning the *CBL* locus in individual patients. **b**, The clonal and subclonal 11qLOH in two representative cases G4 and G11. The red rectangle indicates the *CBL* locus. BAF, B allele frequency; LRR, Log2 R ratio; CN, copy number.
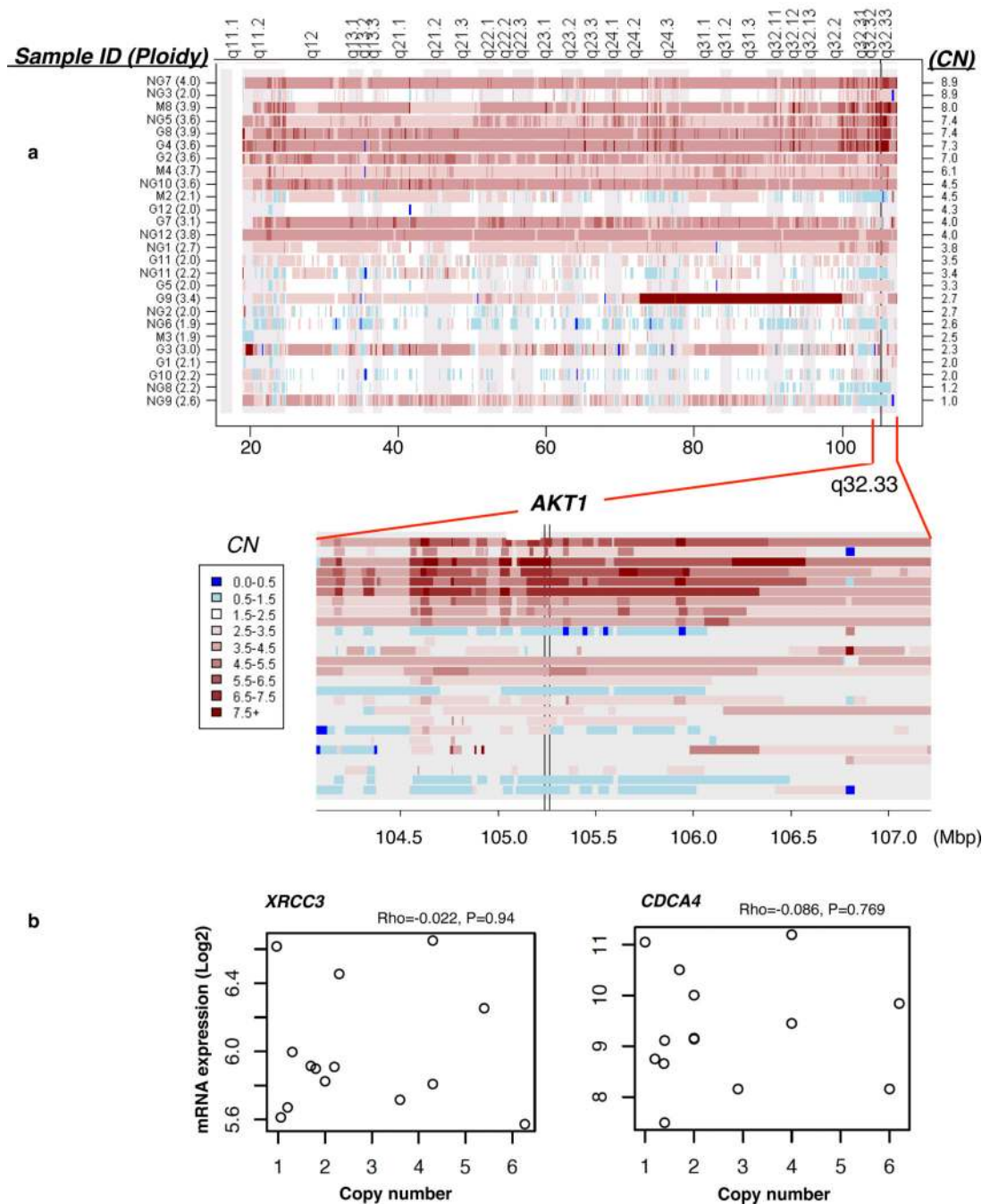
Extended Data Figure 6. Recurrent DNA copy number gains at 14q32 identified in the discovery study

a, Recurrent focal amplification of 14q32.33 spanning the *AKT1* locus. Regions of absolute DNA copy number are plotted for 14q (top panel) and 14q32.33 spanning the *AKT1* locus (bottom panel). Each row represents an individual tumor. Tumors were sorted descendingly by their absolute copy number within the boundaries of *AKT1* (indicated to the right). The mean copy number across chromosomes 1–22 (ploidy) is indicated alongside the tumor IDs on the left. The *x*-axis shows chromosome 14 genomic locations in megabase pairs (Mbp).

CN, copy number inferred by Omni2.5 SNP array. **b**, The correlation between the DNA copy number status and levels of mRNA expression of two representative genes, *XRCC3 and CDCA4*, presented in focal amplified region *14q32.33*. The P-value across all groups was calculated by Spearman's rank-order correlation analysis. Rho, the Spearman's correlation coefficient.

Total germline variants: 611,167 (15,534 genes)

328,033 (silent variants)

Non-silent variants: 283,134 (11,877 genes)

13,380 (low confidence variants)

High-confident variants: 269,754 (11,661 genes)

254,069 (common variants)

Novel variants: 15,677 (5,056 genes)

3,180 (genes without or with low frequency in COSMIC)

With COSMIC evidence: 1,876 genes

1,857 (genes not enriched in IGCTs)

Significantly enriched in Japanese IGCT patients: 19 genes

Extended Data Figure 7. The workflow for filtering of germline variants
Germline variants were filtered step by step to pick up the potentially interesting candidates. First, select the non-silent variants including missense, nonsense, frameshift, and splice-site variants. Second, select the high-confident variants that meet the following criteria: 1) variant allele fraction in both tumor and normal equal to or greater than 0.20; 2) variant calling were supported by at least 4 sequencing reads for both tumor and normal samples. Third, select novel variants that have not been reported in dbSNP database (dbSNP135).

Then, select genes with COSMIC evidence, i.e. genes for which mutations have been reported in COSMIC database in at least 100 times. After that, for all 1876 genes left in the above list, calculate the fold of enrichment of the germline variants in Japanese IGCT patients by comparing its frequency to that of Japanese patients in the control cohort and performed Fisher's exact test to calculate the p values. Then, select potentially interesting genes based on the IGCT frequency bias (>=4) and significant p-values (<0.05).
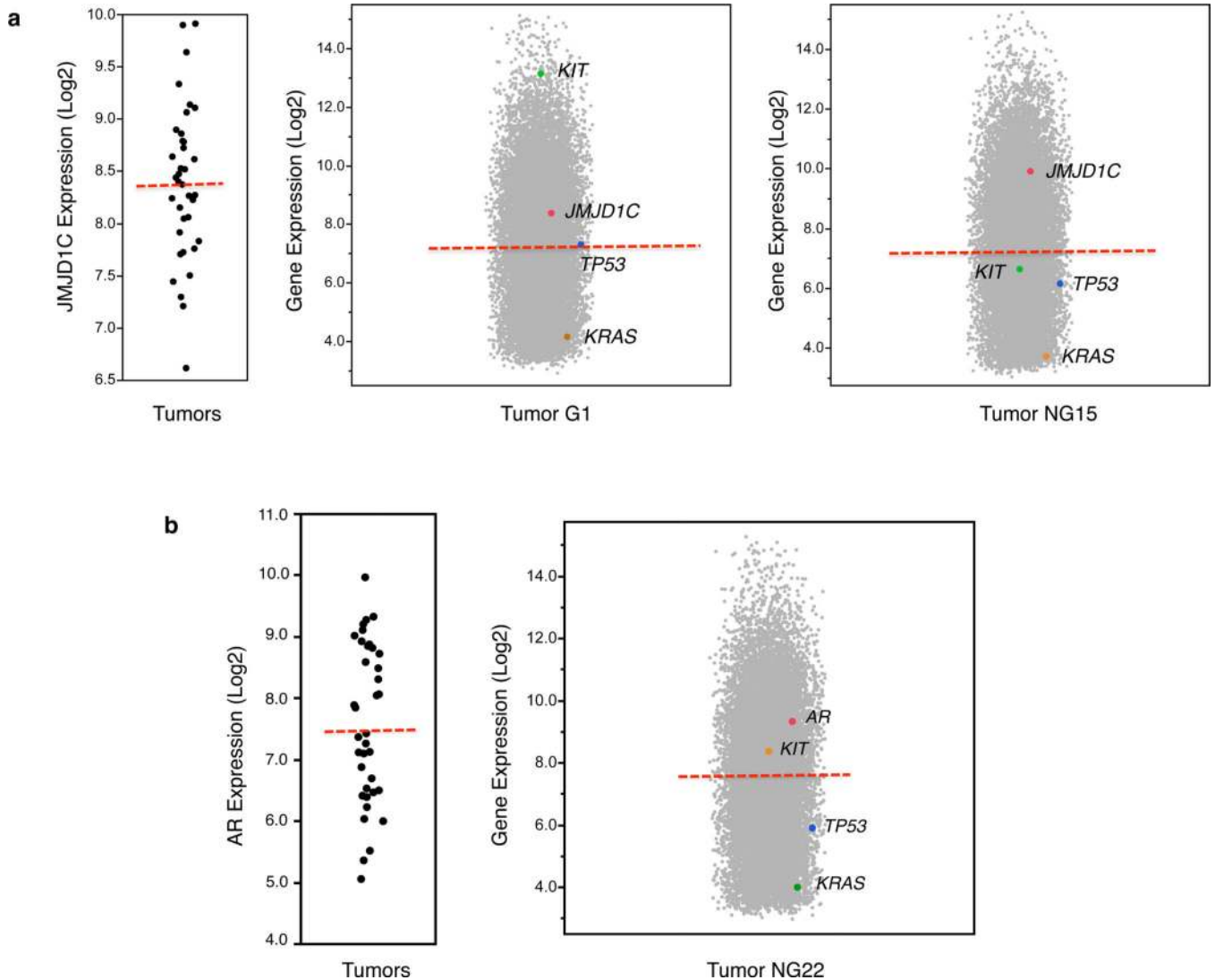

Extended Data Figure 8. The mRNA expression levels of *JMJD1C* and *AR* in IGCT

a, The mRNA expression levels of *JMJD1C*. **b**, The mRNA expression levels of *AR*. The mRNA expression levels were determined by Affymetrix U133Plus2 human gene expression array for 37 out of the total 62 tumors with available RNA. Left, the expression level of *JMJD1C* or *AR* in all IGCT tumors analyzed; middle and right, the expression level of *JMJD1C* or *AR* comparing to other known genes in representative tumors. Selected genes were highlighted in different colors and the remaining genes were colored in gray. The red dash lines indicate median values of expression.
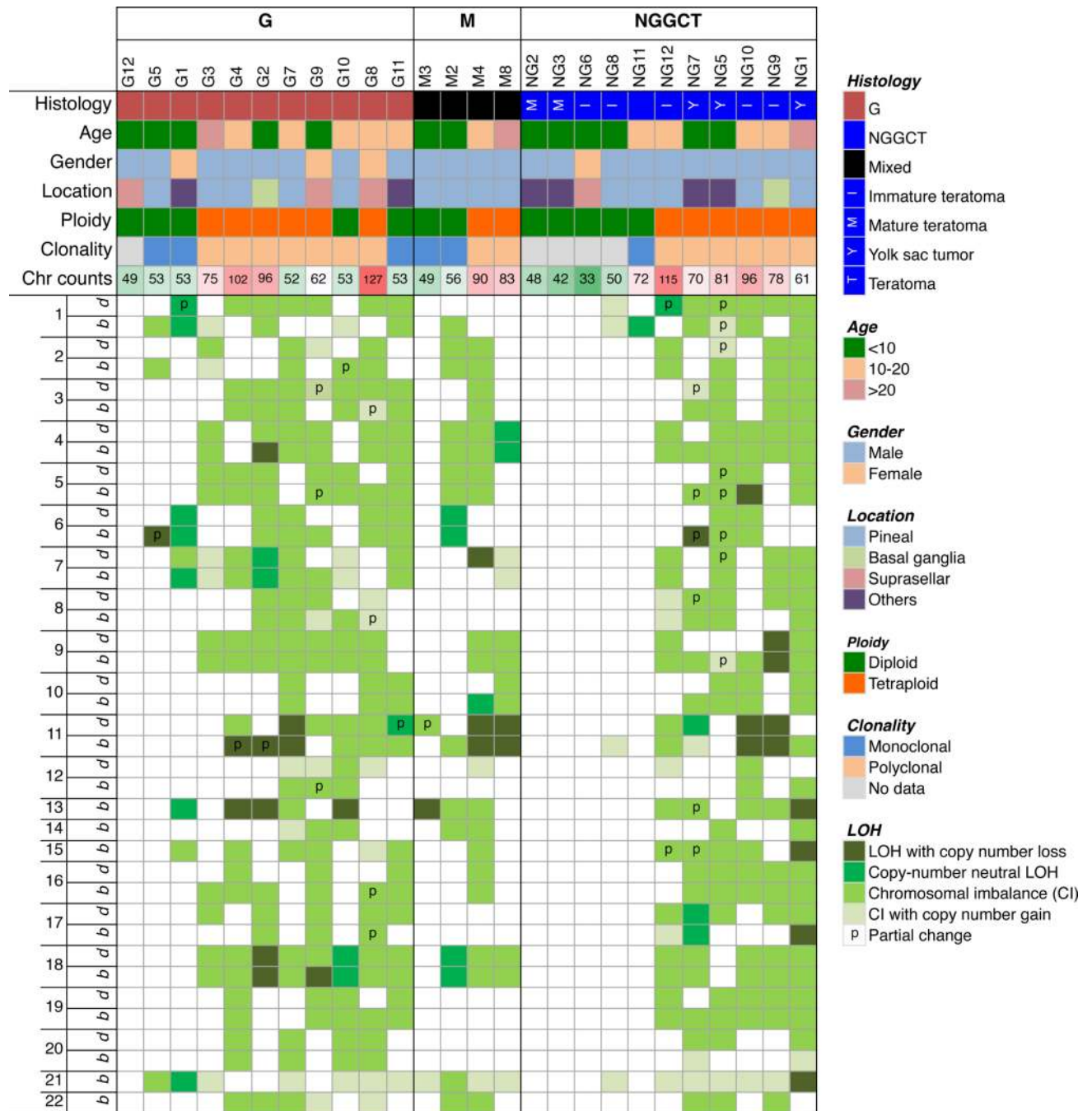
Extended Data Figure 9. Subgroup specificity of LOH and chromosomal imbalance in IGCT
Summary of the gross chromosomal alterations based on genome-wide Illumina Omni2.5
SNP array. Ploidy was predicted by Genome Alteration Print (GAP) algorithm[30].
Chromosomal imbalances are represented by the change of B-allele frequency (BAF) pattern
with or without the loss of heterozygosity (LOH). G, germinoma; NGGCT,
nongerminomatous germ cell tumor; M, Mixed GCTs with germinoma component.
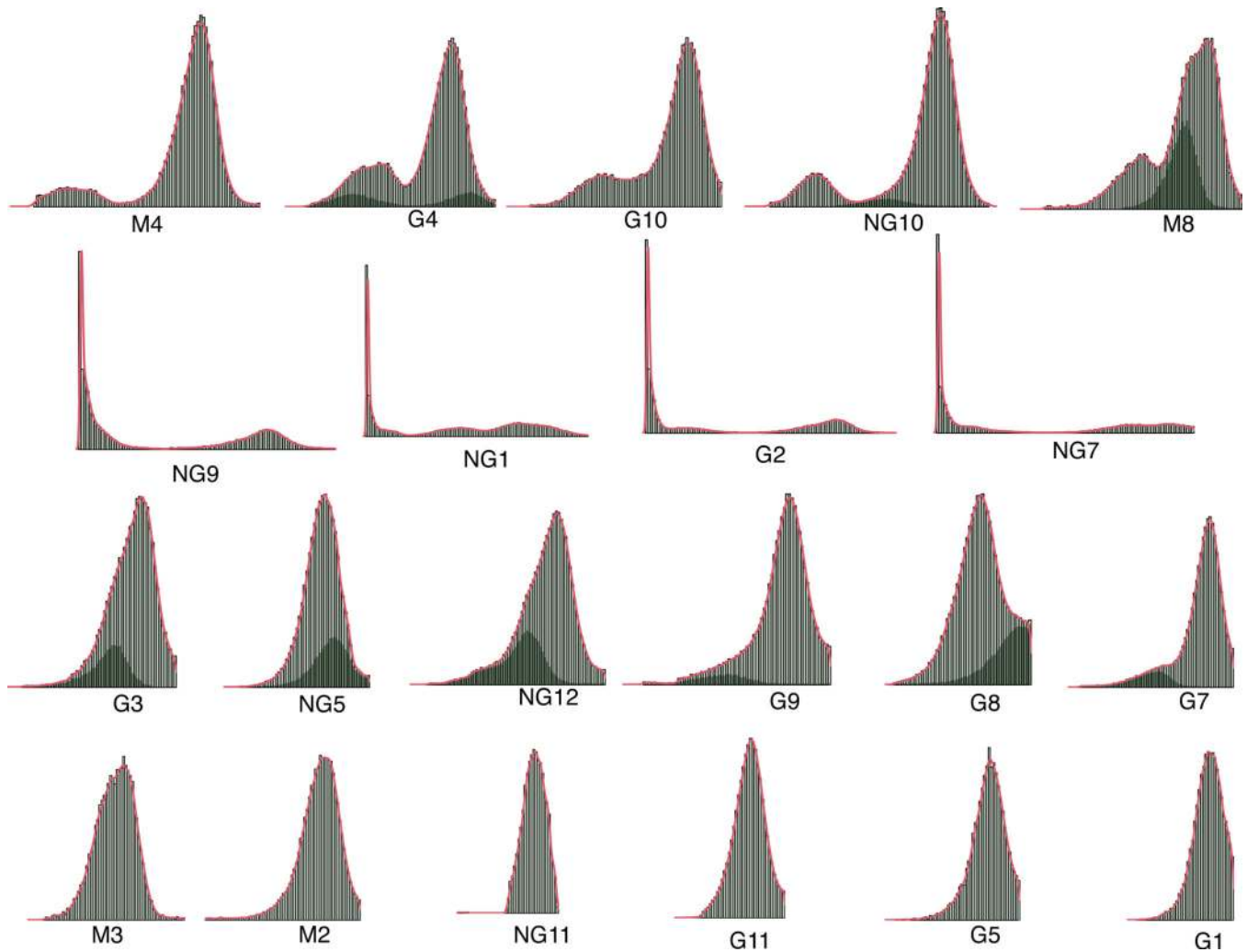
Extended Data Figure 10. An overview of the subclonal signatures across all tumors in the discovery study

Those cases without SNP array data or without detectable copy number changes were excluded for clonality analysis. Each peak in the plot indicates a subclone. The *x* axis indicates mBAF and the *y* axis indicate the number of heterozygous SNPs. Those cases with single peak are monoclonal and those with multiple peaks are polyclonal. Some subclones were highlighted, such as the subclone in G4, for a better visualization. The amplitude of the peaks in the plot has nothing to do with the fractions of cells that are harboring each event.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

REFERENCES

1. Packer RJ, Cohen BH, Cooney K. Intracranial germ cell tumors. The oncologist. 2000; 5:312–320. [PubMed: 10964999]

2. Jennings MT, Gelman R, Hochberg F. Intracranial germ-cell tumors: natural history and pathogenesis. Journal of neurosurgery. 1985; 63:155–167. [PubMed: 2991485]

3. McCarthy BJ, et al. Primary CNS germ cell tumors in Japan and the United States: an analysis of 4 tumor registries. Neuro-oncology. 2012; 14:1194–1200. [PubMed: 22869621]

4. Matsutani M, et al. Primary intracranial germ cell tumors: a clinical analysis of 153 histologically verified cases. Journal of neurosurgery. 1997; 86:446–455. [PubMed: 9046301]

5. Kamakura Y, Hasegawa M, Minamoto T, Yamashita J, Fujisawa H. C-kit gene mutation: common and widely distributed in intracranial germinomas. Journal of neurosurgery. 2006; 104:173–180. [PubMed: 16572634]

6. Sakuma Y, et al. c-kit gene mutations in intracranial germinomas. Cancer science. 2004; 95:716–720. [PubMed: 15471556]

7. Rickert CH, Simon R, Bergmann M, Dockhorn-Dworniczak B, Paulus W. Comparative genomic hybridization in pineal germ cell tumors. Journal of neuropathology and experimental neurology. 2000; 59:815–821. [PubMed: 11005262]

8. Schneider DT, et al. Molecular genetic analysis of central nervous system germ cell tumors with comparative genomic hybridization. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc. 2006; 19:864–873.

9. Terashima K, et al. Genome-wide analysis of DNA copy number alterations and loss of heterozygosity in intracranial germ cell tumors. Pediatric blood & cancer. 2014; 61:593–600. [PubMed: 24249158]

10. Kemmer K, et al. KIT mutations are common in testicular seminomas. The American journal of pathology. 2004; 164:305–313. [PubMed: 14695343]

11. Corless CL, Barnett CM, Heinrich MC. Gastrointestinal stromal tumours: origin and molecular oncology. Nature reviews. Cancer. 2011; 11:865–878. [PubMed: 22089421]

12. Gajiwala KS, et al. KIT kinase mutants show unique mechanisms of drug resistance to imatinib and sunitinib in gastrointestinal stromal tumor patients. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106:1542–1547. [PubMed: 19164557]

13. Janne PA, et al. Selumetinib plus docetaxel for KRAS-mutant advanced non-small-cell lung cancer: a randomised, multicentre, placebo-controlled, phase 2 study. The lancet oncology. 2013; 14:38–47. [PubMed: 23200175]

14. Hatzivassiliou G, et al. ERK inhibition overcomes acquired resistance to MEK inhibitors. Molecular cancer therapeutics. 2012; 11:1143–1154. [PubMed: 22402123]

15. Thien CB, Walker F, Langdon WY. RING finger mutations that abolish c-Cbl-directed polyubiquitination and downregulation of the EGF receptor are insufficient for cell transformation. Molecular cell. 2001; 7:355–365. [PubMed: 11239464]

16. Makishima H, et al. Mutations of e3 ubiquitin ligase cbl family members constitute a novel common pathogenic lesion in myeloid malignancies. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2009; 27:6109–6116. [PubMed: 19901108]

17. Makishima H, et al. CBL mutation-related patterns of phosphorylation and sensitivity to tyrosine kinase inhibitors. Leukemia. 2012; 26:1547–1554. [PubMed: 22246246]

18. Liu LZ, et al. AKT1 amplification regulates cisplatin resistance in human lung cancer cells through the mammalian target of rapamycin/p70S6K1 pathway. Cancer research. 2007; 67:6325–6332. [PubMed: 17616691]

19. Li M, et al. Somatic mutations in the transcriptional corepressor gene BCORL1 in adult acute myelogenous leukemia. Blood. 2011; 118:5914–5917. [PubMed: 21989985]

20. Pagan JK, et al. A novel corepressor, BCoR-L1, represses transcription through an interaction with CtBP. The Journal of biological chemistry. 2007; 282:15248–15257. [PubMed: 17379597]

21. Heery DM, Kalkhoven E, Hoare S, Parker MG. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. Nature. 1997; 387:733–736. [PubMed: 9192902]

22. Jin G, et al. Genome-wide association study identifies a new locus JMJD1C at 10q21 that may influence serum androgen levels in men. Human molecular genetics. 2012; 21:5222–5228. [PubMed: 22936694]

23. Kim SM, et al. Regulation of mouse steroidogenesis by WHISTLE and JMJD1C through histone methylation balance. Nucleic acids research. 2010; 38:6389–6403. [PubMed: 20530532]

24. Kuroki S, et al. JMJD1C, a JmjC domain-containing protein, is required for long-term maintenance of male germ cells in mice. Biology of reproduction. 2013; 89:93. [PubMed: 24006281]

25. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. American journal of human genetics. 2008; 83:311–321. [PubMed: 18691683]

26. Madsen BE, Browning SR. A groupwise association test for rare mutations using a weighted sum statistic. PLoS genetics. 2009; 5:e1000384. [PubMed: 19214210]

27. Katoh M, Katoh M. Identification and characterization of TRIP8 gene in silico. International journal of molecular medicine. 2003; 12:817–821. [PubMed: 14533015]

28. Wolf SS, Patchev VK, Obendorf M. A novel variant of the putative demethylase gene, s-JMJD1C, is a coactivator of the AR. Archives of biochemistry and biophysics. 2007; 460:56–66. [PubMed: 17353003]

29. Okada Y, Nishikawa R, Matsutani M, Louis DN. Hypomethylated X chromosome gain and rare isochromosome 12p in diverse intracranial germ cell tumors. Journal of neuropathology and experimental neurology. 2002; 61:531–538. [PubMed: 12071636]

30. Popova T, et al. Genome Alteration Print (GAP): a tool to visualize and mine complex cancer genomic profiles obtained by SNP arrays. Genome biology. 2009; 10:R128. [PubMed: 19903341]

SUPPLEMENTARY REFERENCES

31. TCGA. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012; 487:330–337. [PubMed: 22810696]

32. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25:1754–1760. [PubMed: 19451168]

33. DePristo MA, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature genetics. 2011; 43:491–498. [PubMed: 21478889]

34. Shen Y, et al. A SNP discovery method to assess variant allele probability from next-generation resequencing data. Genome Res. 2010; 20:273–280. [PubMed: 20019143]

35. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics. 2009; 25:2865–2871. [PubMed: 19561018]

36. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic acids research. 2010; 38:e164. [PubMed: 20601685]

37. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature protocols. 2009; 4:1073–1081. [PubMed: 19561590]

38. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. PLoS one. 2012; 7:e46688. [PubMed: 23056405]

39. Adzhubei IA, et al. A method and server for predicting damaging missense mutations. Nature methods. 2010; 7:248–249. [PubMed: 20354512]

40. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinformatics. 2007; 23:657–663. [PubMed: 17234643]

41. Staaf J, et al. Segmentation-based detection of allelic imbalance and loss-of-heterozygosity in cancer cells using whole genome SNP arrays. Genome biology. 2008; 9:R136. [PubMed: 18796136]
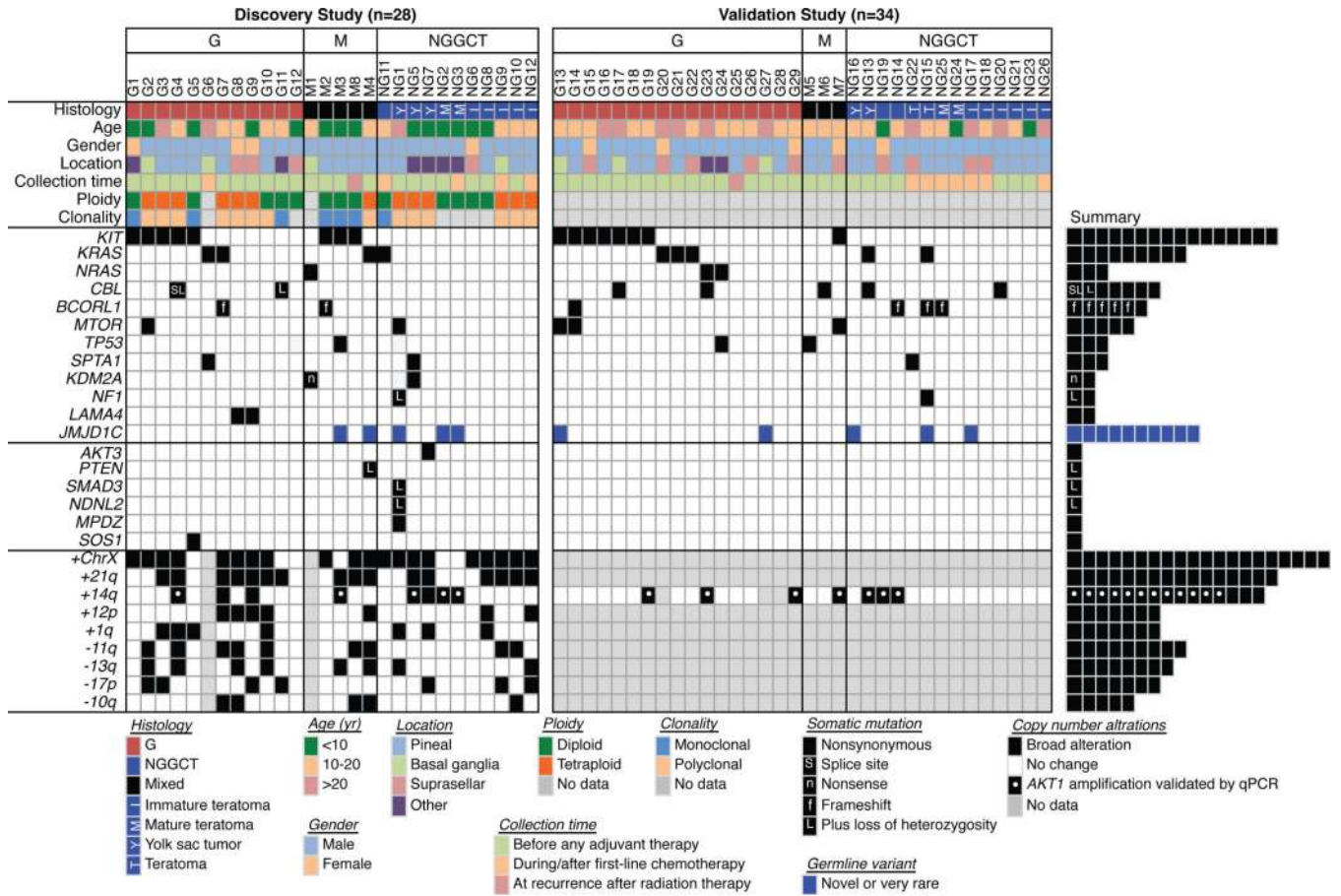
Figure 1. Subgroup specificity of the recurrent genetic alterations identified in 62 IGCT patients
G, germinoma; NGGCT, nongerminomatous germ cell tumor; M, mixed germ cell tumor with germinoma component. Several genes were included that are mutated once but considered biologically important by one of the following criteria: involvement in KIT/RAS or AKT/mTOR pathways, known interaction with *KIT*, tumor suppressor genes with two hits. Germline variants in *JMJD1C* were either novel or rare polymorphisms with minor allele frequency less than 0.005. qPCR of *AKT1* was validated if the gene count adjusted by ploidy was great than 3.
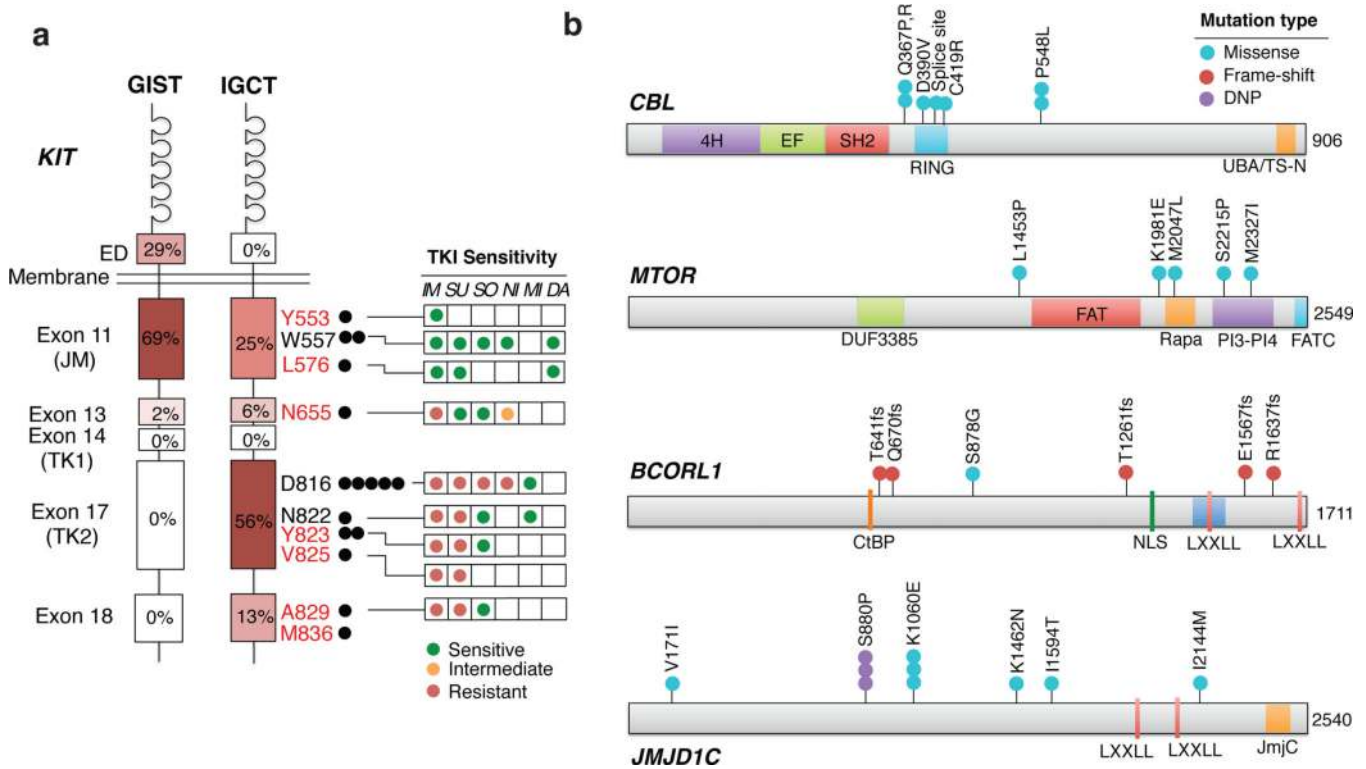
Figure 2. Novel recurrent somatic and germline mutations in IGCT

a, Somatic *KIT* mutations. Red lettering, the novel *KIT* mutations identified in IGCT; black lettering, reported *KIT* mutations; Black filled circles, the number of mutations identified at each mutation site. The primary *KIT* mutations reported in gastrointestinal stromal tumors are shown for comparison. Functional domains: ED, extracellular domain; JM, juxtamembrane domain; TK1, tyrosine kinase I, the ATP binding domain; TK2, tyrosine kinase II, the kinase activation loop (A-loop). The sensitivity of known tyrosine kinase inhibitors (TKIs), corresponding to each mutation site from previous studies performed in other tumor types are shown on the right. IM, Imatinib; SU, Sunitinib, SO, sorafenib; NI, nilotinib; MI, Midostaurin; DA, Dasatanib. *KIT* exon-11 mutations are generally sensitive to Imatinib. Certain Imatinib-resistant *KIT* mutations respond to Sunitinib and Sorafenib. More than half of *KIT* mutations in IGCTs reside in the A-loop (**Supplementary Figure 3**). The D816 mutation causes *KIT* to be constitutively activated by altering the structure of the JM domain and destabilizing the A-loop inactive conformation[12]. Tumors with D816-mutated *KIT* respond well to Midostaurin. **b**, Schematic representation of somatic mutations in *CBL, MTOR, BCORL1*, and germline variants identified in *JMJD1C*. DNP, dinucleotide polymorphism; Rapa, rapamycin binding site; PI3-PI4, PI3-PI4 kinase; CtBP, CtBP binding site; NLS, nuclear localization signal; LXXLL motif, L is Leucine and X is any amino acid. Only novel or rare *JMJD1C* germline variants (MAF<0.005) are shown.
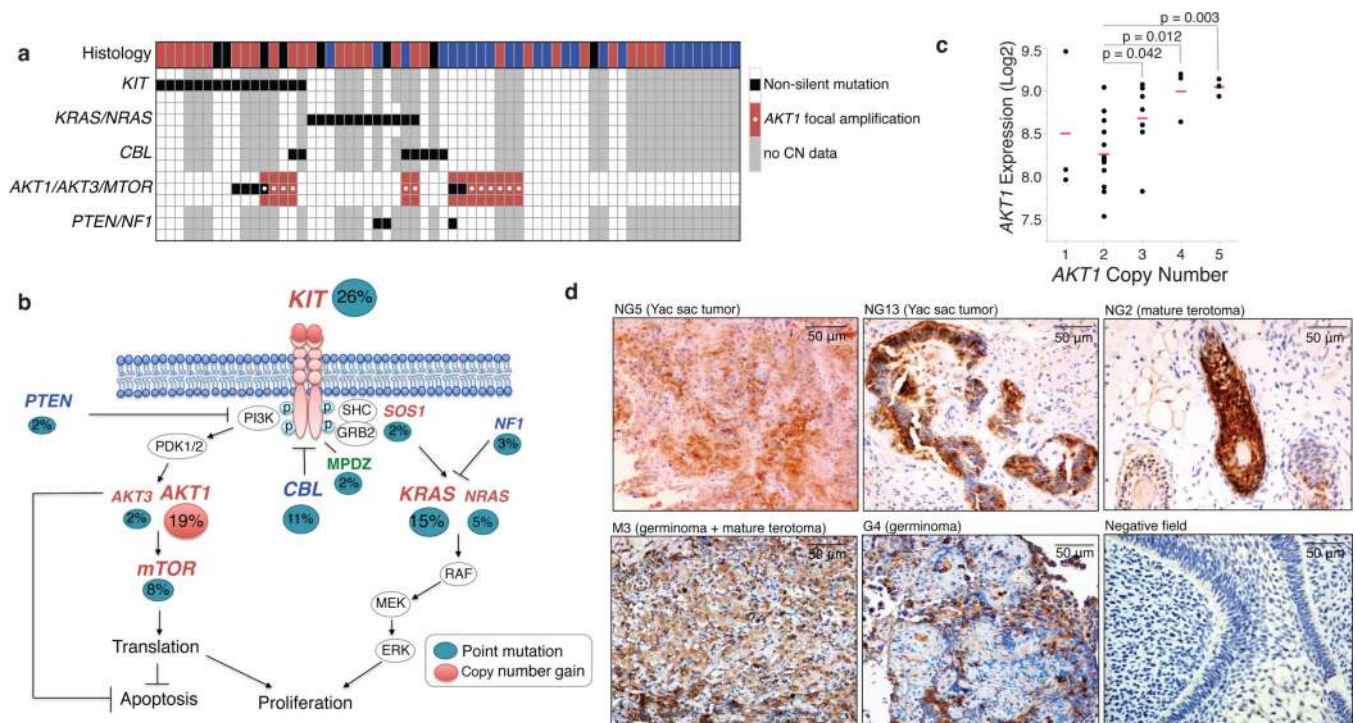
Figure 3. Frequent genetic alteration of KIT/RAS and AKT/mTOR signaling pathways

a, Summary of the somatic events. **b**, KIT/RAS and AKT/mTOR pathway interactions showing frequencies of somatic alterations in key genes. Alteration frequencies are expressed as a percentage of all IGCT patients. Red lettering, protein positively regulates signaling; blue lettering, protein negatively regulates signaling, and green lettering, physically interacting protein. **c**, the correlation of *AKT1* copy number status and levels of *AKT1* mRNA expression. *AKT1* copy number status was assayed by SNP array and validated by qPCR. The mRNA expression levels were determined by Affymetrix U133Plus2 human gene expression array. The red lines indicate mean values of expression. The P-value across all groups calculated by Spearman's rank-order correlation analysis is 0.001 and the correlation coefficient is 0.5614. The P values between two different groups calculated by one-way ANOVA analysis are shown. **d**, Immunohistochemical staining of *AKT1* in *AKT1* amplified tumors. Immunostaining was carried out with AKT1-specifc goat antibody D-17 (sc7126, Santa-Cruz Biotechnology) at 1:75 dilution. Magnification: 400X. Scale bar: 50 μm. Cases M3 and NG5 showed strong and diffuse cytoplasmic and nuclear staining while cases G4, NG2 and NG13 showed strong but focal cytoplasmic and nuclear staining.
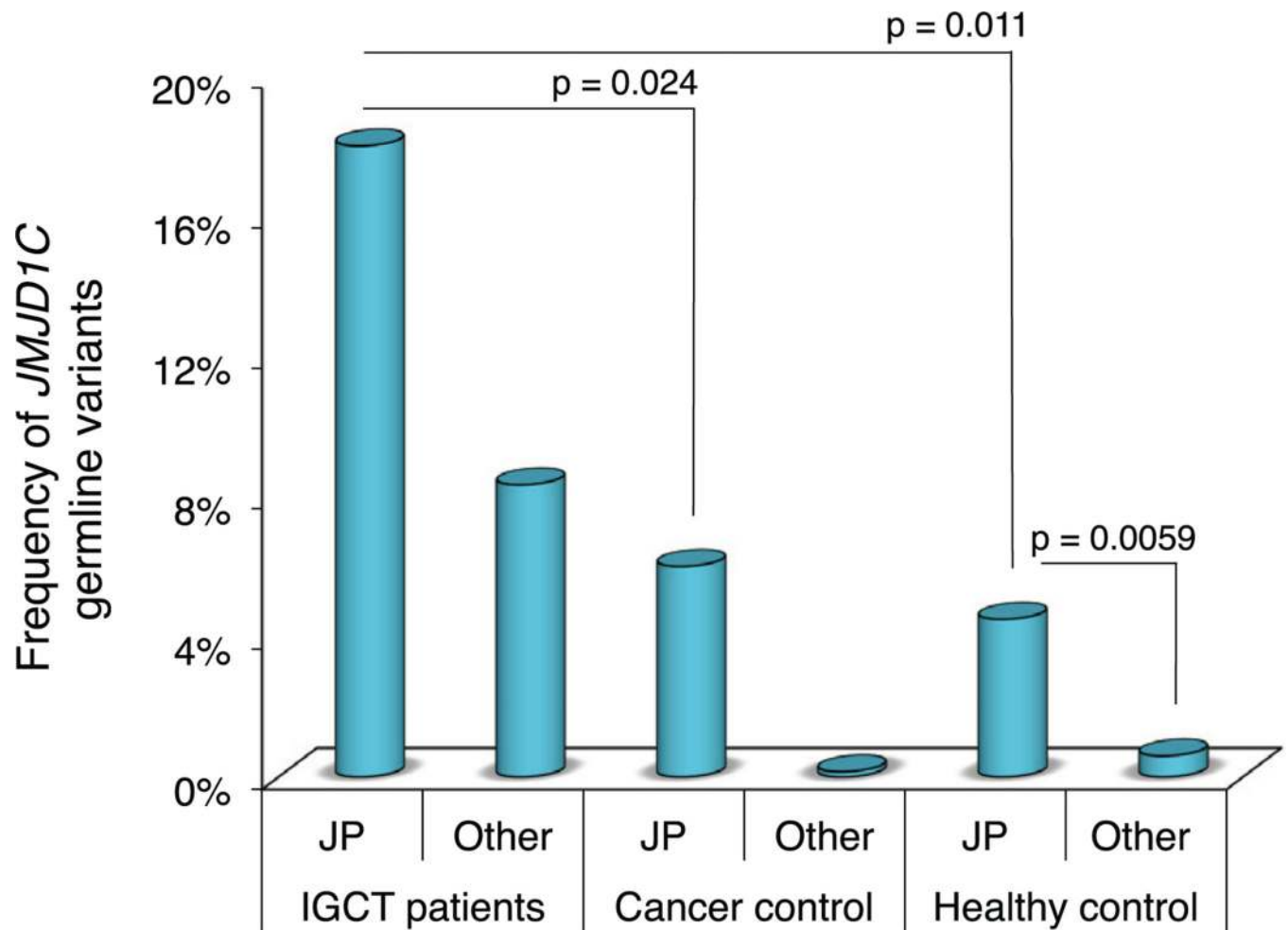
Figure 4. Enrichment of germline *JMJD1C* variants in IGCT
Only validated novel or rare (MAF< 0.005) non-silent germline variants were counted. JP, Japanese. Healthy controls were from the 1000 Genomes Project (n=1092, http://www. 1000genomes.org/). A panel of 778 unrelated cancers sequenced at HGSC (cancer types =10, from both Asia and United States) were used as the cancer control. The one-sided Fisher's exact test was performed to test the enrichment of the minor alleles in IGCT cases.