Childhood Cancer Survivor Study (CCSS)
Concept

Title: The Genomic Landscape of Subsequent Malignant Neoplasms in Childhood Cancer Survivors

Working Groups: Genetics (primary); Second Malignancy (secondary); Epidemiology/Biostatistics (secondary)

Roster of Investigators

Investigator	Role	Expertise	Institution	Contact
Samuel W. Brady, PhD	PI	Cancer genomics	St. Jude Children's Research Hosp.	samuelw.brady@stjude.org
Greg T Armstrong, MD, MS	PI	Survivorship, PI, CCSS	St. Jude Children's Research Hosp.	greg.armstrong@stjude.org
Michael Arnold, MD	Co-I	Pathology	Children's Colorado	michael.arnold@childrenscolorado.org
Miriam Conces, MD	Co-I	Pathology	Nationwide Children's Hospital	miriam.conces@nationwidechildrens.org
Joseph P. Neglia, MD, MPH	Co-I	SMN	Univ. of Minnesota	jneglia@UMN.edu
Lucie Turcotte, MD	Co-I	SMN	Univ. of Minnesota	turc0023@UMN.edu
Wendy Leisenring	Co-I	Biostatistics	Fred Hutchison Cancer Center	wleisenr@fredhutch.org
Smita Bhatia, MD, MPH	Co-I	Genetic epidemiology	University of Alabama at Birmingham	smitabhatia@uabmc.edu
Jinghui Zhang, PhD	Co-I	Cancer genomics	St. Jude Children's Research Hosp.	jinghui.zhang@stjude.org
Wentao Yang, PhD	Co-I	Cancer genomics	St. Jude Children's Research Hosp.	wentao.yang@stjude.org
Yutaka Yasui, PhD	Co-I	Statistical genetics	St. Jude Children's Research Hosp.	yutaka.yasui@stjude.org
Les Robison, PhD	Co-I	Survivorship	St. Jude Children's Research Hosp.	les.robison@stjude.org

BACKGROUND AND SIGNIFICANCE

Subsequent Malignant Neoplasms are a Relatively Common and Potentially Fatal Consequence of Primary Cancer Therapy. Although >85% of children diagnosed with cancer will become long-term (>5-year) survivors[1,2], their risk for a subsequent malignant neoplasm (SMN) is five times greater (standardized incidence ratio 5.4, 95% CI 5.1-5.7) than risk for malignancy in the general population[3]. Additionally, SMNs are among the most feared of late effects as they are the most common non-relapse cause of late mortality, accounting for approximately half of non-relapse deaths among 5-year survivors of childhood cancer[4,5]. Other than non-melanoma skin cancer, breast cancer, thyroid cancer, and meningioma (both benign and malignant) are the most common subsequent neoplasms in pediatric cancer survivors[6]. While thyroid malignancies and meningiomas are often responsive to therapy[7,8], 27% of female survivors of childhood cancer will not survive a subsequent breast cancer due in part to the fact that survivors commonly have high cumulative exposures (from childhood) to key treatments such as chest-directed radiation and anthracycline chemotherapy, limiting therapeutic options[9]. New targets are needed.

Treatment-Related Risk Factors are Well-Established, but the Genomic Landscape is Unknown. Risk factors for SMNs include (1) prior exposure to DNA-damaging therapies, including radiation[10], alkylating agents[11], anthracyclines[11], and epipodophyllotoxins[12–14], and (2) less frequently, germline predisposition variants[15,16]. Specifically, breast cancer as an SMN is associated with higher doses of chest-directed radiation and anthracycline chemotherapy[17], thyroid SMNs with dose of radiation to the neck/thyroid and alkylating agent exposure[18], and meningiomas with higher dose of radiation to brain meningeal tissue[19]. While associations with treatment exposures are well established, little is known about the somatic genomic landscapes of SMNs among pediatric cancer survivors and how specific therapies shape these landscapes. Elucidating this could inform underlying mechanisms of SMN development and provide new targets for future therapies or reductions/alternatives to primary therapy that could reduce risk for SMNs. Further, while most cytotoxic therapies for pediatric cancer alter immune function[20], the impact of therapy-associated immune modulation on SMN development has not been previously evaluated.

The CCSS is a Unique Resource for Evaluation of SMN Etiology. Since 1994, CCSS has confirmed self-reported SMNs with a pathology report and to date collected 645 formalin-fixed paraffin-embedded (FFPE) SMN somatic tissue specimens. The Childhood Cancer Data Initiative (CCDI) of the National Cancer Institute recently funded genomic sequencing of these SMN samples. This includes paired tumor/germline whole-genome

sequencing (WGS) and tumor whole-exome sequencing (WES) on breast cancer (n=111) and thyroid cancer SMNs (n=79) completed in October 2021, with tumor RNA-Seq and germline WES ongoing and expected to be completed by December 2021. Sequencing of CCSS meningiomas by WGS, WES, and RNA-Seq (up to n=111 meningiomas, including both benign and malignant) is also ongoing and expected to be completed by March 2022. This proposal leverages this rich data resource to evaluate the diverse types of somatic variants in breast, thyroid, and meningioma SMNs, elucidate mutational signatures, model the ordering of mutation acquisition, and study immune signatures in SMNs.

SPECIFIC AIMS

Aim 1: Assess the effects of prior therapy on the genomic landscapes of breast, thyroid and meningioma SMNs among survivors of childhood cancer.

- **Aim 1a.** Call variants and filter FFPE artifacts.
- **Aim 1b.** Identify driver genes in breast, thyroid and meningioma SMNs and compare drivers across prior treatment exposures/doses and vs. *de novo* cancers.
- **Aim 1c.** Elucidate the mutational signature landscape of SMNs.

Aim 2: Model the timing of mutation acquisition in SMNs among survivors of childhood cancer.

- **Aim 2a.** Model the timing of mutation acquisition in SMNs.
- **Aim 2b.** Compare timing of mutation acquisition in SMNs to corresponding *de novo* cancers.

Exploratory Aim: Elucidate the immune infiltration signatures of SMNs among survivors.

In summary, this work will uncover the novel genomic landscape of common SMNs and how primary cancer therapy shapes that landscape, which could provide a model for evaluating other SMNs in the future. Together, this work will improve the understanding SMN etiologic mechanisms, and potentially identify targets for treatment and prevention of SMNs, and their often-fatal consequences for survivors of childhood cancer.

PRELIMINARY DATA

Identification of therapy-induced mutational signatures. We have previously employed mutational signature analysis to discover the impact of therapy-induced mutagenesis on pediatric cancer genomes, and this experience will facilitate analysis of treatment-induced alterations in SMNs. Indeed, each mutational process/cause, such as smoking, ultraviolet light, or prior therapy leaves a distinct mutational imprint (signature) on the genome based on a preference to mutate at specific sequence motifs[21], and thus genomic analysis can reveal specific causes of mutagenesis. We previously identified a novel mutational signature induced by thiopurine treatment in relapsed acute lymphoblastic leukemia (ALL), and showed that thiopurine treatment can induce driver mutations including in *TP53*[22,23]. We also found that platinum treatment can nearly double the mutation burden in osteosarcoma[24], and identified thiopurine and platinum mutational signatures in secondary myeloid neoplasms[25] and small numbers of solid tumor SMNs[26,27]. In summary, this work suggests that prior treatment exposures in pediatric cancer survivors may induce mutations in normal cells that promote SMN development, which could be revealed by mutational signature analysis.

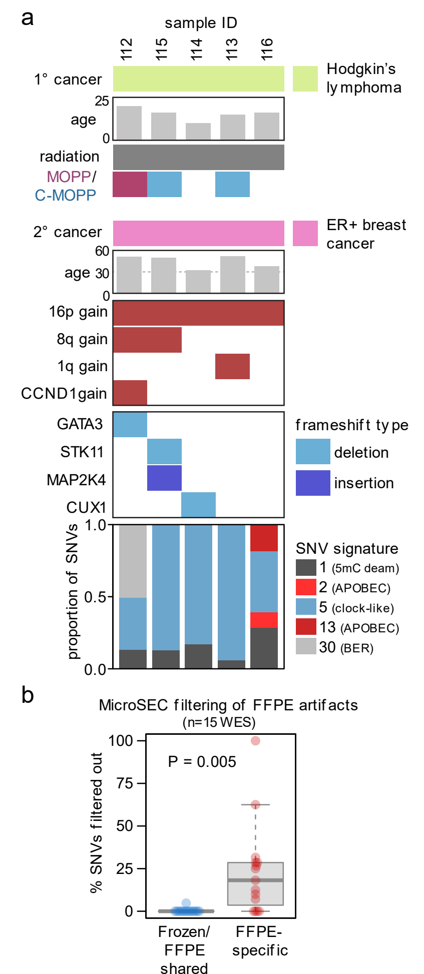

Figure 1. Preliminary analysis of FFPE samples. (a) Genomics of 5 FFPE SMNs. (b) MicroSEC filters FFPE artifacts based on matched FFPE/frozen WES of 15 pediatric cancers. FFPE-specific SNVs are primarily removed.

Preliminary genomic analysis of FFPE-derived SMN tissue. To confirm that our analytical approach is applicable for SMN studies from FFPE tissue, we

performed WGS and WES of 5 breast cancer SMNs from CCSS, and matched germline tissue. As expected for breast cancer we observed driver loss-of-function mutations in *GATA3, STK11, MAP2K4*, and *CUX1*[28]; chromosome 16p and 8q copy gains[28]; and the APOBEC mutational signature[29] (Fig. 1a). The small sample size (n=5) was insufficiently powered to identify novel driver genes or therapy-related mutational signatures, which may be uncovered when analyzing a larger sample set. Further, given that formalin induces artifactual variants[30,31], we have shown that new methods such as MicroSEC[32] can identify likely formalin artifacts in a series of 15 pediatric cancers with WES of matched frozen and FFPE tissue (Fig. 1b). These preliminary analyses indicate the feasibility of genomic variant detection from SMN FFPE tissues.

INNOVATION

Whole Genome Sequencing to Assess SMN Genomic Landscape. Comparable work on this scale has not been undertaken to study SMN whole genomes in pediatric cancer survivors, aside from subsequent myeloid leukemias[25] which are under-represented in CCSS. For example, The Cancer Genome Atlas (TCGA) thyroid cancer study included WES, rather than WGS, of only 26 SMNs (among ~400 evaluated samples)[33], with unspecified age at primary diagnosis. Most SMN studies have used primarily targeted and/or exome sequencing of small cohorts (n=25-36 samples[34–38]) and among these only studies of glioma and meningioma SMNs have focused on survivors of childhood cancer[36–38]. WGS has only been used for solid tumor SMNs, to our knowledge, in a study of 12 radiation-induced gliomas[37] and in a study of 12 radiation-associated SMNs of various histologies (primarily sarcomas)[39]. The latter demonstrated the utility of WGS in SMN analysis, as it revealed radiation-associated indel and rearrangement signatures, and driver alterations in *TP53* and *CASP8* which were likely radiation-induced[39] and unlikely to be detected by WES. WGS of this clinically annotated SMN population of survivors of childhood cancer will provide, to our knowledge, the most comprehensive assessment of the full landscape of breast, thyroid, and meningioma SMN driver alterations, including single-nucleotide variants (SNVs), indels, copy number, structural variants (SVs), and mutational signatures induced by therapy, as WES inefficiently detects SVs and mutational signatures.

Use of Innovative Analytical Approaches for both Intra- and Inter-Cohort Mutational Signature Analyses. Our approach to mutational signature analysis may enable identification of novel therapy-induced mutational signatures and their etiology, through dose-response association of mutational signatures in SMNs with their prior treatment exposures (intra-cohort statistical analysis), and by comparing SMN signatures to signatures in *de novo* breast cancers, thyroid cancers, and meningiomas (from the International Cancer Genome Consortium [ICGC] and TCGA and published studies[28,33,40]; inter-cohort analysis). We have previously used this approach to discover that a novel mutational signature in relapsed leukemia was induced by thiopurine treatment[22]. Additionally, mutational signature analysis can predict driver mutations likely induced by specific therapies, as our prediction that thiopurine treatment induces *TP53* mutations in relapsed leukemias was experimentally validated[23]. We will employ this technique to identify which driver variants are likely therapy-induced in SMNs. We will also perform single-sample evolutionary timing analysis using WGS[41], as we have done previously[42], to construct the evolutionary history of SMNs and compare them to *de novo* cancers. These evolutionary and mutational signature analyses will together offer unprecedented resolution into the therapy-induced evolution of SMNs and may suggest approaches to prevent their development. Finally, our exploratory analysis of the immune infiltrate of SMNs using RNA-Seq (and comparison with *de novo* cancers) may identify immunomodulatory interventions to prevent or treat SMNs. In summary, these analyses will elucidate the impact of therapy on SMN development, which may facilitate improved prevention, surveillance, and treatment approaches.

ANALYSIS FRAMEWORK

Patient cohort and sequencing. Among 334 pediatric cancer survivors with breast or thyroid SMNs in the CCSS Biopathology Center, WGS and WES have been completed and RNA-Seq is in process (expected December 2021) on 111 breast (of 218 available) and 79 thyroid SMNs (of 116). Matched germline WGS has also been performed for all samples with germline WES underway and expected December 2021. Primary cancers were most frequently Hodgkin lymphoma, sarcoma, and leukemia, with selected treatment exposures noted in Table 1. Sequenced SMNs had ≥30% pathologically-assessed tumor purity and ≥75% non-necrotic tissue in tumor regions, and DNA sequencing was performed by Hudson Alpha (60x WGS and 100x WES using an Illumina NovaSeq 6000 instrument), while RNA-Seq is being performed by the St. Jude Department of

Computational Biology genomics lab. For each survivor, treatment exposures/doses have been extensively characterized through medical record abstraction (Table 1; deidentified via an honest broker) including cumulative dosage for 29 chemotherapeutic agents and both region- and organ-specific radiation dosimetry based on 3D phantoms[43], for determination of treatment associations with therapy-induced mutational signatures and driver alterations. Matched tumor/germline WGS and WES are also underway for 111 meningioma SMNs, along with RNA-Seq of somatic tissue, and medical record abstraction will be performed on all meningiomas with successful sequencing data.

A control cohort of non-SMN (*de novo*) breast and thyroid cancers will be obtained from the ICGC and TCGA, accessible through the ICGC Data Portal, which we have used previously[22], including *de novo* breast (n=506 WGS samples) and thyroid cancers (n= 794; 50 WGS and 744 WES). Control meningioma WES data will be obtained from a published study including 15 *de novo* meningiomas[40]. This will be used to compare driver alterations and other features between SMNs and *de novo* cancers.

SMN clinical features		Breast SMN		Thyroid SMN	
		n	%	n	%
	Total	111	100	79	100
Age (1° diag)	0-9	16	14.4	44	55.7
	10-20	95	85.6	35	44.3
1° Diagnosis	Acute lymphoblastic leukemia	10	9.0	23	29.1
	Acute myeloid leukemia	5	4.5	2	2.5
	Other leukemia	2	1.8	5	6.3
	CNS tumor	1	0.9	9	11.4
	Hodgkin lymphoma	57	51.4	19	24.1
	Non-Hodgkin lymphoma	6	5.4	5	6.3
	Sarcoma	21	18.9	12	15.2
	Other solid tumor	9	8.1	4	5.1
Breast/chest RT	Yes	86	77.5	45	57.0
	No	15	13.5	15	19.0
	Unknown	10	9.0	19	24.1
Thyroid RT	Yes	87	78.4	60	75.9
	No	15	13.5	15	19.0
	Unknown	9	8.1	4	5.1
Alkylator	Yes	43	38.7	47	59.5
	No	52	46.8	22	27.8
	Unknown	16	14.4	10	12.7
Anthracycline	Yes	45	40.5	33	41.8
	No	58	52.3	39	49.4
	Unknown	8	7.2	7	8.9

Table 1. Clinical features of sequenced SMNs. Therapy given prior to SMN diagnosis is noted. RT, radiation.

Outcomes of interest. We expect to determine (1) the landscape of driver alterations across breast, thyroid, and meningioma SMNs, (2) how these drivers' frequencies differ between SMNs and *de novo* cancers, (3) how therapy-induced mutagenesis shapes the SMN genome through mutational signature analysis, (4) how the timing of mutation acquisition differs between SMNs and *de novo* cancers, and (5) how the immune infiltrate of SMNs differs from *de novo* cancers.

Specific Aim 1: Assess the effects of prior therapy on the genomic landscapes of breast, thyroid and meningioma SMNs among survivors of childhood cancer.

Aim 1a. *Call variants and filter FFPE artifacts.* We will first map WGS and WES raw sequencing reads to the human genome using Burroughs-Wheeler Alignment (BWA)[44] and then identify somatic variants, including SNVs (using Bambino[45]), indels (MuTect[46] and Strelka[47]), copy number alterations (or CNAs; CONSERTING[48] for WGS and CNVkit[49] for WES), and structural variants (or SVs; CREST[50] and Manta[51]). RNA-Seq data will be aligned with STAR[52] in two-pass mode. False-positive indels will be filtered by using the intersection of two indel callers (MuTect[46] and Strelka[47]) and using a novel indel post-processor (indelPost[53]) we developed. Somatic SNVs and indels in coding regions will be manually curated by inspecting raw sequencing data in BAM files to remove artifacts and filtered by intersecting WGS and WES calls as needed. Germline SNVs and indels will also be identified and pathogenicity will be assessed with PeCanPIE, our Cloud-based tool that classifies the germline variant pathogenicity[54], and ACMG guidelines. Somatic CNAs derived from WGS will be corroborated with WES-based CNA profiles using CNVkit[49], which we have found can reduce artifactual CNA segmentation from FFPE samples to identify arm-level CNAs. Somatic SVs will also be corroborated with CNA data and cross-validated with RNA-Seq data to observe effects on expression. Fusion genes identified from SV data will be corroborated with RNA-Seq fusion analysis (CICERO[55] and ARRIBA[56]). We will also analyze kataegis and chromothripsis as we have done previously[57,58].

Further, as the CCSS SMN samples are stored as FFPE tissue, we will perform additional filtering to remove artifactual variants caused by formalin fixation. These artifacts include characteristic C>T variants with a low variant allele fraction, artifacts induced by fragmented DNA[30], and microhomology-related artifacts which can be filtered with MicroSEC (Fig. 1b)[32] and other approaches. We will use 15 previously sequenced tumors (Fig. 1b), analyzed from both frozen and matched FFPE tissues at St. Jude, as a training set to further identify FFPE

artifacts. This analysis will be facilitated by our prior experience identifying artifact-induced mutational signatures[42] and with artifact removal in general[59,60].

Aim 1b. *Identify driver genes in breast, thyroid and meningioma SMNs and compare drivers across prior treatment exposures/doses and vs. de novo cancers.* After filtering somatic variants, we will identify significantly mutated genes (SMGs) and CNA regions in breast, thyroid and meningioma SMNs using MutSigCV[61], GRIN[62], and GISTIC[63], as we have done previously[42,60]. Our breast cancer SMN analysis (n=111) has 80% probability to detect a recurrent gene alteration in ≥5, ≥4 or ≥3 of 111 samples if the alteration exists in 6.0%, 5.0%, or 3.9% of samples or more, based on binomial sampling. The corresponding values in the thyroid SMN analysis (n=79) are 8.4%, 6.9%, and 5.4%. Power analysis is not calculated for meningiomas as the number of samples that will be successfully sequenced (passing quality control metrics, etc.) among the 111 available samples is not yet known.

We will next summarize the most common drivers across each SMN disease type, using novel SMGs identified and known[28,33] driver genes. To determine the impact of treatment on SMN genomic profiles, we will stratify breast, thyroid, and meningioma SMNs by their somatic driver alterations and assess each driver's association with treatment exposures/doses (e.g., radiation treatment, anthracycline treatment) by Fisher's exact test for dichotomous exposures and Mann-Whitney test for continuous doses. We expect differing prior treatments to result in different mutational landscapes due to therapy-specific mutagenesis. Table 2 shows the minimal detectable relative risk of a specific treatment associated with a specific driver among the 79 thyroid SMNs (the smaller sample size) with 80% power (1-sided, α=0.05); breast SMNs and continuous doses would have higher power. This indicates sufficient power to detect driver-treatment associations for a wide range of driver and treatment exposure frequencies. Thyroid SMN driver gene frequencies will also be compared by sex (56 females, 23 males).

Freq. of Thyroid SMN w/ vs. w/o the specific driver gene	% Tx exposed in SMNs w/o the specific driver gene			
	10%	30%	50%	70%
5 vs. 74	6.7	3.0	---	---
10 vs.69	5.1	2.5	1.8	---
15 vs. 64	4.3	2.3	1.7	1.4
20 vs. 59	3.9	2.1	1.6	1.4

Table 2. Minimal detectable driver gene-treatment relative risk in thyroid SMNs.

We will also compare SMN landscapes with those of *de novo* breast cancer[28], thyroid cancer[33], and meningioma[40]. For this analysis, we will analyze mutation (SNV/indel), SV, and CNA data from the ICGC and TCGA accessible through the ICGC Data Portal, which we have used previously[22], including *de novo* breast (n=506 WGS samples) and thyroid cancers (n=794; 50 WGS and 744 WES), plus WES of 15 *de novo* meningiomas from a published study[40]. Based on the influence of prior treatment, we expect specific genes or pathways to be mutationally enriched/depleted in breast, thyroid and meningioma SMNs compared to *de novo* cancers. As breast cancer, thyroid cancer, and meningioma are divided into transcriptional subtypes (e.g. luminal vs. basal breast cancer)[33,40,64] we will also compare the prevalence of these subtypes and their landscapes between SMNs and *de novo* cancers. Even in the lower-powered comparison of thyroid SMNs vs. *de novo* thyroid cancer, we have 80% power (2-sided, α=0.05) to detect a relative risk of 2.4, 2.1, and 2.0 of carrying a specific driver alteration, when 10%, 20%, and 30%, respectively, of the *de novo* cancers bear the mutation.

Aim 1c. *Elucidate the mutational signature landscape of SMNs.* Many DNA-damaging therapies induce characteristic mutational signatures, which we have detected in relapsed cancers[22–24], secondary myeloid cancers[25], and a few solid tumor SMNs[26,27]. Therefore, to identify the effects of prior therapy on SMN genomes, we will perform mutational signature analysis using WGS data. SNV signatures will be analyzed by ascertaining the trinucleotide context (5' and 3' flanking bases) of each SNV, resulting in 96 mutation types, followed by non-negative matrix factorization to discover novel signatures, and subsequent quantification of novel and known (COSMIC[65] v3.2) signatures in each SMN (using SigProfiler)[21,22]. Indel, dinucleotide, and SV signatures will also be analyzed[21,28]. We expect this to identify known and possibly novel therapy-related signatures in SMNs.

To identify the therapy-related causes of mutational signatures identified above, we will correlate each signature's presence/absence with prior therapeutic exposures, which may reveal specific therapies inducing novel signatures. We will also perform dose-response analysis comparing signature strength with doses of therapies received to identify signature etiologies[22]. Additionally, mutational signatures in SMNs will be compared with those from *de novo* breast and thyroid cancers from ICGC/TCGA (excluding meningioma as only WES, rather than WGS, is available for *de novo* meningiomas[40]) and with multiple relapsed cancer datasets with known treatment history, including ovarian[66] and breast cancer[29], neuroblastoma[42], and leukemia as we have done previously[22], to narrow down therapies potentially causing novel signatures. (For example, a novel signature detected in SMNs but absent in relapsed ovarian cancer is likely not caused by ovarian cancer therapies). The power analysis in Aim 1b applies to these analyses by replacing driver genes with mutational signatures.

Further, the specific driver variants caused by a DNA-damaging agent can be predicted by comparing the sequence context of mutations caused by the therapy to the sequence context of driver variants[22,25]. We have

used this approach to identify thiopurine-induced *TP53* variants in relapsed ALL, a prediction that was validated experimentally[23]. Therefore, we will mathematically predict the likelihood that each driver variant was induced by each mutational signature[22], which may reveal therapy-induced alterations leading to SMNs. These results may suggest alternative treatment regimens for primary cancers to prevent therapy-induced driver mutations[23].

Finally, we will integrate the above analyses with mutation timing analysis (Aim 2) and treatment history in individual patients to identify variants likely occurring before, during, or after primary cancer therapy.

Aim 1 Potential Problems and Alternatives. It may be necessary to develop machine learning methods to filter FFPE artifacts by identifying features of variants detected by both WGS and WES in SMNs (likely true-positives), or using published matched FFPE/fresh-frozen WGS datasets[67]. Further, artifact abundance varies by fixation time and sample age[68], and samples with excess artifacts may need to be excluded. Finally, race may need to be included as a covariate when comparing SMNs vs. *de novo* cancers if differing racial backgrounds are observed between the two groups (race can be inferred using germline WGS data).

Specific Aim 2: Model the timing of mutation acquisition in SMNs among survivors of childhood cancer.

Aim 2a. *Model timing mutation acquisition in SMNs.* Genomic analysis of single tumor samples can reveal the order of variant acquisition based on the following: (1) SNVs/indels with high variant allele fraction (VAF; mutant / total reads) occur evolutionarily earlier than those with low VAF[69] (Fig. 2a); (2) in copy-gained regions, SNVs/indels on ≥2 copies occurred before the copy gain, while those on 1 copy often occurred after[41] (Fig. 2b); and (3) certain mutational signatures have constant mutation rates over time (Fig. 2c), which can be used along with (2) to infer the absolute time (in years before diagnosis) of genomic aberrations in some cases[41,70]. We have used these principles to elucidate the evolution of pediatric cancers[22,42,71] revealing, for example, that copy gains in neuroblastoma precede most SNVs[42]. Remarkably, these single-tumor analyses can recapitulate evolutionary schemes known from sequential precancerous-to-cancerous sample analyses[41], such as Vogelstein's *APC-KRAS-TP53* colon model[72]. Further, they can reveal the likely time interval in which specific alterations occurred in a patient's lifetime[41], which may indicate alterations occurring during prior treatment.

We will therefore perform these evolutionary timing analyses on breast, thyroid, and meningioma SMNs using WGS and WES data and determine the likely time interval during which genomic aberrations occurred. We will compare this to prior treatment history to identify whether alterations occurred during treatment (potentially therapy-induced), before (related to therapy selection), or after. This will be corroborated with therapy-related mutational signature analysis (Aim 1). Power analysis is not conducted as this is not a statistical inferential analysis.

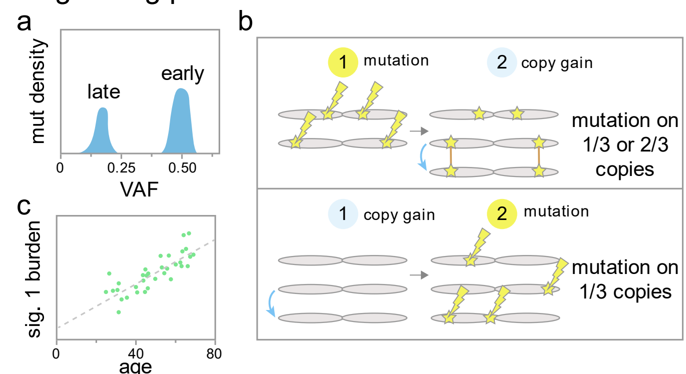

Figure 2. Principles used in timing analysis (simulated data).

Aim 2b. *Compare timing of mutation acquisition in SMNs to corresponding de novo cancers.* We will also perform evolutionary timing analysis on *de novo* breast cancer, thyroid cancer, and meningioma data from ICGC and TCGA and published work[40] (mutation data obtained as in Aim 1) and compare them to SMNs. We expect this will reveal different mutation acquisition trajectories in SMNs vs. *de novo* cancers (e.g., variants acquired with different ordering in SMNs vs. *de novo* cancers, such as on-treatment). This will indicate how therapy alters SMN evolution compared to *de novo* cancer evolution, which may suggest primary treatment adjustments to prevent SMN development. The Aim 1b power analysis applies here by replacing driver gene alterations with evolutionary trajectories.

Aim 2 Potential Problems and Alternatives. This analysis relies on accurate CNA profiles, sufficient SNVs detected in copy-gained regions, and sufficient tumor purity so that VAFs and CNA profiles can be detected with good resolution. Hence samples with unreliable CNA profiles, few copy gains, or low tumor purity will be excluded from aspects of this analysis.

Exploratory Aim: Elucidate the immune infiltration signatures of SMNs among survivors.

Pediatric cancer treatment regimens substantially modulate immune function, and it is not known how long immune modulation persists after treatment cessation[20], or whether it promotes SMN development. RNA-Seq can infer tumor-infiltrating immune cell abundance, cell type, and activation status in tumor specimens[73]. RNA-Seq is ongoing on all breast, thyroid and meningioma SMNs sequenced by WGS and WES from which we will analyze how prior treatment shapes SMN immune signatures.

As we are interested in tumor-infiltrating immune cells, we will exclude samples with low tumor purity which may have immune cells from adjacent tissues. First, samples with low purity (<70%) as inferred from WGS/WES-based somatic CNA profiles, allelic imbalance, and SNV VAFs[24,42,74] will be excluded. Second, RNA-Seq samples passing this filter will be subjected to additional checks, including verifying that VAFs of SNVs in RNA-Seq data are similar to VAFs detected by WGS/WES (as sequential tissue sections used for WGS/WES vs. RNA-Seq may vary in purity) and do not display evidence of adjacent normal tissue contamination in their expression profiles.

To compare immune signatures in SMNs vs. *de novo* cancers, we will obtain RNA-Seq from *de novo* ICGC and TCGA breast (n=1,080) and thyroid (n=500) cancers using the ICGC Data Portal, and from 15 *de novo* meningiomas from a published study[40]. We will process all raw RNA-Seq data with STAR[52] alignment and HTSeq[75] to obtain gene expression counts. SMN and *de novo* cancer data will be batch-corrected within each cancer type using ComBat[76] and iterative t-distributed stochastic neighbor embedding (t-SNE) analysis[26] until breast cancer, thyroid cancer, and meningioma transcriptional clusters overlap between *de novo* and SMN samples by t-SNE. We will then infer the relative abundance of tumor-infiltrating immune cell types and activation status using single-sample gene set enrichment analysis (ssGSEA) of published immune signatures from the Molecular Signatures Database (C7)[77,78], as done by others[73]. Importantly, RNA-Seq of FFPE samples accurately identifies pathway-level expression changes despite FFPE-associated noise in individual genes[79], indicating the feasibility of this multigene (ssGSEA)-based approach. We will also corroborate results with CIBERSORT, which infers immune cell composition from RNA-Seq[80]. We expect SMNs may have dampened immune activation compared to *de novo* cancers. To validate the immune signatures identified using RNA-Seq, we will perform spatial transcriptomics using the 10X Visium platform on 10 breast, thyroid, or meningioma SMNs as validation. We will correlate RNA-Seq-inferred immune signatures with spatial transcriptomics as validation.

Exploratory Aim Potential Problems and Alternatives. Batch effects may confound ICGC/TCGA vs. SMN comparisons. Therefore, we may also perform intra-SMN-cohort analysis of samples stratified by prior treatment to see if certain therapies correlate with decreased immune infiltration/activation.

Funding: We have applied for an R21 through PAR-20-277 to fund this proposal. St. Jude Developmental Funds also support this analysis through June 2022 with a possible one year extension, after which internal funds will be used if the R21 is not funded.

REFERENCES

1. Noone A.M., Howlader N., Krapcho M., Miller D., Brest A., Yu M., Tatalovich Z., Marriotto A., Lewid D.R., Chen H.S., Feuer E.J., Cronin K.A. (eds). SEER Cancer Statistics Review, 1975-2015, National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/csr/1975_2015/, based on November 2017 SEER data submission, posted to the SEER web site, April 2018.
2. Laetsch, T. W., DuBois, S. G., Bender, J. G., Macy, M. E. & Moreno, L. Opportunities and Challenges in Drug Development for Pediatric Cancers. *Cancer Discov.* **11**, 1–15 (2020).
3. Turcotte, L. M., Liu, Q., Yasui, Y., Arnold, M. A., Hammond, S., Howell, R. M., Smith, S. A., Weathers, R. E., Henderson, T. O., Gibson, T. M., Leisenring, W., Armstrong, G. T., Robison, L. L. & Neglia, J. P. Temporal Trends in Treatment and Subsequent Neoplasm Risk Among 5-Year Survivors of Childhood Cancer, 1970-2015. *JAMA* **317**, 814–824 (2017).
4. Mertens, A. C., Liu, Q., Neglia, J. P., Wasilewski, K., Leisenring, W., Armstrong, G. T., Robison, L. L. & Yasui, Y. Cause-Specific Late Mortality Among 5-Year Survivors of Childhood Cancer: The Childhood Cancer Survivor Study. *JNCI J. Natl. Cancer Inst.* **100**, 1368–1379 (2008).
5. Armstrong, G. T., Chen, Y., Yasui, Y., Leisenring, W., Gibson, T. M., Mertens, A. C., Stovall, M., Oeffinger, K. C., Bhatia, S., Krull, K. R., Nathan, P. C., Neglia, J. P., Green, D. M., Hudson, M. M. & Robison, L. L. Reduction in Late Mortality among 5-Year Survivors of Childhood Cancer. *N. Engl. J. Med.* **374**, 833–842 (2016).
6. Meadows, A.T., Friedman, D.L., Neglia, J.P., Mertens, A.C., Donaldson, S.S., Stovall, M., Hammond, S., Yasui, Y., Inskip, P.D. Second neoplasms in survivors of childhood cancer: findings from the Childhood Cancer Survivor Study cohort. *J. Clin. Oncol.* **27**, 2356–2362 (2009).
7. McLeod, D. S. A., Sawka, A. M. & Cooper, D. S. Controversies in primary treatment of low-risk papillary thyroid cancer. *Lancet* **381**, 1046–1057 (2013).
8. Preusser, M., Brastianos, P. K. & Mawrin, C. Advances in meningioma genetics: novel therapeutic opportunities. *Nat. Rev. Neurol. 2018 142* **14**, 106–115 (2018).
9. Moskowitz, C. S., Chou, J. F., Neglia, J. P., Partridge, A. H., Howell, R. M., Diller, L. R., Friedman, D. N.,

Barnea, D., Morton, L. M., Turcotte, L. M., Arnold, M. A., Leisenring, W. M., Armstrong, G. T., Robison, L. L., Oeffinger, K. C. & Henderson, T. O. Mortality After Breast Cancer Among Survivors of Childhood Cancer: A Report From the Childhood Cancer Survivor Study. *J. Clin. Oncol.* **37**, 2120 (2019).

10. Armstrong, G. T., Stovall, M. & Robison, L. L. Long-term effects of radiation exposure among adult survivors of childhood cancer: Results from the childhood cancer survivor study. *Radiation Research* vol. 174 840–850 (2010).
11. Henderson, T. O., Moskowitz, C. S., Chou, J. F., Bradbury, A. R., Neglia, J. P., Dang, C. T., Onel, K., Friedman, D. N., Bhatia, S., Strong, L. C., Stovall, M., Kenney, L. B., Barnea, D., Lorenzi, E., Hammond, S., Leisenring, W. M., Robison, L. L., Armstrong, G. T., Diller, L. R. & Oeffinger, K. C. Breast cancer risk in childhood cancer survivors without a history of chest radiotherapy: A report from the childhood cancer survivor study. *J. Clin. Oncol.* **34**, 910–918 (2016).
12. Borgmann, A., Zinn, C., Hartmann, R., Herold, R., Kaatsch, P., Escherich, G., Möricke, A., Henze, G. & von Stackelberg, A. Secondary malignant neoplasms after intensive treatment of relapsed acute lymphoblastic leukaemia in childhood. *Eur. J. Cancer* **44**, 257–268 (2008).
13. Kim, S. H., Shin, K. H., Seok, S. O., Cho, Y. J., Noh, J. K., Suh, J. S. & Yang, W. I. Secondary Malignant Neoplasms after Osteosarcoma: Early Onset and Cumulative Alkylating Agent Dose Dependency. *Ann. Surg. Oncol.* **22**, 859–865 (2015).
14. Pole, J. D., Gu, L. Y., Kirsh, V., Greenberg, M. L. & Nathan, P. C. Subsequent malignant neoplasms in a population- based cohort of pediatric cancer patients: A focus on the first 5 years. *Cancer Epidemiol. Biomarkers Prev.* **24**, 1585–1592 (2015).
15. Zhang, J., Walsh, M. F., Wu, G., Edmonson, M. N., Gruber, T. A., Easton, J., Hedges, D., Ma, X., Zhou, X., Yergeau, D. A., Wilkinson, M. R., Vadodaria, B., Chen, X., McGee, R. B., Hines-Dowell, S., Nuccio, R., Quinn, E., Shurtleff, S. A., Rusch, M., Patel, A., Becksfort, J. B., Wang, S., Weaver, M. S., Ding, L., Mardis, E. R., Wilson, R. K., Gajjar, A., Ellison, D. W., Pappo, A. S., Pui, C.-H., Nichols, K. E. & Downing, J. R. Germline Mutations in Predisposition Genes in Pediatric Cancer. *N. Engl. J. Med.* **373**, 2336–2346 (2015).
16. Wang, Z., Wilson, C. L., Easton, J., Thrasher, A., Mulder, H., Liu, Q., Hedges, D. J., Wang, S., Rusch, M. C., Edmonson, M. N., Levy, S., Lanctot, J. Q., Caron, E., Shelton, K., Currie, K., Lear, M., Patel, A., Rosencrance, C., Shao, Y., Vadodaria, B., Yergeau, D., Sapkota, Y., Brooke, R. J., Moon, W., Rampersaud, E., Ma, X., Chang, T. C., Rice, S. V., Pepper, C., Zhou, X., Chen, X., Chen, W., Jones, A., Boone, B., Ehrhardt, M. J., Krasin, M. J., Howell, R. M., Phillips, N. S., Lewis, C., Srivastava, D., Pui, C. H., Kesserwan, C. A., Wu, G., Nichols, K. E., Downing, J. R., Hudson, M. M., Yasui, Y., Robison, L. L. & Zhang, J. Genetic risk for subsequent neoplasms among long-term survivors of childhood cancer. in *Journal of Clinical Oncology* vol. 36 2078–2087 (American Society of Clinical Oncology, 2018).
17. Moskowitz, C. S., Ronckers, C. M., Chou, J. F., Smith, S. A., Friedman, D. N., Barnea, D., Kok, J. L., Vries, S. de, Wolden, S. L., Henderson, T. O., Pal, H. J. H. van der, Kremer, L. C. M., Neglia, J. P., Turcotte, L. M., Howell, R. M., Arnold, M. A., Schaapveld, M., Aleman, B., Janus, C., Versluys, B., Leisenring, W., Sklar, C. A., Begg, C. B., Pike, M. C., Armstrong, G. T., Robison, L. L., Leeuwen, F. E. van & Oeffinger, K. C. Development and Validation of a Breast Cancer Risk Prediction Model for Childhood Cancer Survivors Treated With Chest Radiation: A Report From the Childhood Cancer Survivor Study and the Dutch Hodgkin Late Effects and LATER Cohorts. *https://doi.org/10.1200/JCO.20.02244* JCO.20.02244 (2021) doi:10.1200/JCO.20.02244.
18. Veiga, L. H. S., Holmberg, E., Anderson, H., Pottern, L., Sadetzki, S., Adams, M. J., Sakata, R., Schneider, A. B., Inskip, P., Bhatti, P., Johansson, R., Neta, G., Shore, R., de Vathaire, F., Damber, L., Kleinerman, R., Hawkins, M. M., Tucker, M., Lundell, M. & Lubin, J. H. Thyroid Cancer after Childhood Exposure to External Radiation: An Updated Pooled Analysis of 12 Studies. *Radiat. Res.* **185**, 473–484 (2016).
19. Taylor, A. J., Little, M. P., Winter, D. L., Sugden, E., Ellison, D. W., Stiller, C. A., Stovall, M., Frobisher, C., Lancashire, E. R., Reulen, R. C. & Hawkins, M. M. Population-Based Risks of CNS Tumors in Survivors of Childhood Cancer: The British Childhood Cancer Survivor Study. *https://doi.org/10.1200/JCO.2009.27.0090* **28**, 5287–5293 (2010).
20. Guilcher, G. M. T., Rivard, L., Huang, J. T., Wright, N. A. M., Anderson, L., Eissa, H., Pelletier, W., Ramachandran, S., Schechter, T., Shah, A. J., Wong, K. & Chow, E. J. Immune function in childhood cancer survivors: a Children's Oncology Group review. *Lancet Child Adolesc. Heal.* **5**, 284–294 (2021).
21. Alexandrov, L. B., Kim, J., Haradhvala, N. J., Huang, M. N., Tian Ng, A. W., Wu, Y., Boot, A., Covington, K. R., Gordenin, D. A., Bergstrom, E. N., Islam, S. M. A., Lopez-Bigas, N., Klimczak, L. J., McPherson,

J. R., Morganella, S., Sabarinathan, R., Wheeler, D. A., Mustonen, V., Alexandrov, L. B., Bergstrom, E. N., Boot, A., Boutros, P., Chan, K., Covington, K. R., Fujimoto, A., Getz, G., Gordenin, D. A., Haradhvala, N. J., Huang, M. N., Islam, S. M. A., Kazanov, M., Kim, J., Klimczak, L. J., Lopez-Bigas, N., Lawrence, M., Martincorena, I., McPherson, J. R., Morganella, S., Mustonen, V., Nakagawa, H., Tian Ng, A. W., Polak, P., Prokopec, S., Roberts, S. A., Rozen, S. G., Sabarinathan, R., Saini, N., Shibata, T., Shiraishi, Y., Stratton, M. R., Teh, B. T., Vázquez-García, I., Wheeler, D. A., Wu, Y., Yousif, F., Yu, W., Getz, G., Rozen, S. G. & Stratton, M. R. The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).

22. Li, B., Brady, S. W., Ma, X., Shen, S., Zhang, Y., Li, Y., Szlachta, K., Dong, L., Liu, Y., Yang, F., Wang, N., Flasch, D. A., Myers, M. A., Mulder, H. L., Ding, L., Liu, Y., Tian, L., Hagiwara, K., Xu, K., Zhou, X., Sioson, E., Wang, T., Yang, L., Zhao, J., Zhang, H., Shao, Y., Sun, H., Sun, L., Cai, J., Sun, H. Y., Lin, T. N., Du, L., Li, H., Rusch, M., Edmonson, M. N., Easton, J., Zhu, X., Zhang, J., Cheng, C., Raphael, B. J., Tang, J., Downing, J. R., Alexandrov, L. B., Zhou, B. B. S., Pui, C. H., Yang, J. J. & Zhang, J. Therapy-induced mutations drive the genomic landscape of relapsed acute lymphoblastic leukemia. *Blood* **135**, 41–55 (2020).
23. Yang, F., Brady, S. W., Tang, C., Sun, H., Du, L., Barz, M. J., Ma, X., Chen, Y., Fang, H., Li, X., Kolekar, P., Pathak, O., Cai, J., Ding, L., Wang, T., von Stackelberg, A., Shen, S., Eckert, C., Klco, J. M., Chen, H., Duan, C., Liu, Y., Li, H., Li, B., Kirschner-Schwabe, R., Zhang, J. & Zhou, B.-B. S. Chemotherapy and mismatch repair deficiency cooperate to fuel TP53 mutagenesis and ALL relapse. *Nat. Cancer 2021* 1–16 (2021) doi:10.1038/s43018-021-00230-8.
24. Brady, S. W., Ma, X., Bahrami, A., Satas, G., Wu, G., Newman, S., Rusch, M., Putnam, D. K., Mulder, H. L., Yergeau, D., Edmonson, M. N., Easton, J., Alexandrov, L. B., Chen, X., Mardis, E. R., Wilson, R. K., Downing, J. R., Pappo, A. S., Raphael, B. J., Dyer, M. A. & Zhang, J. The Clonal Evolution of Metastatic Osteosarcoma as Shaped by Cisplatin Treatment. *Mol. Cancer Res.* molcanres.0620.2018 (2019) doi:10.1158/1541-7786.MCR-18-0620.
25. Schwartz, J. R., Ma, J., Kamens, J., Westover, T., Walsh, M. P., Brady, S. W., Robert Michael, J., Chen, X., Montefiori, L., Song, G., Wu, G., Wu, H., Branstetter, C., Hiltenbrand, R., Walsh, M. F., Nichols, K. E., Maciaszek, J. L., Liu, Y., Kumar, P., Easton, J., Newman, S., Rubnitz, J. E., Mullighan, C. G., Pounds, S., Zhang, J., Gruber, T., Ma, X. & Klco, J. M. The acquisition of molecular drivers in pediatric therapy-related myeloid neoplasms. *Nat. Commun.* **12**, 985 (2021).
26. McLeod, C., Gout, A. M., Zhou, X., Thrasher, A., Rahbarinia, D., Brady, S. W., Macias, M., Birch, K., Finkelstein, D., Sunny, J., Mudunuri, R., Orr, B. A., Treadway, M., Davidson, B., Ard, T. K., Chiao, A., Swistak, A., Wiggins, S., Foy, S., Wang, J., Sioson, E., Wang, S., Michael, J. R., Liu, Y., Ma, X., Patel, A., Edmonson, M. N., Wilkinson, M. R., Frantz, A. M., Chang, T.-C., Tian, L., Lei, S., Islam, S. M. A., Meyer, C., Thangaraj, N., Tater, P., Kandali, V., Ma, S., Nguyen, T., Serang, O. A., McGuire, I., Robison, N., Gentry, D., Tang, X., Palmer, L. E., Wu, G., Suh, E., Tanner, L., McMurry, J., Lear, M., Pappo, A. S., Wang, Z., Wilson, C. L., Cheng, Y., Meshinchi, S., Alexandrov, L. B., Weiss, M. J., Armstrong, G. T., Robison, L. L., Yasui, Y., Nichols, K. E., Ellison, D. W., Bangur, C., Mullighan, C. G., Baker, S. J., Dyer, M. A., Miller, G., Newman, S., Rusch, M., Daly, R., Perry, K., Downing, J. R. & Zhang, J. St. Jude Cloud-a Pediatric Cancer Genomic Data Sharing Ecosystem. *Cancer Discov.* CD-20-1230 (2021) doi:10.1158/2159-8290.cd-20-1230.
27. Newman, S., Nakitandwe, J., Kesserwan, C. A., Azzato, E. M., Wheeler, D. A., Rusch, M., Shurtleff, S., Hedges, D. J., Hamilton, K. V, Foy, S. G., Edmonson, M. N., Thrasher, A., Bahrami, A., Orr, B. A., Klco, J. M., Gu, J., Harrison, L. W., Wang, L., Clay, M. R., Ouma, A., Silkov, A., Liu, Y., Zhang, Z., Liu, Y., Brady, S. W., Zhou, X., Chang, T.-C., Pande, M., Davis, E., Becksfort, J., Patel, A., Wilkinson, M. R., Rahbarinia, D., Kubal, M., Maciaszek, J. L., Pastor, V., Knight, J., Gout, A. M., Wang, J., Gu, Z., Mullighan, C. G., McGee, R. B., Quinn, E. A., Nuccio, R., Mostafavi, R., Gerhardt, E. L., Taylor, L. M., Valdez, J. M., Hines-Dowell, S. J., Pappo, A. S., Robinson, G., Johnson, L.-M., Pui, C.-H., Ellison, D. W., Downing, J. R., Zhang, J. & Nichols, K. E. Genomes for Kids: The scope of pathogenic mutations in pediatric cancer revealed by comprehensive DNA and RNA sequencing. *Cancer Discov.* candisc.1631.2020 (2021) doi:10.1158/2159-8290.CD-20-1631.
28. Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorena, I., Alexandrov, L. B., Martin, S., Wedge, D. C., Van Loo, P., Ju, Y. S., Smid, M., Brinkman, A. B., Morganella, S., Aure, M. R., Lingjærde, O. C., Langerød, A., Ringnér, M., Ahn, S. M., Boyault, S., Brock, J. E., Broeks, A., Butler, A., Desmedt, C., Dirix, L., Dronov, S., Fatima, A., Foekens, J. A., Gerstung, M., Hooijer, G. K. J., Jang, S. J., Jones, D. R., Kim, H. Y., King, T. A., Krishnamurthy, S., Lee, H. J., Lee, J. Y., Li, Y.,

McLaren, S., Menzies, A., Mustonen, V., O'Meara, S., Pauporté, I., Pivot, X., Purdie, C. A., Raine, K., Ramakrishnan, K., Rodríguez-González, F. G., Romieu, G., Sieuwerts, A. M., Simpson, P. T., Shepherd, R., Stebbings, L., Stefansson, O. A., Teague, J., Tommasi, S., Treilleux, I., Van Den Eynden, G. G., Vermeulen, P., Vincent-Salomon, A., Yates, L., Caldas, C., Veer, L. V. t., Tutt, A., Knappskog, S., Tan, B. K. T., Jonkers, J., Borg, Å., Ueno, N. T., Sotiriou, C., Viari, A., Futreal, P. A., Campbell, P. J., Span, P. N., Van Laere, S., Lakhani, S. R., Eyfjord, J. E., Thompson, A. M., Birney, E., Stunnenberg, H. G., Van De Vijver, M. J., Martens, J. W. M., Børresen-Dale, A. L., Richardson, A. L., Kong, G., Thomas, G. & Stratton, M. R. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).

29. Yates, L. R., Knappskog, S., Wedge, D., Farmery, J. H. R., Gonzalez, S., Martincorena, I., Alexandrov, L. B., Van Loo, P., Haugland, H. K., Lilleng, P. K., Gundem, G., Gerstung, M., Pappaemmanuil, E., Gazinska, P., Bhosle, S. G., Jones, D., Raine, K., Mudie, L., Latimer, C., Sawyer, E., Desmedt, C., Sotiriou, C., Stratton, M. R., Sieuwerts, A. M., Lynch, A. G., Martens, J. W., Richardson, A. L., Tutt, A., Lønning, P. E. & Campbell, P. J. Genomic Evolution of Breast Cancer Metastasis and Relapse. *Cancer Cell* **32**, 169-184.e7 (2017).
30. Bhagwate, A. V., Liu, Y., Winham, S. J., McDonough, S. J., Stallings-Mann, M. L., Heinzen, E. P., Davila, J. I., Vierkant, R. A., Hoskin, T. L., Frost, M., Carter, J. M., Radisky, D. C., Cunningham, J. M., Degnim, A. C. & Wang, C. Bioinformatics and DNA-extraction strategies to reliably detect genetic variants from FFPE breast tissue samples. *BMC Genomics* **20**, 689 (2019).
31. Wong, S. Q., Li, J., Tan, A. Y. C., Vedururu, R., Pang, J. M. B., Do, H., Ellul, J., Doig, K., Bell, A., Macarthur, G. A., Fox, S. B., Thomas, D. M., Fellowes, A., Parisot, J. P. & Dobrovic, A. Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively parallel sequencing. *BMC Med. Genomics* **7**, 1–10 (2014).
32. Ikegami, M., Kohsaka, S., Hirose, T., Ueno, T., Kanomata, N., Yamauchi, H., Kobayashi, H. & Tanaka, S. *MicroSEC: Sequence error filtering pipeline for formalin-fixed and paraffin-embedded 2 samples 3 4*. https://www.researchsquare.com (2021) doi:10.21203/RS.3.RS-153650/V1.
33. Agrawal, N., Akbani, R., Aksoy, B. A., Ally, A., Arachchi, H., Asa, S. L., Auman, J. T., Balasundaram, M., Balu, S., Baylin, S. B., Behera, M., Bernard, B., Beroukhim, R., Bishop, J. A., Black, A. D., Bodenheimer, T., Boice, L., Bootwalla, M. S., Bowen, J., Bowlby, R., Bristow, C. A., Brookens, R., Brooks, D., Bryant, R., Buda, E., Butterfield, Y. S. N., Carling, T., Carlsen, R., Carter, S. L., Carty, S. E., Chan, T. A., Chen, A. Y., Cherniack, A. D., Cheung, D., Chin, L., Cho, J., Chu, A., Chuah, E., Cibulskis, K., Ciriello, G., Clarke, A., Clayman, G. L., Cope, L., Copland, J. A., Covington, K., Danilova, L., Davidsen, T., Demchok, J. A., DiCara, D., Dhalla, N., Dhir, R., Dookran, S. S., Dresdner, G., Eldridge, J., Eley, G., El-Naggar, A. K., Eng, S., Fagin, J. A., Fennell, T., Ferris, R. L., Fisher, S., Frazer, S., Frick, J., Gabriel, S. B., Ganly, I., Gao, J., Garraway, L. A., Gastier-Foster, J. M., Getz, G., Gehlenborg, N., Ghossein, R., Gibbs, R. A., Giordano, T. J., Gomez-Hernandez, K., Grimsby, J., Gross, B., Guin, R., Hadjipanayis, A., Harper, H. A., Hayes, D. N., Heiman, D. I., Herman, J. G., Hoadley, K. A., Hofree, M., Holt, R. A., Hoyle, A. P., Huang, F. W., Huang, M., Hutter, C. M., Ideker, T., Iype, L., Jacobsen, A., Jefferys, S. R., Jones, C. D., Jones, S. J. M., Kasaian, K., Kebebew, E., Khuri, F. R., Kim, J., Kramer, R., Kreisberg, R., Kucherlapati, R., Kwiatkowski, D. J., Ladanyi, M., Lai, P. H., Laird, P. W., Lander, E., Lawrence, M. S., Lee, D., Lee, E., Lee, S., Lee, W., Leraas, K. M., Lichtenberg, T. M., Lichtenstein, L., Lin, P., Ling, S., Liu, J., Liu, W., Liu, Y., LiVolsi, V. A., Lu, Y., Ma, Y., Mahadeshwar, H. S., Marra, M. A., Mayo, M., McFadden, D. G., Meng, S., Meyerson, M., Mieczkowski, P. A., Miller, M., Mills, G., Moore, R. A., Mose, L. E., Mungall, A. J., Murray, B. A., Nikiforov, Y. E., Noble, M. S., Ojesina, A. I., Owonikoko, T. K., Ozenberger, B. A., Pantazi, A., Parfenov, M., Park, P. J., Parker, J. S., Paull, E. O., Pedamallu, C. S., Perou, C. M., Prins, J. F., Protopopov, A., Ramalingam, S. S., Ramirez, N. C., Ramirez, R., Raphael, B. J., Rathmell, W. K., Ren, X., Reynolds, S. M., Rheinbay, E., Ringel, M. D., Rivera, M., Roach, J., Robertson, A. G., Rosenberg, M. W., Rosenthal, M., Sadeghi, S., Saksena, G., Sander, C., Santoso, N., Schein, J. E., Schultz, N., Schumacher, S. E., Seethala, R. R., Seidman, J., Senbabaoglu, Y., Seth, S., Sharpe, S., Shaw, K. R. M., Shen, J. P., Shen, R., Sherman, S., Sheth, M., Shi, Y., Shmulevich, I., Sica, G. L., Simons, J. V., Sinha, R., Sipahimalani, P., Smallridge, R. C., Sofia, H. J., Soloway, M. G., Song, X., Sougnez, C., Stewart, C., Stojanov, P., Stuart, J. M., Sumer, S. O., Sun, Y., Tabak, B., Tam, A., Tan, D., Tang, J., Tarnuzzer, R., Taylor, B. S., Thiessen, N., Thorne, L., Thorsson, V., Tuttle, R. M., Umbricht, C. B., Van Den Berg, D. J., Vandin, F., Veluvolu, U., Verhaak, R. G. W., Vinco, M., Voet, D., Walter, V., Wang, Z., Waring, S., Weinberger, P. M., Weinhold, N., Weinstein, J. N., Weisenberger, D. J., Wheeler, D., Wilkerson, M. D., Wilson, J., Williams, M., Winer, D. A., Wise, L., Wu, J., Xi, L., Xu, A. W., Yang, L.,

Yang, L., Zack, T. I., Zeiger, M. A., Zeng, D., Zenklusen, J. C., Zhao, N., Zhang, H., Zhang, J., Zhang, J., Zhang, W., Zmuda, E. & Zou, L. Integrated Genomic Characterization of Papillary Thyroid Carcinoma. *Cell* **159**, 676–690 (2014).

34. Gonin-Laurent, N., Gibaud, A., Huygue, M., Lefèvre, S. H., Le Bras, M., Chauveinc, L., Sastre-Garau, X., Doz, F., Lumbroso, L., Chevillard, S. & Malfoy, B. Specific TP53 mutation pattern in radiation-induced sarcomas. *Carcinogenesis* **27**, 1266–1272 (2006).
35. Joachim, T., Ram, Z., Rappaport, Z. H., Simon, M., Schramm, J., Wiestler, O. D. & Von Deimling, A. Comparative analysis of the NF2, TP53, PTEN, KRAS, NRAS and HRAS genes in sporadic and radiation-induced human meningiomas. *Int. J. Cancer* **94**, 218–221 (2001).
36. Deng, M.Y., Sturm, D., Pfaff, E., Sill, M., Stichel, D., Balasubramanian, G.P., Tippelt, S., Kramm, C., Donson, A.M., Green, A.L., Jones, C., Schittenhelm, J., Ebinger, M., Schuhmann, M.U., Jones, B.C., van Tilburg, C.M., Wittmann, A., Golanov, A., Ryzhova, M., Ecker, J., Milde, T., Witt, O., Sahm, F., Reuss, D., Sumerauer, D., Zamecnik, J., Korshunov, A., von Deimling, A., Pfister, S.M., Jones, D.T.W. Radiation-induced gliomas represent H3-/IDH-wild type pediatric gliomas with recurrent PDGFRA amplification and loss of CDKN2A/B. *Nat. Commun.* **12**, 5530 (2021).
37. DeSisto, J., Lucas, J. T., Xu, K., Donson, A., Lin, T., Sanford, B., Wu, G., Tran, Q. T., Hedges, D., Hsu, C.-Y., Armstrong, G. T., Arnold, M., Bhatia, S., Flannery, P., Lemma, R., Hardie, L., Schüller, U., Venkataraman, S., Hoffman, L. M., Dorris, K., Mulcahy Levy, J. M., Hankinson, T. C., Handler, M., Liu, A. K., Foreman, N., Vibhakar, R., Jones, K., Allen, S., Zhang, J., Baker, S. J., Merchant, T. E., Orr, B. A. & Green, A. L. Comprehensive molecular characterization of pediatric radiation-induced high-grade glioma. *Nat. Commun. 2021 121* **12**, 1–16 (2021).
38. Agnihotri, S., Suppiah, S., Tonge, P. D., Jalali, S., Danesh, A., Bruce, J. P., Mamatjan, Y., Klironomos, G., Gonen, L., Au, K., Mansouri, S., Karimi, S., Sahm, F., von Deimling, A., Taylor, M. D., Laperriere, N. J., Pugh, T. J., Aldape, K. D. & Zadeh, G. Therapeutic radiation for childhood cancer drives structural aberrations of NF2 in meningiomas. *Nat. Commun. 2017 81* **8**, 1–7 (2017).
39. Behjati, S., Gundem, G., Wedge, D. C., Roberts, N. D., Tarpey, P. S., Cooke, S. L., Van Loo, P., Alexandrov, L. B., Ramakrishna, M., Davies, H., Nik-Zainal, S., Hardy, C., Latimer, C., Raine, K. M., Stebbings, L., Menzies, A., Jones, D., Shepherd, R., Butler, A. P., Teague, J. W., Jorgensen, M., Khatri, B., Pillay, N., Shlien, A., Futreal, P. A., Badie, C., McDermott, U., Bova, G. S., Richardson, A. L., Flanagan, A. M., Stratton, M. R., Campbell, P. J., Cooper, C. S., Eeles, R. A., Easton, D., Foster, C., Neal, D. E., Brewer, D. S., Hamdy, F., Lu, Y. J., Lynch, A. G., Massi, C. E., Ng, A., Whitaker, H. C., Yu, Y., Zhang, H., Bancroft, E., Berney, D., Camacho, N., Corbishley, C., Dadaev, T., Dennis, N., Dudderidge, T., Edwards, S., Fisher, C., Ghori, J., Gnanapragasam, V. J., Greenman, C., Hawkins, S., Hazell, S., Howat, W., Karaszi, K., Kay, J., Kote-Jarai, Z., Kremeyer, B., Kumar, P., Lambert, A., Leongamornlert, D., Livni, N., Luxton, H., Matthews, L., Mayer, E., Merson, S., Nicol, D., Ogden, C., O'Meara, S., Pelvender, G., Shah, N. C., Tavare, S., Thomas, S., Thompson, A., Verrill, C., Warren, A. & Zamora, J. Mutational signatures of ionizing radiation in second malignancies. *Nat. Commun.* **7**, 1–8 (2016).
40. Prager, B. C., Vasudevan, H. N., Dixit, D., Bernatchez, J. A., Wu, Q., Wallace, L. C., Bhargava, S., Lee, D., King, B. H., Morton, A. R., Gimple, R. C., Pekmezci, M., Zhu, Z., Siqueira-Neto, J. L., Wang, X., Xie, Q., Chen, C., Barnett, G. H., Vogelbaum, M. A., Mack, S. C., Chavez, L., Perry, A., Raleigh, D. R. & Rich, J. N. The meningioma enhancer landscape delineates novel subgroups and drives druggable dependencies. *Cancer Discov.* **10**, 1722–1741 (2020).
41. Gerstung, M., Jolly, C., Leshchiner, I., Dentro, S. C., Gonzalez, S., Rosebrock, D., Mitchell, T. J., Rubanova, Y., Anur, P., Yu, K., Tarabichi, M., Deshwar, A., Wintersinger, J., Kleinheinz, K., Vázquez-García, I., Haase, K., Jerman, L., Sengupta, S., Macintyre, G., Malikic, S., Donmez, N., Livitz, D. G., Cmero, M., Demeulemeester, J., Schumacher, S., Fan, Y., Yao, X., Lee, J., Schlesner, M., Boutros, P. C., Bowtell, D. D., Zhu, H., Getz, G., Imielinski, M., Beroukhim, R., Sahinalp, S. C., Ji, Y., Peifer, M., Markowetz, F., Mustonen, V., Yuan, K., Wang, W., Morris, Q. D., Spellman, P. T., Wedge, D. C. & Van Loo, P. The evolutionary history of 2,658 cancers. *Nature* **578**, 122–128 (2020).
42. Brady, S. W., Liu, Y., Ma, X., Gout, A. M., Hagiwara, K., Zhou, X., Wang, J., Macias, M., Chen, X., Easton, J., Mulder, H. L., Rusch, M., Wang, L., Nakitandwe, J., Lei, S., Davis, E. M., Naranjo, A., Cheng, C., Maris, J. M., Downing, J. R., Cheung, N. K. V., Hogarty, M. D., Dyer, M. A. & Zhang, J. Pan-neuroblastoma analysis reveals age- and signature-associated driver alterations. *Nat. Commun.* **11**, 1–13 (2020).
43. Howell, R. M., Smith, S. A., Weathers, R. E., Kry, S. F. & Stovall, M. Adaptations to a Generalized

Radiation Dose Reconstruction Methodology for Use in Epidemiologic Studies: An Update from the MD Anderson Late Effect Group. *Radiat. Res.* **192**, 169–188 (2019).

44. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–60 (2009).
45. Edmonson, M. N., Zhang, J., Yan, C., Finney, R. P., Meerzaman, D. M. & Buetow, K. H. Bambino: a variant detector and alignment viewer for next-generation sequencing data in the SAM/BAM format. *Bioinformatics* **27**, 865–866 (2011).
46. Cibulskis, K., Lawrence, M. S., Carter, S. L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E. S. & Getz, G. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–9 (2013).
47. Saunders, C. T., Wong, W. S. W., Swamy, S., Becq, J., Murray, L. J. & Cheetham, R. K. Strelka: accurate somatic small-variant calling from sequenced tumor–normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).
48. Chen, X., Gupta, P., Wang, J., Nakitandwe, J., Roberts, K., Dalton, J. D., Parker, M., Patel, S., Holmfeldt, L., Payne, D., Easton, J., Ma, J., Rusch, M., Wu, G., Patel, A., Baker, S. J., Dyer, M. A., Shurtleff, S., Espy, S., Pounds, S., Downing, J. R., Ellison, D. W., Mullighan, C. G. & Zhang, J. CONSERTING: integrating copy-number analysis with structural-variation detection. *Nat. Methods* **12**, 527–530 (2015).
49. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLOS Comput. Biol.* **12**, e1004873 (2016).
50. Wang, J., Mullighan, C. G., Easton, J., Roberts, S., Heatley, S. L., Ma, J., Rusch, M. C., Chen, K., Harris, C. C., Ding, L., Holmfeldt, L., Payne-Turner, D., Fan, X., Wei, L., Zhao, D., Obenauer, J. C., Naeve, C., Mardis, E. R., Wilson, R. K., Downing, J. R. & Zhang, J. CREST maps somatic structural variation in cancer genomes with base-pair resolution. *Nat. Methods* **8**, 652–654 (2011).
51. Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A. J., Kruglyak, S. & Saunders, C. T. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).
52. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. & Gingeras, T. R. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
53. Hagiwara, K., Edmonson, M. N., Wheeler, D. A. & Zhang, J. indelPost: harmonizing ambiguities in simple and complex indel alignments. *Bioinformatics* (2021) doi:10.1093/BIOINFORMATICS/BTAB601.
54. Edmonson, M. N., Patel, A. N., Hedges, D. J., Wang, Z., Rampersaud, E., Kesserwan, C. A., Zhou, X., Liu, Y., Newman, S., Rusch, M. C., McLeod, C. L., Wilkinson, M. R., Rice, S. V., Soussi, T., Taylor, J. P., Benatar, M., Becksfort, J. B., Nichols, K. E., Robison, L. L., Downing, J. R. & Zhang, J. Pediatric Cancer Variant Pathogenicity Information Exchange (PeCanPIE): a cloud-based platform for curating and classifying germline variants. *Genome Res.* **29**, 1555–1565 (2019).
55. Tian, L., Li, Y., Edmonson, M. N., Zhou, X., Newman, S., McLeod, C., Thrasher, A., Liu, Y., Tang, B., Rusch, M. C., Easton, J., Ma, J., Davis, E., Trull, A., Michael, J. R., Szlachta, K., Mullighan, C., Baker, S. J., Downing, J. R., Ellison, D. W. & Zhang, J. CICERO: a versatile method for detecting complex and diverse driver fusions using cancer RNA sequencing data. *Genome Biol.* **21**, 126 (2020).
56. Uhrig, S., Ellermann, J., Walther, T., Burkhardt, P., Fröhlich, M., Hutter, B., Toprak, U. H., Neumann, O., Stenzinger, A., Scholl, C., Fröhling, S. & Brors, B. Accurate and efficient detection of gene fusions from RNA sequencing data. *Genome Res.* **31**, gr.257246.119 (2021).
57. Wu, G., Diaz, A. K., Paugh, B. S., Rankin, S. L., Ju, B., Li, Y., Zhu, X., Qu, C., Chen, X., Zhang, J., Easton, J., Edmonson, M., Ma, X., Lu, C., Nagahawatte, P., Hedlund, E., Rusch, M., Pounds, S., Lin, T., Onar-Thomas, A., Huether, R., Kriwacki, R., Parker, M., Gupta, P., Becksfort, J., Wei, L., Mulder, H. L., Boggs, K., Vadodaria, B., Yergeau, D., Russell, J. C., Ochoa, K., Fulton, R. S., Fulton, L. L., Jones, C., Boop, F. A., Broniscer, A., Wetmore, C., Gajjar, A., Ding, L., Mardis, E. R., Wilson, R. K., Taylor, M. R., Downing, J. R., Ellison, D. W., Zhang, J. & Baker, S. J. The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma. *Nat. Genet.* **46**, 444 (2014).
58. Chen, X., Bahrami, A., Pappo, A., Easton, J., Dalton, J., Hedlund, E., Ellison, D., Shurtleff, S., Wu, G., Wei, L., Parker, M., Rusch, M., Nagahawatte, P., Wu, J., Mao, S., Boggs, K., Mulder, H., Yergeau, D., Lu, C., Ding, L., Edmonson, M., Qu, C., Wang, J., Li, Y., Navid, F., Daw, N. C., Mardis, E. R., Wilson, R. K., Downing, J. R., Zhang, J. & Dyer, M. A. Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. *Cell Rep.* **7**, 104–112 (2014).
59. Ma, X., Shao, Y., Tian, L., Flasch, D. A., Mulder, H. L., Edmonson, M. N., Liu, Y., Chen, X., Newman, S.,

Nakitandwe, J., Li, Y., Li, B., Shen, S., Wang, Z., Shurtleff, S., Robison, L. L., Levy, S., Easton, J. & Zhang, J. Analysis of error profiles in deep next-generation sequencing data. *Genome Biol.* **20**, 50 (2019).
60. Ma, X., Liu, Y., Liu, Y., Alexandrov, L. B., Edmonson, M. N., Gawad, C., Zhou, X., Li, Y., Rusch, M. C., Easton, J., Huether, R., Gonzalez-Pena, V., Wilkinson, M. R., Hermida, L. C., Davis, S., Sioson, E., Pounds, S., Cao, X., Ries, R. E., Wang, Z., Chen, X., Dong, L., Diskin, S. J., Smith, M. A., Guidry Auvil, J. M., Meltzer, P. S., Lau, C. C., Perlman, E. J., Maris, J. M., Meshinchi, S., Hunger, S. P., Gerhard, D. S. & Zhang, J. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature* **555**, 371–376 (2018).
61. Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V, Cibulskis, K., Sivachenko, A., Carter, S. L., Stewart, C., Mermel, C. H., Roberts, S. A., Kiezun, A., Hammerman, P. S., McKenna, A., Drier, Y., Zou, L., Ramos, A. H., Pugh, T. J., Stransky, N., Helman, E., Kim, J., Sougnez, C., Ambrogio, L., Nickerson, E., Shefler, E., Cortés, M. L., Auclair, D., Saksena, G., Voet, D., Noble, M., DiCara, D., Lin, P., Lichtenstein, L., Heiman, D. I., Fennell, T., Imielinski, M., Hernandez, B., Hodis, E., Baca, S., Dulak, A. M., Lohr, J., Landau, D.-A., Wu, C. J., Melendez-Zajgla, J., Hidalgo-Miranda, A., Koren, A., McCarroll, S. A., Mora, J., Crompton, B., Onofrio, R., Parkin, M., Winckler, W., Ardlie, K., Gabriel, S. B., Roberts, C. W. M., Biegel, J. A., Stegmaier, K., Bass, A. J., Garraway, L. A., Meyerson, M., Golub, T. R., Gordenin, D. A., Sunyaev, S., Lander, E. S., Getz, G. & Getz, G. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
62. Pounds, S., Cheng, C., Li, S., Liu, Z., Zhang, J. & Mullighan, C. A genomic random interval model for statistical analysis of genomic lesion data. *Bioinformatics* **29**, 2088–2095 (2013).
63. Mermel, C. H., Schumacher, S. E., Hill, B., Meyerson, M. L., Beroukhim, R. & Getz, G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).
64. Russnes, H. G., Lingjærde, O. C., Børresen-Dale, A. L. & Caldas, C. Breast Cancer Molecular Stratification: From Intrinsic Subtypes to Integrative Clusters. *Am. J. Pathol.* **187**, 2152–2162 (2017).
65. Forbes, S. A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S., Kok, C. Y., Jia, M., De, T., Teague, J. W., Stratton, M. R., McDermott, U. & Campbell, P. J. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* **43**, D805–D811 (2015).
66. Patch, A.-M., Christie, E. L., Etemadmoghadam, D., Garsed, D. W., George, J., Fereday, S., Nones, K., Cowin, P., Alsop, K., Bailey, P. J., Kassahn, K. S., Newell, F., Quinn, M. C. J., Kazakoff, S., Quek, K., Wilhelm-Benartzi, C., Curry, E., Leong, H. S., Hamilton, A., Mileshkin, L., Au-Yeung, G., Kennedy, C., Hung, J., Chiew, Y.-E., Harnett, P., Friedlander, M., Quinn, M., Pyman, J., Cordner, S., O'Brien, P., Leditschke, J., Young, G., Strachan, K., Waring, P., Azar, W., Mitchell, C., Traficante, N., Hendley, J., Thorne, H., Shackleton, M., Miller, D. K., Arnau, G. M., Tothill, R. W., Holloway, T. P., Semple, T., Harliwong, I., Nourse, C., Nourbakhsh, E., Manning, S., Idrisoglu, S., Bruxner, T. J. C., Christ, A. N., Poudel, B., Holmes, O., Anderson, M., Leonard, C., Lonie, A., Hall, N., Wood, S., Taylor, D. F., Xu, Q., Fink, J. L., Waddell, N., Drapkin, R., Stronach, E., Gabra, H., Brown, R., Jewell, A., Nagaraj, S. H., Markham, E., Wilson, P. J., Ellul, J., McNally, O., Doyle, M. A., Vedururu, R., Stewart, C., Lengyel, E., Pearson, J. V, Waddell, N., deFazio, A., Grimmond, S. M. & Bowtell, D. D. L. Whole-genome characterization of chemoresistant ovarian cancer. *Nature* **521**, 489–94 (2015).
67. Robbe, P., Popitsch, N., Knight, S. J. L., Antoniou, P., Becq, J., He, M., Kanapin, A., Samsonova, A., Vavoulis, D. V, Ross, M. T., Kingsbury, Z., Cabes, M., Ramos, S. D. C., Page, S., Dreau, H., Ridout, K., Jones, L. J., Tuff-Lacey, A., Henderson, S., Mason, J., Buffa, F. M., Verrill, C., Maldonado-Perez, D., Roxanis, I., Collantes, E., Browning, L., Dhar, S., Damato, S., Davies, S., Caulfield, M., Bentley, D. R., Taylor, J. C., Turnbull, C. & Schuh, A. Clinical whole-genome sequencing from routine formalin-fixed, paraffin-embedded specimens: pilot study for the 100,000 Genomes Project. *Genet. Med. 2018 2010* **20**, 1196–1205 (2018).
68. Kim, S., Park, C., Ji, Y., Kim, D. G., Bae, H., van Vrancken, M., Kim, D. H. & Kim, K. M. Deamination Effects in Formalin-Fixed, Paraffin-Embedded Tissue Samples in the Era of Precision Medicine. *J. Mol. Diagnostics* **19**, 137–146 (2017).
69. Tarabichi, M., Salcedo, A., Deshwar, A. G., Ni Leathlobhair, M., Wintersinger, J., Wedge, D. C., Van Loo, P., Morris, Q. D. & Boutros, P. C. A practical guide to cancer subclonal reconstruction from DNA sequencing. *Nat. Methods* **18**, 144–155 (2021).
70. Alexandrov, L. B., Jones, P. H., Wedge, D. C., Sale, J. E., Campbell, P. J., Nik-Zainal, S. & Stratton, M.

R. Clock-like mutational processes in human somatic cells. *Nat. Genet.* **47**, 1402–1407 (2015).

71. Ma, X., Edmonson, M., Yergeau, D., Muzny, D. M., Hampton, O. A., Rusch, M., Song, G., Easton, J., Harvey, R. C., Wheeler, D. A., Ma, J., Doddapaneni, H., Vadodaria, B., Wu, G., Nagahawatte, P., Carroll, W. L., Chen, I.-M., Gastier-Foster, J. M., Relling, M. V., Smith, M. A., Devidas, M., Auvil, J. M. G., Downing, J. R., Loh, M. L., Willman, C. L., Gerhard, D. S., Mullighan, C. G., Hunger, S. P. & Zhang, J. Rise and fall of subclones from diagnosis to relapse in pediatric B-acute lymphoblastic leukaemia. *Nat. Commun.* **6**, 6604 (2015).
72. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* vol. 61 759–767 (1990).
73. Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., Ou Yang, T. H., Porta-Pardo, E., Gao, G. F., Plaisier, C. L., Eddy, J. A., Ziv, E., Culhane, A. C., Paull, E. O., Sivakumar, I. K. A., Gentles, A. J., Malhotra, R., Farshidfar, F., Colaprico, A., Parker, J. S., Mose, L. E., Vo, N. S., Liu, J., Liu, Y., Rader, J., Dhankani, V., Reynolds, S. M., Bowlby, R., Califano, A., Cherniack, A. D., Anastassiou, D., Bedognetti, D., Rao, A., Chen, K., Krasnitz, A., Hu, H., Malta, T. M., Noushmehr, H., Pedamallu, C. S., Bullman, S., Ojesina, A. I., Lamb, A., Zhou, W., Shen, H., Choueiri, T. K., Weinstein, J. N., Guinney, J., Saltz, J., Holt, R., Rabkin, C. E., Caesar-Johnson, S. J., Demchok, J. A., Felau, I., Kasapi, M., Ferguson, M. L., Hutter, C. M., Sofia, H. J., Tarnuzzer, R., Wang, Z., Yang, L., Zenklusen, J. C., Zhang, J. (Julia), Chudamani, S., Liu, J., Lolla, L., Naresh, R., Pihl, T., Sun, Q., Wan, Y., Wu, Y., Cho, J., DeFreitas, T., Frazer, S., Gehlenborg, N., Getz, G., Heiman, D. I., Kim, J., Lawrence, M. S., Lin, P., Meier, S., Noble, M. S., Saksena, G., Voet, D., Zhang, H., Bernard, B., Chambwe, N., Dhankani, V., Knijnenburg, T., Kramer, R., Leinonen, K., Liu, Y., Miller, M., Reynolds, S., Shmulevich, I., Thorsson, V., Zhang, W., Akbani, R., Broom, B. M., Hegde, A. M., Ju, Z., Kanchi, R. S., Korkut, A., Li, J., Liang, H., Ling, S., Liu, W., Lu, Y., Mills, G. B., Ng, K. S., Rao, A., Ryan, M., Wang, J., Weinstein, J. N., Zhang, J., Abeshouse, A., Armenia, J., Chakravarty, D., Chatila, W. K., de Bruijn, I., Gao, J., Gross, B. E., Heins, Z. J., Kundra, R., La, K., Ladanyi, M., Luna, A., Nissan, M. G., Ochoa, A., Phillips, S. M., Reznik, E., Sanchez-Vega, F., Sander, C., Schultz, N., Sheridan, R., Sumer, S. O., Sun, Y., Taylor, B. S., Wang, J., Zhang, H., Anur, P., Peto, M., Spellman, P., Benz, C., Stuart, J. M., Wong, C. K., Yau, C., Hayes, D. N., Parker, J. S., Wilkerson, M. D., Ally, A., Balasundaram, M., Bowlby, R., Brooks, D., Carlsen, R., Chuah, E., Dhalla, N., Jones, S. J. M., Kasaian, K., Lee, D., Ma, Y., Marra, M. A., Mayo, M., Moore, R. A., Mungall, A. J., Mungall, K., Robertson, A. G., Sadeghi, S., Schein, J. E., Sipahimalani, P., Tam, A., Thiessen, N., Tse, K., Wong, T., Berger, A. C., Beroukhim, R., Cherniack, A. D., Cibulskis, C., Gabriel, S. B., Ha, G., Meyerson, M., Schumacher, S. E., Shih, J., Kucherlapati, M. H., Kucherlapati, R. S., Baylin, S., Cope, L., Danilova, L., Bootwalla, M. S., Lai, P. H., Maglinte, D. T., Van Den Berg, D. J., Weisenberger, D. J., Auman, J. T., Balu, S., Bodenheimer, T., Fan, C., Hoadley, K. A., Hoyle, A. P., Jefferys, S. R., Jones, C. D., Meng, S., Mieczkowski, P. A., Mose, L. E., Perou, A. H., Perou, C. M., Roach, J., Shi, Y., Simons, J. V., Skelly, T., Soloway, M. G., Tan, D., Veluvolu, U., Fan, H., Hinoue, T., Laird, P. W., Shen, H., Zhou, W., Bellair, M., Chang, K., Covington, K., Creighton, C. J., Dinh, H., Doddapaneni, H. V., Donehower, L. A., Drummond, J., Gibbs, R. A., Glenn, R., Hale, W., Han, Y., Hu, J., Korchina, V., Lee, S., Lewis, L., Li, W., Liu, X., Morgan, M., Morton, D., Muzny, D., Santibanez, J., Sheth, M., Shinbrot, E., Wang, L., Wang, M., Wheeler, D. A., Xi, L., Zhao, F., Hess, J., Appelbaum, E. L., Bailey, M., Cordes, M. G., Ding, L., Fronick, C. C., Fulton, L. A., Fulton, R. S., Kandoth, C., Mardis, E. R., McLellan, M. D., Miller, C. A., Schmidt, H. K., Wilson, R. K., Crain, D., Curley, E., Gardner, J., Lau, K., Mallery, D., Morris, S., Paulauskis, J., Penny, R., Shelton, C., Shelton, T., Sherman, M., Thompson, E., Yena, P., Bowen, J., Gastier-Foster, J. M., Gerken, M., Leraas, K. M., Lichtenberg, T. M., Ramirez, N. C., Wise, L., Zmuda, E., Corcoran, N., Costello, T., Hovens, C., Carvalho, A. L., de Carvalho, A. C., Fregnani, J. H., Longatto-Filho, A., Reis, R. M., Scapulatempo-Neto, C., Silveira, H. C. S., Vidal, D. O., Burnette, A., Eschbacher, J., Hermes, B., Noss, A., Singh, R., Anderson, M. L., Castro, P. D., Ittmann, M., Huntsman, D., Kohl, B., Le, X., Thorp, R., Andry, C., Duffy, E. R., Lyadov, V., Paklina, O., Setdikova, G., Shabunin, A., Tavobilov, M., McPherson, C., Warnick, R., Berkowitz, R., Cramer, D., Feltmate, C., Horowitz, N., Kibel, A., Muto, M., Raut, C. P., Malykh, A., Barnholtz-Sloan, J. S., Barrett, W., Devine, K., Fulop, J., Ostrom, Q. T., Shimmel, K., Wolinsky, Y., Sloan, A. E., De Rose, A., Giuliante, F., Goodman, M., Karlan, B. Y., Hagedorn, C. H., Eckman, J., Harr, J., Myers, J., Tucker, K., Zach, L. A., Deyarmin, B., Hu, H., Kvecher, L., Larson, C., Mural, R. J., Somiari, S., Vicha, A., Zelinka, T., Bennett, J., Iacocca, M., Rabeno, B., Swanson, P., Latour, M., Lacombe, L., Têtu, B., Bergeron, A., McGraw, M., Staugaitis, S. M., Chabot, J., Hibshoosh, H., Sepulveda, A., Su, T., Wang, T., Potapova, O., Voronina, O., Desjardins, L., Mariani, O., Roman-Roman, S., Sastre, X., Stern, M. H., Cheng, F., Signoretti, S., Berchuck, A., Bigner, D., Lipp, E.,

Marks, J., McCall, S., McLendon, R., Secord, A., Sharp, A., Behera, M., Brat, D. J., Chen, A., Delman, K., Force, S., Khuri, F., Magliocca, K., Maithel, S., Olson, J. J., Owonikoko, T., Pickens, A., Ramalingam, S., Shin, D. M., Sica, G., Van Meir, E. G., Zhang, H., Eijckenboom, W., Gillis, A., Korpershoek, E., Looijenga, L., Oosterhuis, W., Stoop, H., van Kessel, K. E., Zwarthoff, E. C., Calatozzolo, C., Cuppini, L., Cuzzubbo, S., DiMeco, F., Finocchiaro, G., Mattei, L., Perin, A., Pollo, B., Chen, C., Houck, J., Lohavanichbutr, P., Hartmann, A., Stoehr, C., Stoehr, R., Taubert, H., Wach, S., Wullich, B., Kycler, W., Murawa, D., Wiznerowicz, M., Chung, K., Edenfield, W. J., Martin, J., Baudin, E., Bubley, G., Bueno, R., De Rienzo, A., Richards, W. G., Kalkanis, S., Mikkelsen, T., Noushmehr, H., Scarpace, L., Girard, N., Aymerich, M., Campo, E., Giné, E., Guillermo, A. L., Van Bang, N., Hanh, P. T., Phu, B. D., Tang, Y., Colman, H., Evason, K., Dottino, P. R., Martignetti, J. A., Gabra, H., Juhl, H., Akeredolu, T., Stepa, S., Hoon, D., Ahn, K., Kang, K. J., Beuschlein, F., Breggia, A., Birrer, M., Bell, D., Borad, M., Bryce, A. H., Castle, E., Chandan, V., Cheville, J., Copland, J. A., Farnell, M., Flotte, T., Giama, N., Ho, T., Kendrick, M., Kocher, J. P., Kopp, K., Moser, C., Nagorney, D., O'Brien, D., O'Neill, B. P., Patel, T., Petersen, G., Que, F., Rivera, M., Roberts, L., Smallridge, R., Smyrk, T., Stanton, M., Thompson, R. H., Torbenson, M., Yang, J. D., Zhang, L., Brimo, F., Ajani, J. A., Gonzalez, A. M. A., Behrens, C., Bondaruk, J., Broaddus, R., Czerniak, B., Esmaeli, B., Fujimoto, J., Gershenwald, J., Guo, C., Lazar, A. J., Logothetis, C., Meric-Bernstam, F., Moran, C., Ramondetta, L., Rice, D., Sood, A., Tamboli, P., Thompson, T., Troncoso, P., Tsao, A., Wistuba, I., Carter, C., Haydu, L., Hersey, P., Jakrot, V., Kakavand, H., Kefford, R., Lee, K., Long, G., Mann, G., Quinn, M., Saw, R., Scolyer, R., Shannon, K., Spillane, A., Stretch, onathan, Synott, M., Thompson, J., Wilmott, J., Al-Ahmadie, H., Chan, T. A., Ghossein, R., Gopalan, A., Levine, D. A., Reuter, V., Singer, S., Singh, B., Tien, N. V., Broudy, T., Mirsaidi, C., Nair, P., Drwiega, P., Miller, J., Smith, J., Zaren, H., Park, J. W., Hung, N. P., Kebebew, E., Linehan, W. M., Metwalli, A. R., Pacak, K., Pinto, P. A., Schiffman, M., Schmidt, L. S., Vocke, C. D., Wentzensen, N., Worrell, R., Yang, H., Moncrieff, M., Goparaju, C., Melamed, J., Pass, H., Botnariuc, N., Caraman, I., Cernat, M., Chemencedji, I., Clipca, A., Doruc, S., Gorincioi, G., Mura, S., Pirtac, M., Stancul, I., Tcaciuc, D., Albert, M., Alexopoulou, I., Arnaout, A., Bartlett, J., Engel, J., Gilbert, S., Parfitt, J., Sekhon, H., Thomas, G., Rassl, D. M., Rintoul, R. C., Bifulco, C., Tamakawa, R., Urba, W., Hayward, N., Timmers, H., Antenucci, A., Facciolo, F., Grazi, G., Marino, M., Merola, R., de Krijger, R., Gimenez-Roqueplo, A. P., Piché, A., Chevalier, S., McKercher, G., Birsoy, K., Barnett, G., Brewer, C., Farver, C., Naska, T., Pennell, N. A., Raymond, D., Schilero, C., Smolenski, K., Williams, F., Morrison, C., Borgia, J. A., Liptay, M. J., Pool, M., Seder, C. W., Junker, K., Omberg, L., Dinkin, M., Manikhas, G., Alvaro, D., Bragazzi, M. C., Cardinale, V., Carpino, G., Gaudio, E., Chesla, D., Cottingham, S., Dubina, M., Moiseenko, F., Dhanasekaran, R., Becker, K. F., Janssen, K. P., Slotta-Huspenina, J., Abdel-Rahman, M. H., Aziz, D., Bell, S., Cebulla, C. M., Davis, A., Duell, R., Elder, J. B., Hilty, J., Kumar, B., Lang, J., Lehman, N. L., Mandt, R., Nguyen, P., Pilarski, R., Rai, K., Schoenfield, L., Senecal, K., Wakely, P., Hansen, P., Lechan, R., Powers, J., Tischler, A., Grizzle, W. E., Sexton, K. C., Kastl, A., Henderson, J., Porten, S., Waldmann, J., Fassnacht, M., Asa, S. L., Schadendorf, D., Couce, M., Graefen, M., Huland, H., Sauter, G., Schlomm, T., Simon, R., Tennstedt, P., Olabode, O., Nelson, M., Bathe, O., Carroll, P. R., Chan, J. M., Disaia, P., Glenn, P., Kelley, R. K., Landen, C. N., Phillips, J., Prados, M., Simko, J., Smith-McCune, K., VandenBerg, S., Roggin, K., Fehrenbach, A., Kendler, A., Sifri, S., Steele, R., Jimeno, A., Carey, F., Forgie, I., Mannelli, M., Carney, M., Hernandez, B., Campos, B., Herold-Mende, C., Jungk, C., Unterberg, A., von Deimling, A., Bossler, A., Galbraith, J., Jacobus, L., Knudson, M., Knutson, T., Ma, D., Milhem, M., Sigmund, R., Godwin, A. K., Madan, R., Rosenthal, H. G., Adebamowo, C., Adebamowo, S. N., Boussioutas, A., Beer, D., Giordano, T., Mes-Masson, A. M., Saad, F., Bocklage, T., Landrum, L., Mannel, R., Moore, K., Moxley, K., Postier, R., Walker, J., Zuna, R., Feldman, M., Valdivieso, F., Dhir, R., Luketich, J., Pinero, E. M. M., Quintero-Aguilo, M., Carlotti, C. G., Dos Santos, J. S., Kemp, R., Sankarankuty, A., Tirapelli, D., Catto, J., Agnew, K., Swisher, E., Creaney, J., Robinson, B., Shelley, C. S., Godwin, E. M., Kendall, S., Shipman, C., Bradford, C., Carey, T., Haddad, A., Moyer, J., Peterson, L., Prince, M., Rozek, L., Wolf, G., Bowman, R., Fong, K. M., Yang, I., Korst, R., Rathmell, W. K., Fantacone-Campbell, J. L., Hooke, J. A., Kovatich, A. J., Shriver, C. D., DiPersio, J., Drake, B., Govindan, R., Heath, S., Ley, T., Van Tine, B., Westervelt, P., Rubin, M. A., Lee, J. Il, Aredes, N. D., Mariamidze, A., Serody, J. S., Demicco, E. G., Disis, M. L., Vincent, B. G. & Shmulevich, Ilya. The Immune Landscape of Cancer. *Immunity* **48**, 812-830.e14 (2018).

74. Brady, S. W., McQuerry, J. A., Qiao, Y., Piccolo, S. R., Shrestha, G., Jenkins, D. F., Layer, R. M., Pedersen, B. S., Miller, R. H., Esch, A., Selitsky, S. R., Parker, J. S., Anderson, L. A., Dalley, B. K., Factor, R. E., Reddy, C. B., Boltax, J. P., Li, D. Y., Moos, P. J., Gray, J. W., Heiser, L. M., Buys, S. S.,

Cohen, A. L., Johnson, W. E., Quinlan, A. R., Marth, G., Werner, T. L. & Bild, A. H. Combating subclonal evolution of resistant cancer phenotypes. *Nat. Commun.* **8**, 1231 (2017).
75. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–9 (2015).
76. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–3 (2012).
77. Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P. & Mesirov, J. P. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–40 (2011).
78. Barbie, D. A., Tamayo, P., Boehm, J. S., Kim, S. Y., Moody, S. E., Dunn, I. F., Schinzel, A. C., Sandy, P., Meylan, E., Scholl, C., Fröhling, S., Chan, E. M., Sos, M. L., Michel, K., Mermel, C., Silver, S. J., Weir, B. A., Reiling, J. H., Sheng, Q., Gupta, P. B., Wadlow, R. C., Le, H., Hoersch, S., Wittner, B. S., Ramaswamy, S., Livingston, D. M., Sabatini, D. M., Meyerson, M., Thomas, R. K., Lander, E. S., Mesirov, J. P., Root, D. E., Gilliland, D. G., Jacks, T. & Hahn, W. C. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–12 (2009).
79. Pennock, N. D., Jindal, S., Horton, W., Sun, D., Narasimhan, J., Carbone, L., Fei, S. S., Searles, R., Harrington, C. A., Burchard, J., Weinmann, S., Schedin, P. & Xia, Z. RNA-seq from archival FFPE breast cancer samples: molecular pathway fidelity and novel discovery. *BMC Med. Genomics 2019 121* **12**, 1–18 (2019).
80. Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., Hoang, C. D., Diehn, M. & Alizadeh, A. A. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods 2015 125* **12**, 453–457 (2015).