**Centrosome linker-induced tetraploid segregation errors links rhabdoid phenotype and lethal colorectal cancers**

Andrea Remo1†, Erminia Manfrin2†, Pietro Parcesepe2†, AlbertoFerrarini3, Hye Seung Han4, Ugnius Mickys5, Carmelo Laudanna 6, Michele Simbolo2, Donatella Malanga6, Duarte Mendez Oliveira6, Elisabetta Baritono1, Tommaso Colangelo7, Lina Sabatino8, Jacopo Giuliani1, Enrico Molinari1, Marianna Garonzi9, Luciano Xumerle9, Massimo Delledonne9,10, Guido Giordano11,12, Claudio Ghimenton2 Fortunato Lonardo13, Fulvio D’angelo14, Federica Grillo15, Luca Mastracci15, Giuseppe Viglietto6, Michele Ceccarelli8,14, Vittorio Colantuoni8, Aldo Scarpa2,16\* and Massimo Pancione8,17\*.

1Pathology Unit, “Mater Salutis” Hospital AULSS9, Legnago (Verona), Italy; 2 Department of Diagnostics and Public Health – Section of Pathology, University and Hospital Trust of Verona, Verona, Italy; 3 Menarini Silicon Biosystems S.p.A, Bologna, Italy, 4Department of Pathology, Konkuk University School of Medicine, Seoul,Korea; 5National Center of Pathology, Affiliate of Vilnius University Hospital Santariskiu Clinics, Vilnius, Lithuania; 6Department of Experimental and Clinical Medicine "Gaetano Salvatore", University “Magna Grecia”,Catanzaro, Italy; 7Institute for Stem-cell Biology, Regenerative Medicine and Innovative Therapies (ISBReMIT), Casa Sollievo della Sofferenza-IRCCS, San Giovanni Rotondo, Italy; 8Department of Sciences and Technologies, University of Sannio, Benevento, Italy; 9Functional Genomics Center, Department of Biotechnology, University of Verona, Verona, Italy; 10Personal Genomics S.r.l., Verona, Italy; 11CRO Aviano National Cancer Center, Aviano, Italy; 12Medical Oncology Unit, San Filippo Neri Hospital, Rome, Italy; 13Medical Cytogenetics and Molecular Genetics Unit, AORN "Gaetano Rummo," Benevento; 14Bioinformatics Laboratory, BIOGEM scrl, Ariano Irpino, Avellino, Italy; 15 Department of Surgical and Diagnostic Sciences (DISC), University of Genova and S. Martino Polyclinic Hospital, Genova, Italy; 16ARC-Net Centre for applied research on cancer, University and Hospital trust of Verona, Verona, Italy; 17Department of Biochemistry and Molecular Biology II, Faculty of Pharmacy, Complutense University,Madrid, Spain.

† These authors contributed equally to this work.

\*Correspondence should be addressed to: Massimo Pancione, ([massimo.pancione@unisannio.it](mailto:massimo.pancione@unisannio.it)) Department of Sciences and Technologies, University of Sannio, Via Port’Arsa, 1182100 Benevento, Italy; Department of Biochemistry and Molecular Biology II, Faculty of Pharmacy, Complutense University, Madrid, Spain. Phone: +39 0824 305157; +34913941785 Fax +39 0824 305147 or Aldo Scarpa, ([aldo.scarpa@univr.it](mailto:aldo.scarpa@univr.it)), ARC-NET Research Centre, Policlinico GB Rossi, Piazzale L.A. Scuro, 10, Phone.: +39 045 8124043, Fax:+39 045 8127432,

**MATERIALS AND METHODS**

**Human research ethical approval.** Approval n. 997CESC from the Ethics Comittee (Comitato Etico di Verona e Rovigo dell’Azienda Ospedaliera Universitaria Integrata) approved on 7 September 2016, documented by the CESC prot. 42160 on 9 September 2016 and formalized by the General Manager with deliberation n. 458 of 16 September 2016, communicated with protocol 51319 on 23 September 2016.

**Patient and tissue cohort.** Rhabdoid colorectal cancers:formalin-fixed paraffin-embedded (FFPE) samples from 7 cases of primary rhabdoid colorectal cancers (RC) and matched normal colonic mucosa were studied (cases RC1 to RC7 **Supplemental Table 1).** Five of these cases had been previously reported [9,10,42,43].Five additional cases were used as an independent validation series to screen the mutational status of newly identified genes (cases RC8 to RC12 **Supplemental Table 1).** Rhabdoid tumours of infants/young adults: FFPE samples from 7 rhabdoid tumours arised in central nervous system of patients between 2 months and 19 years of age were collected from the files of the Azienda Ospedaliera Universitaria Integrata, Verona, Italy **(Supplemental Table 4).** These pediatric/young adults rhabdoid tumours are indicated as Rhabdoid of infancy (RI) throughout the article. Classic type colorectal cancer cohorts: two independent datasets of patients with classic type sporadic colorectal cancer were analysed: Dataset A (26,27) including 141 primary cancers and Dataset B (28,29) including 102 primary cancers. FFPE blocks were available for both cohorts. Matched normal mucosa was available for 61 cases of cohort A and 22 cases of cohort B. For 82 patients from the Cohort A, snap-frozen paired tumour-normal tissues stored at −80°C were also available. No patient received chemotherapy or radiation therapy prior to surgery. Patients who had a family history of intestinal dysfunction or CRC or had taken non-steroidal anti-inflammatory drugs on a regular basis were not included. Clinical pathological data of all patients are summarized in **(Supplemental Table 5).** Overall Survival and Progression Free Survival were recorded. The latest update was on May 1st, 2015. All cases were assessed for microsatellite instability, *KRAS* and *BRAFV600E* mutations (27,44). Tissue microarrays were assembled and screened for the immunohistochemical expression of mismatch repair proteins Mlh1, Msh2, Msh6 and Pms2 (27,44).

**Cell lines.** Human colon cancer cell lines HCT116, HT29, CaCo-2, LoVo, RKO, T84, DLD1, SW480 and SW620 were purchased from American Type Culture Collection (ATCC). BJ human fibroblasts derived from normal foreskin used as non-neoplastic control andG401 cells derived from a pediatric rhabdoid tumour were a gift by Dr. Roberta Maestro (CRO, Aviano, Italy). All cell lines were mycoplasma free as tested with the mycoAlert Mycoplasma Detection Kit (Lonza). The genotype of cell lines assessed by short tandem repeat (STR) genotyping was consistent with published genotype for each cell lines.

**DNA extraction from formalin-fixed paraffin-embedded samples.** The tumoural areas with rhabdoid morphology were enriched by manual microdissection from H&E stained sections (45). DNA was purified using QIAamp DNA FFPE Tissue Kit (Qiagen), quantified and qualified with both spectrophotometric Nanodrop 1000 (Thermofisher) and fluorometric Picogreen dsDNA kit (Invitrogen) (46). To further verify DNA integrity a multiplex PCR test was performed (47). Tumour cellularity was determined using SNP arrays (Illumina) and the allele-specific copy number analysis of tumours (ASCAT) algorithm (48).

**Whole-Exome Sequencing.** Whole-exome sequencing with 100-bp paired reads was performed with a HiSEQ1000 (Illumina), using 1.3 µg genomic DNA (based on fluorometric Picogreen dsDNA quantification) and enrichment for whole exome according to TruSeq Exome Enrichment Guide  (Illumina). Sequences were pre-processed by removing reads with a percentage of undetermined bases (N) higher than 10% of the read length and more than 50 bases with a quality ≤7. Adapters were clipped using Scythe v0.980 (https://github.com/vsbuffalo/scythe), and 3′ ends with a quality score <20 over a window of 10 bases were trimmed using Sickle v0.940 (https://github.com/vsbuffalo/sickle), entirely removing the fragment if the final length of one of the reads was lower than 50 bp. Filtered reads were mapped to reference hg19 genome using BWA 0.6.2 (49,50) and BAM files are available in the EGA (accession number: EGAS00001002480). Duplicates were marked with Picard tools and reads were realigned in correspondence of indels and recalibrated using GATK 2.6-5 (51,52).Variants were called with Mutect v1.1.4 and GATK 2.6-5 UnifiedGenotyper module with “-glm BOTH” parameter (53). Functional annotation of mutations was performed by integrating 5 prediction tools reported in **(Supplemental Table 2)** (54). Somatic single nucleotide variants (SNVs), indels, and copy number changes were visualized using circus (55). Structural variants not classifiable as deletions and duplications were omitted for clarity.

**Splice-site prediction tool and driver genes analysis.** A splice-site software SpliceFinder (20-22) based on two online publicly available tools (<http://www.fruitfly.org/seq_tools/splice.html>) and (<http://www.cbs>.dtu.dk/services/NetGene2/) was used for the prediction of splicing variants derived from whole-exome sequencing data. The SNP variant was accepted as deleterious splicing mutation when the prediction was confirmed by both programs. The computational tool, driver genes and pathways (DrGaP) (16) with a (false discovery rate (FDR) ≤0.01) was used to identify genes and driver signaling pathways on muated genes shared by the two rhabdoid cancers.

**Sanger sequencing validation**. Direct DNA sequencing of PCR-amplified DNAs served to validate *CROCC* mutation identified by WES **(Supplemental Fig. 3 and Supplemental Table 6,)** as described (56).

***CROCC* targeted sequencing.** An Ion Ampliseq Custom panel (Thermofisher) was devised, using AmpliSeq designer v2.1, to target the coding regions and flanking intron–exon junctions of *CROCC* (NM\_014675.3) and *SMARCB1* genes **(Supplemental Table 7, A and B).** Twenty nanograms of DNA were used per multiplex PCR. The quality of obtained libraries was evaluated by the Agilent 2100 Bioanalyzer on chip electrophoresis. Emulsion PCR was performed with the OneTouch OT2 system (Thermofisher). Sequencing was run on Ion Torrent Personal Genome Machine (PGM, Thermofisher) loaded with 318 chips v2. Filtered variants were annotated using a custom pipeline based on vcflib (https://github.com/ekg/vcflib), SnpSift, the Variant Effect Predictor (VEP) softwares and NCBI RefSeq database. Alignments were visually verified with the Integrative Genomics Viewer: IGV v2.3 (Broad Institute) (57-59).

**LOH analysis at *CROCC* and *SMARCB1* chromosomal loci.** DNA from rhabdoid tumour/normal matched samples were analyzed for loss of heterozygosity (LOH) at the *CROCC* locus (1p36.13) using three markers: D1S3391 (proximal), D1S1443 (median), D1S3669 (distal). Markers at the *SMARCB1* locus (22q11.2) were D22S301 and D22S345 (60).LOH analysis was done using fluorescent forward primers, followed by fragmental analysis compared to normal mucosa on Rotor-gene Q 5plex HRM (Qiagen) as described (61).

**DNA mismatch repair analysis.** DNA mismatch repair was analysed by microsatellite instability (MSI) molecular test using polyA markers and tissue immunostaining for DNA mismatch repair proteins Mlh1, Msh2, Msh6, and Pms2, as described (27).

**CpG island methylator phenotype analysis**. Sodium bisulphite modified DNA was subjected to quantitative methylation-specific PCR for 6 genes (*CDKN2A*, *IGF2, MLH1, NEUROG1, RUNX3, SOCS1*) and tumours with at least three methylated loci, with methylation levels greater than 15%, were classified as CpG island methylator phenotype (CIMP)-positive (27,44).

***KRAS* and *BRAFV600E* mutation analysis.** *KRAS* (exons 2, 3 and 4) and *BRAF* mutations were investigated by Sanger sequencing on DNA from FFPE tumour tissues as described (56).

***CROCC* mRNA expression analysis on FFPE tissues and CRC cell lines.** Total RNA was isolated from three 10-µm FFPE paired tumour/normal tissue sections using RNeasy FFPE Kit (Qiagen). After DNase I treatment, reverse transcription was performed using the first-strand cDNA synthesis kit (General Electric). For RT-qPCR analysis, a total cDNA amount corresponding to 10–20 ng of starting RNA was used for each reaction. Primers for RT-PCR are listed in **(Supplemental Table 6).** Fast SYBR Green Master Mix (Lifetechnologies) and 10 μM for each primer pair were used. The relative amount of mRNA was calculated using the comparative Ct method after normalization to rRNA 18S expression. qPCR reactions were performed on an Applied Biosystems 7500 Real-Time PCR machine using the standard amplification protocol. All reactions were performed in duplicate.

**Immunostaining of tissues and cell lines for centrosome markers and FISH analysis.** The primary antibodies used are listed in **(Supplemental Table 8).** Staining was performed as previously reported for immunohistochemistry (27) and immunofluorescence (62).Fallopian tubes were used as positive control for CROCC immunohistochemistry. In normal cells, CROCC and γ-tubulin staining consisted of 1 or 2 dot-like signals. The expression pattern in cancer was stratified into three categories: a) Loss, <1 dot per cell; b) Normal, 1-2 dots per cell; c) Amplified, >2 signals per cell. Two hundred cells in triplicate sections were blindly scored by three authors (A.R, E.M, M.P). CROCC and γ-tubulin staining was also considered structurally abnormal if showed: 1) a diameter greater than twice the diameter of dots present in normal epithelium within the same section; 2) changes in shape, size or perinuclear position, 3) a diffuse intense staining in the cytoplasm of individual tumour cells, as described (19,63). For fluorescent in situ hybridization (FISH), centromeric probes for chromosomes 1, 12 and 17 (CEN1, MDM2/CEN12, ERBB2/CEP17 respectively, ZytoVision, Bremerhaven, DE) were used to visualize and determine chromosome number alterations in interphase cells. The results were interpreted following enumeration of the signals in >500 nuclei by two independent experiments.

**Immunoblotting.** Immunoblots using protein extracts from matched tumour/normal fresh-frozen tissues or cell lines were performed as reported (27). Proteins from FFPE tissues were isolated using Qproteome FFPE Tissue Kit (Qiagen). The primary antibodies used are listed in **(Supplemental Table 8).** The secondary antibodies were anti-mouse sc-2031 and anti rabbit sc-2004 (Santa Cruz).

**Use of the Public Databases as a Reference.** As rhabdoid colorectal cancer is a rare variant of colorectal cancers, no public databases are available as a reference. To identify novel candidate genes, we compared common mutated genes between rhabdoid colorectal cancers found in our analysis with whole exome sequencing (WES) data from patients with classical colorectal cancer by using TCGA data set (https://tcga-data.nci.nih.gov) for which RNA-sequencing data was also available (14,64). *CROCC* mutations were filtered for exons and at splice sites (±3 bp). Ploidy was estimated from TCGA data set by calculating the weighted median copy number across all copy number segments, with weights equal to the segment length. To estimate *CROCC* mRNA variation in tumor tissues we selected colorectal cancer data sets for which gene expression from normal colon epithelium was also available. To this end, we compared *CROCC* gene expression changes found in our analysis with data from human gene-expression arrays from the Gene Expression Omnibus (GEO) database (GSE20916, GSE41258) (<http://www.ncbi.nlm.nih.gov/geo>), TCGA (14), and cBioPortal for Cancer Genomics (<http://www.cbioportal.org>). To identify a relation between *CROCC* expression and the level of chromosome instability (CIN), an independent dataset **(**GSE30540) of colorectal carcinomas characterized for CIN phenotype (CIN-high and CIN-low) was analyzed (32). The Cancer Cell Line Encyclopedia (CCLE) was used to visualize DNA copy number, mRNA expression and mutation data (25). Colorectal cancer cell lines with a mutation prevalence >25 per 106 bases and a chromosome copy number 2n were indicated as CIN low (25,65). By contrast, nonhypermutated cells with a chromosome copy number ranging from 2n to 4n were classified as CIN high (25,65). To define the gene-array expression thresholds used to separate cancer cell lines harboring 1p36.13 deletion from those retaining 1p36.13 locus, supervised clustering analysis with FDR<0.01 based on Spearman and Ward linkage was used. Genomics of Drug Sensitivity in Cancer project (Sanger panel, http://www.cancerrxgene.org/) was used to see if cell lines are more sensitive to specific drugs. The deriving *P* values were adjusted for multiple testing with Benjamini and Hochberg method (66).

**Gene ontology enrichment analysis.** The enrichment analysis was performed with the TopGO package using the classic algorithm and Fisher’s test with the same cutoffs described above (25,65,66). Genes tested for differential expression were used as the background. To create enrichment treemaps, parent categories that had enriched chil­dren were first removed, and maps were then created with the Treemap pack­age, color coding categories according to the combination of non-overlapping parent categories accounting for the largest proportion of plotted categories. All reported *P values* were calculated using the Benjamini-Hochberg method. To understand causal connections between diseases, genes and networks of upstream or downstream regulators the samples were subjected to Ingenuity Pathways Analysis (IPA, Ingenuity Systems) which was used as a starting point for building biological networks.

***CROCC* silencing experiment.** RKO cells at 70-80% confluence were transiently transfected with SureSilencing control or CROCC shRNA expression plasmids KH23140P (Qiagen) containing the puromycin resistance cassette. After selection with 0.8 μg/ml puromycin (Thermofisher) for 1 week, single colonies were amplified and assessed for efficient CROCC silencing by quantitative PCR (qPCR).

***CROCC* rescue experiment.** T84 and HT29 colon cancer cells were transfected with the full-length CROCC coding sequence “clone 6150861 pEGFP Rootletin, Nigg pFL2-CW499” (gift of Prof. Erich Nigg, Biozentrum of the University of Basel, Switzerland) or a truncate form (1–494aa) cloned with GFP epitope or GFP alone (used as control); after selection in 2 mg/ml G418 CROCC-GFP expressing cells were flow-sorted (11). For long-term experiments CROCC-GFP+ cells were maintained in 0.6 mg/ml G418. All transfections were performed with Lipofectamine 3000 (Thermofisher).

**Metaphase spreads and clonal FISH preparation.** For metaphase spreads, CRC cells were collected after 1-h treatment with 10mM colcemid Gibco KaryoMAX Colcemid (Lifetechnologies), and swelled with KCl (0.4%, 37 °C, 7 min) before fixation in 3:1 methanol:acetic acid. Cells were dropped onto glass slides and aged for 2 weeks. For clonal FISH, 500 cells were expanded into colonies of 30–60 cells on glass slides before KCl treatment and fixation in 3:1 methanol:acetic acid. Slides were denatured (70 uC in 23 sodium citrate (SSC)/75% formamide, 2 min, quenched in ice-cold 70% ethanol) and dehydrated through an ethanol series. Subsequently, the cells were stained with 49,6-diamidino-2-phenylindole (DAPI) and anti-centromere antibodies (ACAs) or using alternatively the alpha-satellite repetitive centromere DNA for chromosome X (CEPX, Abbott). The probes were denatured (90 uC, 6 min) and hybridized to slides (16 h, 37°C), and then washed. Slides were dehydrated and mounted in Vectashield hardset plus DAPI mounting medium (H-1500).

**Nocodazole washout assay.** Cells were incubated for 5, 15 and 30 min with the microtubule destabilizer nocodazole (10 μg/ml) at 37°C, washed five times with PBS at room temperature. To determine recovery times, slides were then fixed in -20°C methanol. Microtubule structures were detected using antibodies against anti γ-tubulin (1:500 in 1% BSA for 1 hr). The secondary antibody was a FITC labeled sheep antimouse antibody (1:500, 1 h, Sigma-Aldrich) and cells counterstained with propidium iodide. Asters were counted when clear signals at the periphery of the nucleus were detected. The completion of microtubule regrowth was scored when the pattern of immunofluorescence resembled the one prior to the nocodazole induced depolymerization. The amount of time required for the reformation of microtubules from the endogenous tubulin was assessed for normal human fibroblasts BJ cells used as control.

**CAsy cell counter and Proliferation assay.** Number and density of viable cells were determined using CAsy Cell Counter (Roche Innovatis). Each cell suspension was prepared three times in CAsyTon buffer (Roche Innovatis), followed by triplicate measurements of 200 μl sample volume. All counts of a size smaller than 10 μm (dead cells and debris) were excluded. For cell proliferation assay, 5,000 cells/well were seeded in 96-well plates. The number of proliferating cells was evaluated by 3-(4,5-dimethyl-2- thiazolyl)-2,5-diphenyl-2H-tetrazolium bromide (MTT). At the indicated times, MTT solution (Sigma-Aldrich) in complete medium (0.28-mg/ml final concentration) was added and incubated at 37°C for 4 h. The medium was discarded, and the formazan salts were dissolved in 4 mM HCl, 0.1% NP40 in isopropanol. The colorimetric substrate was quantified at 560 nm in an enzyme-linked immunosorbent assay plate reader.

**Cell morphology, wound healing and invasion assay.** Cell motility was evaluated by the wound-healing assay. Briefly, cells were grown to confluence and a wound made through the monolayer using a p1000 tip. Accurate measures of the wounds were taken during the time course to calculate the migration rate according to the equation: *percentage wound healing = ((wound length at 0 h) - (wound length at 24, 48 or 72 h))/(wound length at 0 h) x 100.* Phase-contrast images were acquired every2 h for 24 h and the percentage of migrating cells was determinedautomatically. Three independent experiments were performed. For invasion assay,2×104 cells were added to the upper compartment of a 24-well BioCoat Matrigel Invasion Chamber (Corning) in serum-free DMEM. After 24 h, invading cells were fixed, stained with crystal violet 0.1% and counted. Hematoxylin&Eosin staining was used to reveal morphological changes.

**Flow cytometry and clonogenic assay analysis.** Cell cycle analysis was performed three days after seeding on both attached and floating cells using the BD Cycletest Plus DNA reagent Kit (Cat. 340242; BD Biosciences). Propidium Iodide stained cells (>20.000 events) were analyzed by flow cytometry on FACSVerse (BD Biosciences) equipped with FACSuite Software v.1.0.5.3841 (BD Biosciences). Debris and doublet cells were excluded and only single cells were considered for cell cycle analysis. Results were reported as percentage of cells in G1, S and G2/M phases calculating mean ± standard deviation (SD) of different biological replicates. For colony formation assay, control or transfected cells were seeded at low confluence. Next, cells were fixed, stained, and photographed after 10 days of culture.

**Statistics.** Patient subsets were compared for survival outcomes, using both Kaplan-Meier survival curves and Cox proportional hazards method. Differences in Kaplan-Meier curves were tested for statistical significance using the log-rank test. Insight into molecular mechanisms underlying rhabdoid colorectal cancer (RC) has been difficult because of its extreme rarity.Our analysis identified a total of 23 RC for whom follow-up or molecular data were available for 20 (86.9%) and 9 (39%) cases, respectively **(Supplemental Table 9).** This suggests that our cohort of RCs was the largest used so far and highly informative. Overall Survival (OS) was defined as the time elapsed between the first colon surgery and death. For the metastatic series, OS was defined as the time elapsed between the first chemotherapy and death as reported (27,29). We censored those patients who were alive without tumour recurrence or dead at last contact. Expression levels, genetic or phenotypic changes among subgroups, were evaluated using boxplots and tested for statistical significance using a 2-sample t-test (2-tailed). Pearson’s t-test and odds-ratios (OR) together with 95% confidence intervals (CI) were used to compare expression profiles. Wilcoxon-Mann-Whitney and Kruskal-Wallis tests with median differences at 95% confidence interval (CI) were also used. The Spearman rank test was used to assess the correlation between continuous variables, and the Pearson X2 test for the association between categorical variables. Data are presented with mean, medians and ranges. The *P* values were calculated two sided. Statistical analyses were conducted were performed by GeneSpring R/bioconductor v.12.5 and R based package, SPSS v15 and GraphPad Prism 5.

**References**

42 Lee, S. H. *et al.* Rhabdoid colorectal carcinomas: reports of two cases. *Korean J Pathol* **47**, 372-377 (2013).

43 Samalavicius, N. E. *et al.* Rhabdoid carcinoma of the rectum. *Ann Coloproctol* **29**, 252-255 (2013).

44 Pancione, M. *et al.* Right-sided rhabdoid colorectal tumors might be related to the serrated pathway. *Diagn Pathol* **8**, 31 (2013).

45 Mafficini, A. *et al.* Reporting Tumor Molecular Heterogeneity in Histopathological Diagnosis. *PLoS One* **9**, e104979 (2014).

46 Simbolo, M. *et al.* DNA qualification workflow for next generation sequencing of histopathological samples. *PLoS One* **8**, e62692 (2013).

47 Zamo, A. *et al.* Application of microfluidic technology to the BIOMED-2 protocol for detection of B-cell clonality. *J Mol Diagn* **14**, 30-37 (2012).

48 Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci U S A* **107**, 16910-16915 (2010).

49 Douville, C. *et al.* CRAVAT: cancer-related analysis of variants toolkit. *Bioinformatics* **29**, 647-648 (2013).

50 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).

51 Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* **31**, 213-219 (2013).

52 DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491-498 (2011).

53 Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).

54 Dong, C. *et al.* Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet* **24**, 2125-2137 (2015).

55 Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639-1645 (2009).

56 Scarpa, A. *et al.* Molecular typing of lung adenocarcinoma on cytological samples using a multigene next generation sequencing panel. *PLoS One* **8**, e80478 (2013).

57 Cingolani, P. *et al.* Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Front Genet* **3**, 35 (2012).

58 McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069-2070 (2010).

59 Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26 (2011).

60 Jackson, E. M. *et al.* High-density single nucleotide polymorphism array analysis in patients with germline deletions of 22q11.2 and malignant rhabdoid tumor. *Hum Genet* **122**, 117-127 (2007).

61 Montgomery, J. L., Sanford, L. N. & Wittwer, C. T. High-resolution DNA melting analysis in clinical research and diagnostics. *Expert Rev Mol Diagn* **10**, 219-240 (2010).

62 Caracciolo, V. *et al.* Differential expression and cellular distribution of gamma-tubulin and betaIII-tubulin in medulloblastomas and human medulloblastoma cell lines. *J Cell Physiol* **223**, 519-529 (2010).

63 Teixido-Travesa, N., Roig, J. & Luders, J. The where, when and how of microtubule nucleation - one ring to rule them all. *J Cell Sci* **125**, 4445-4456 (2012).

64 Sadanandam, A. *et al.* A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat Med* **19**, 619-625 (2013).

65 Mouradov, D. *et al.* Colorectal cancer cell lines are representative models of the main molecular subtypes of primary cancer. *Cancer Res* **74**, 3238-3247 (2014).

66 Noble, W. S. How does multiple testing correction work? *Nat Biotechnol* **27**, 1135-1137 (2009).