Supplementary data

Transcriptomics and transposon mutagenesis identify multiple mechanisms of resistance to the FGFR inhibitor AZD4547

Sjors M. Kas, Julian R. de Ruiter, Koen Schipper, Eva Schut, Lorenzo Bombardelli, Ellen Wientjens, Anne Paulien Drenth, Renske de Korte-Grimmerink, Sunny Mahakena, Christopher Phillips, Paul D. Smith, Sjoerd Klarenbeek, Koen van de Wetering, Anton Berns, Lodewyk F.A. Wessels, Jos Jonkers

Supplementary experimental details

Cell viability assay

WESB cells were seeded in triplicate with 500 cells per well in 96-well plates. After 24 hours, the medium was refreshed with either DMSO (as a control) or AZD4547 with the indicated concentrations. Three days later, cell viability was assayed in an Envision plate reader (Perkin Elmer) using resazurin (cell titer blue; Promega).

Clonogenic assay

MEFs were trypsinized and 2000 cells were seeded in 6-well plates. After 24 hours, the medium was refreshed with either DMSO (as a control) or AZD4547 with the indicated concentrations. Eleven days later, the cells were fixed with 4% formalin in PBS and stained with 0.1% crystal violet in demineralized water. Quantification was performed by dissolving the crystal violet with 10% acetic acid in demineralized water and determining the absorbance at 590 nm. The experiment was performed three times.

Competition assay

WESB-*Fgfr2* cells were transfected with pX330.pgkpur constructs containing three independent sgRNAs targeting *Rasa1* (sg*Rasa1*) or a non-targeting as control (sgNT) using Lipofectamine 2000 (ThermoFisher Scientific 11668027) according to manufacturer's protocol. Transfected cells were selected using puromycin (4 µg/ml) for 48 hours. The pX330.pgkpur construct is a modified version of the pX330 backbone (1), which contains a puromycin resistance ORF under the hPGK promoter (2). The pX330-U6-Chimeric_BB-CBh-hSpCas9 construct was a gift from Dr Feng Zhang (Addgene plasmid #42230). WESB-*Fgfr2* cells containing sgNT or sg*Rasa1* were mixed and seeded at a one-to-one ratio on 6-well plates in medium supplemented with 5% FBS and either DMSO or 2 µM AZD4547. The medium was refreshed every 4 days and DNA was isolated at days 0, 7, 10 and 12 using the Gentra Puregene genomic DNA isolation kit (Qiagen). PCR amplifications of *Rasa1* exon 2, 6 and 8 were performed with specific primers spanning the target sites and 100-200 ng DNA template, using the Q5 High-Fidelity PCR kit (NEB M0492). Amplification PCR reactions were diluted 20 times with Milli-Q and subsequently Sanger sequenced using the FW primers. CRISPR/Cas9-induced editing efficacy was quantified using the TIDE algorithm (3). Cells with only sgNT were used as a negative control in all genomic DNA amplifications and only TIDE outputs with R^2>0.9 were considered.

sgRNA sequence *Rasa1*-1: 5'-TTATAAGAGAGAGTGATCGG-3'
sgRNA sequence *Rasa1*-2: 5'-CGAGAAGAAGATCCACACGA -3'
sgRNA sequence *Rasa1*-3: 5'-ATCTCCAGGAGTATTATCTG-3'

Rasa1 sgRNA1 PCR FW 5'-TTGTGTTCTCACAGACCTGAAT-3' (557 bp)
Rasa1 sgRNA1 PCR RV 5'-TCAATCTGTGATCTCCAAGCC-3' (557 bp)
Rasa1 sgRNA2 PCR FW 5'-TGTAGGCAAGAGAGCCAAATTA-3' (697 bp)
Rasa1 sgRNA2 PCR RV 5'-GTTCAAGGCCAGTCTGATCTAC-3' (697 bp)
Rasa1 sgRNA3 PCR FW 5'-GAGTTCTTTCAGAGAGCGAAGG-3' (406 bp)
Rasa1 sgRNA3 PCR RV 5'-GAGTTCTTTCAGAGAGCGAAGG-3' (406 bp)

Preparation of membrane vesicles and vesicular transport assays

Membrane vesicles from Sf9 cells were obtained after infection with a control or a human ABCG2-containing baculovirus at a multiplicity of infection of 1 (4). After incubation at 27°C for 3 days, cells were harvested by centrifugation at 500 x *g* for 5 min. Cells were then resuspended in ice-cold hypotonic buffer (0.5 mM sodium phosphate and 0.1 mM EDTA, pH 7.4) supplemented with a protease inhibitor cocktail (Roche, Basel, Switzerland) and incubated at 4°C for 90 min under constant agitation. Thereafter, the cell lysate was homogenized using a tight-fitting Dounce homogenizer. Next, cell debris and nuclei were removed by slow-speed centrifugation at 500 x *g* (4°C for 10 min). The supernatant was collected and centrifuged at 4°C at 100,000 x *g* for 40 min. The membrane pellet was resuspended in TS buffer (50 mM Tris-HCl and 250 mM sucrose, pH 7.4) and passed through a 27-gauge needle 25 times. The vesicles were dispensed in aliquots, snap-frozen in liquid nitrogen, and stored at -80°C until use. Vesicular transport assays were performed using the rapid filtration method as previously described (4,5). Briefly, ABCG2 or control Sf9 membrane vesicles containing 20 µg of protein were incubated with 1 µM [^{3}H]-MTX in 50 µl of TS buffer in the presence of 4 mM ATP or AMP, 10 mM $MgCl_2$, 10 mM creatine phosphate, and 100 g of creatine kinase/ml. After 10 minutes, 40 µl of the reaction mixture was diluted in 200 µl of ice-cold TS buffer and immediately filtrated using a MultiScreen$_{HTS}$ vacuum manifold in combination with Multiscreen$_{HTS}$ FB 96-well filter plates (Millipore, Bedford, MA). Membranes were washed four times with 200 µl of ice-cold phosphate-buffered saline and the radioactivity retained on the membranes was counted by liquid scintillation counting.

Immunohistochemistry

Tissues were formalin-fixed and paraffin-embedded (FFPE) by routine procedures. Immunohistochemical stainings of MET, ABCG2 (BCRP) and cleaved Caspase-3 were processed as previously described (6,7). The following primary antibodies were used for the respective proteins: MET (1:100, R&D Systems AF527), BCRP (1:400, Abcam 24115) and cleaved Caspase-3 (1:400, CST 9661). Citrate buffer was used as antigen retrieval for MET and BCRP. TRIS/EDTA pH 9.0 was used for cleaved Caspase-3. Immunohistochemical staining of IGF1R was performed on a Discovery Ultra autostainer (Ventana Medical Systems). Briefly, paraffin sections were cut at 3 µm, heated at 75°C for 28 minutes and deparaffinised in the instrument with EZ prep solution (Ventana Medical Systems). Heat-induced antigen retrieval was carried out using Cell Conditioning 1 (CC1, Ventana

Medical Systems) for 64 minutes at 95^0C. IGF1R was detected using clone G11 (Ready-to-use, 16 minutes at 37^0C, Ventana Medical Systems), bound antibody was detected using the OMap anti-Rb HRP (Ventana Medical Systems) for 12 minutes after which the ChromoMap DAB Kit (Ventana Medical Systems) was applied. Slides were counterstained with Hematoxylin and Bluing Reagent (Ventana Medical Systems). All slides were digitally processed using the Aperio ScanScope (Aperio, Vista, CA, USA) and captured using ImageScope software version 12.3.2.8013 (Aperio). Cleaved Caspase-3 and ABCG2 immunohistochemical stainings were reviewed and scored by a veterinary pathologist (Sjoerd Klarenbeek) in a blinded manner. The images on the slides were captured using an Axioskop 40 microscope and an AxioCam MRc5 camera (Zeiss) and analyzed using the ZEN lite 2012 (Blue edition) software. The number of cleaved Caspase-3 positive cells were counted in four independent fields (0.34 mm^2) per tumor and the average number of positive cells per mm^2 was calculated. Necrotic areas in these tumors were excluded from the analysis.

Immunoblotting

Protein lysates were made using lysis buffer (25 mM Tris-HCl pH 7.6, 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS in Milli-Q) complemented with protease and phosphatase inhibitors (Roche) and quantified using the BCA protein assay kit (Pierce). Equal amounts of proteins were separated on a 4-12% Bis-Tris gradient gel (Invitrogen) and transferred overnight onto nitrocellulose membrane (Bio-Rad) in 1x transfer buffer (25 mM Tris, 2 M Glycine, 20% methanol in demineralized water). Membranes were blocked in 5% w/v bovine serum albumin (BSA) in PBS-T (pH 7.5, 0.005% Tween-20 in demineralized water) and incubated overnight with the primary antibodies in 5% w/v BSA in PBS-T. Membranes were washed three times and incubated with the secondary antibodies goat anti-rabbit-HRP (1:2000, Dako P0448), rabbit anti-mouse-HRP (1:5000, Dako P0260), rabbit anti-rat-HRP (1:2000, Invitrogen 61-9520) or donkey anti-mouse IRDye 680nm (1:5000, Li_COR 926-32222) in 5% w/v BSA in PBS-T. Stained membranes were washed three times in PBS-T and then developed using ECL (Pierce 32209), ECL 2 Substrate (Pierce 80196) or captured using the Li-Cor Odyssey Infrared Imaging System and analyzed using Odyssey Application software version 3.0.16. The intensities of the bands were quantified using ImageJ software version 2.0.0-rc-65/1.52b.

RNA sequencing and analysis

Illumina TruSeq mRNA libraries were generated and sequenced with 50-65 base single reads on a HiSeq 2500 using V4 chemistry (Illumina Inc., San Diego) as previously described by Boelens *et al.* (8). The resulting reads were trimmed using Cutadapt (version 1.13) to remove any remaining adapter sequences (9), filtering reads shorter than 20 bp after trimming to ensure good mappability. The trimmed reads were aligned to the GRCm38 reference genome using STAR (version 2.5.3a) (10). QC statistics from Fastqc (version 0.11.5; http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and the above-mentioned tools were collected and summarized using Multiqc (version 1.1) (11). Gene expression counts were generated by featureCounts (version 1.5.2) using gene definitions from

Ensembl GRCm38 version 89 (12). Normalized expression values were obtained by correcting for differences in sequencing depth between samples using DESeqs median-of-ratios approach and then log-transforming the normalized counts (13). Differentially expressed genes were identified using DIDS (version 0.10.1) (14), using a threshold of $p < 0.05$ for statistical significance. Variants in RTKs and genes involved in downstream FGFR signaling were called using Vardict (version 2017.04.18) and annotated using Ensembl VEP (version 90.7) (15,16). The entire analysis pipeline (including the alignment, expression estimation and variant calling) was implemented using Snakemake and is freely available on GitHub [https://github.com/jrderuiter/snakemake-rnaseq] (17).

Amplification of *SB* transposon insertions

Transposon insertions were amplified following a previously described tagmentation-based DNA sequencing protocol (18). Briefly, recombinant Tn5 transposase was prepared as previously described by Picelli *et al.* (19), and diluted in glycerol buffer to a final concentration of 3.7 µM. The Tn5-adapter complex was prepared by incubating 30 minutes at 37°C equimolar amounts of Tn5 and separately annealed adapters pairs Tn5ME-A+ 3'dT5P-oligo and Tn5ME-B+3'dT5P-oligo as previously described by David L. Stern (18). Each tagmentation reaction was prepared by combining 2 µl of genomic DNA (100 ng in total), 4 µl of 5x TAPS-PEG buffer (19), 1 µl of Tn5-adapters complex and 13 µl water and incubated for 10 minutes at 55°C. Tn5 was stripped off from DNA by adding 4 µl of 0.2% SDS and incubating the reaction 5 min at 55°C. The enrichment PCR was performed by combining 3 µl of tagmented DNA, 1 µl of enrichment primer at 1 µM, 6 µl water and 10 µl Phusion Flash 2x mix (cat.# F548L, Thermo Scientific, Waltham, MA USA). PCR1 was performed by combining 5 µl of enrichment PCR reaction, 8 µl water, 1 µl of P5-indexed primer and 1 µl of transposon-specific primer SB-PCR1 and 10 µL Phusion Flash 2x mix. PCR2 was performed by combining 2 µl of PCR1 reaction, 8 µl water, 1 µl of P7-indexed primer + 1 µl FC2 primer and 10 µl Phusion Flash 2x mix. Equal amounts of PCR2 products were pooled and run on an agarose gel. Fragments above 600 basepairs were excised from the gel, purified on Qiagen columns and eluted in water. The pool of tagmented DNA was sequenced with 150 base paired-end reads on a MiSeq 300 using the micro kit v2 reagents (Illumina Inc., San Diego). The following primer sequences and PCR cycler settings were used:

Primer sequences:
SB-enrich: GCTTGTGGAAGGCTACTCGAAATGTTTGACCC
SB_pcr1: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTGTATGTAAACTTCCGACTTCAAC
FC2: AATGATACGGCGACCACCGA
Tn5ME-A-adaptor: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
3'dT5P-oligo: CTGTCTCTTATACACATCTGAC (must be 5’ phosphorylated and 3’OH blocked by an inverted thymindine)
Tn5ME-B-adaptor: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG
A-idx-i5-1: AATGATACGGCGACCACCGAGATCTACACTAATGTGGTCGTCGGCAGCGTC
A-idx-i5-2: AATGATACGGCGACCACCGAGATCTACACGCACTCAGTCGTCGGCAGCGTC

A-idx-i5-3: AATGATACGGCGACCACCGAGATCTACACAACAGCGGTCGTCGGCAGCGTC
A-idx-i5-4: AATGATACGGCGACCACCGAGATCTACACCCATATGATCGTCGGCAGCGTC
A-idx-i5-5: AATGATACGGCGACCACCGAGATCTACACTGGAAAGCTCGTCGGCAGCGTC
A-idx-i5-6: AATGATACGGCGACCACCGAGATCTACACAGCAACGCTCGTCGGCAGCGTC
A-idx-i5-7: AATGATACGGCGACCACCGAGATCTACACCCCTTGCATCGTCGGCAGCGTC
A-idx-i5-8: AATGATACGGCGACCACCGAGATCTACACCCCTCTTGTCGTCGGCAGCGTC
A-idx-i5-9: AATGATACGGCGACCACCGAGATCTACACTTCGAGCCTCGTCGGCAGCGTC
A-idx-i5-10: AATGATACGGCGACCACCGAGATCTACACAGTAGTTATCGTCGGCAGCGTC
A-idx-i5-11: AATGATACGGCGACCACCGAGATCTACACAGAAAGTGTCGTCGGCAGCGTC
A-idx-i5-12: AATGATACGGCGACCACCGAGATCTACACTGCCGGTATCGTCGGCAGCGTC
A-idx-i5-13: AATGATACGGCGACCACCGAGATCTACACGCAAACTGTCGTCGGCAGCGTC
A-idx-i5-14: AATGATACGGCGACCACCGAGATCTACACGGTTGAGATCGTCGGCAGCGTC
A-idx-i5-15: AATGATACGGCGACCACCGAGATCTACACATAGATGTTCGTCGGCAGCGTC
A-idx-i5-16: AATGATACGGCGACCACCGAGATCTACACCAAACATTTCGTCGGCAGCGTC
A-idx-i5-17: AATGATACGGCGACCACCGAGATCTACACCTGAGCGTTCGTCGGCAGCGTC
A-idx-i5-18: AATGATACGGCGACCACCGAGATCTACACTAGTCTCTTCGTCGGCAGCGTC
B-idx-i7-1: CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGG
B-idx-i7-2: CAAGCAGAAGACGGCATACGAGATATCCGCATGTCTCGTGGGCTCGG
B-idx-i7-3: CAAGCAGAAGACGGCATACGAGATATTGAAGTGTCTCGTGGGCTCGG
B-idx-i7-4: CAAGCAGAAGACGGCATACGAGATCTCTGCGTGTCTCGTGGGCTCGG
B-idx-i7-5: CAAGCAGAAGACGGCATACGAGATGATACGCAGTCTCGTGGGCTCGG
B-idx-i7-6: CAAGCAGAAGACGGCATACGAGATTACGTTCGGTCTCGTGGGCTCGG
B-idx-i7-7: CAAGCAGAAGACGGCATACGAGATTGAATCCTGTCTCGTGGGCTCGG
B-idx-i7-8: CAAGCAGAAGACGGCATACGAGATGGCTATAAGTCTCGTGGGCTCGG
B-idx-i7-9: CAAGCAGAAGACGGCATACGAGATCACAACCTGTCTCGTGGGCTCGG

PCR cycler settings:
Enrichment PCR (2-step): 98°C 30s; 45 cycles of (98°C 8s, 72°C 35s)
PCR1: 98°C 30s; 15-18 cycles of (98°C 8s, 63°C 5s, 72°C 30s)
PCR2: 98°C 30s; 15-18 cycles of (98°C 8s, 63°C 5s, 72°C 30s)

Insertion site analysis

Insertion sites were identified using the taqmap pipeline in PyIM (version 0.3.0, https://github.com/jrderuiter/pyim). Briefly, this analysis pipeline first trimmed the TaqMap paired-end reads using Cutadapt (version 1.12) to remove any matepairs not containing the transposon and to remove any Nextera transposase sequences (9). The remaining mate pairs were aligned to the mm10 reference genome using Bowtie2 (version 2.3.0) (20). After the alignment, redundant sequences mapping to the same genomic location and belonging to the same tumor were collapsed into a single insertion. To avoid issues with slight variations in the alignment, insertions from the same sample that

occurred within 10 bp of each other were collapsed into a single insertion. Insertions were assigned to genes using the rule-based mapping approach (21) with the *SB* preset and gene definitions from Ensembl GRCm38 89. Support scores were calculated as the number of unique mate pairs supporting a given insertion. Relative support scores (used as a proxy for clonality) were calculated by normalizing support scores to the highest support score of the corresponding sample.

Genes associated with *de novo* resistance were identified by selecting genes that did not have any insertions in untreated tumors (vehicle-treated tumors and the donor tumor) and had insertions in at least two AZD4547-resistant samples. The *de novo* candidate genes were then ranked by their frequency of occurrence. Genes associated with intrinsic resistance were selected by performing a Welch's *t*-test between the clonality scores of insertions in the AZD4547-resistant tumors and the vehicle treated tumors, as well as determining the difference between the means of both groups (to ensure a minimum effect size). Candidate genes were selected by filtering for genes with a difference in means > 0.1 and a *t*-test *P* value < 0.25, after which the candidates were ranked by their mean differences.

Validation of the endogenous *Fgfr2-Tbc1d1* fusion

The *Fgfr2*-*Tbc1d1* fusion was detected in WESB-*Fgfr2* cells as previously described (22). WESB cells were used as negative control. The following primer sequences were used:

Fgfr2 FW: 5´-TGGCCAGGGATATCAACAAC-3´
Tbc1d1 RV: 5´-CCAGGCTGTGAGAAGGATTT-3´

***Met* qPCR copy number analysis**

DNA was isolated from AZD4547-resistant tumors and wild-type FVB spleen as a control. The qPCR was performed on a Quantstudio 6 flex Real-Time PCR system (Applied Biosystems) using low ROX SYBR green (Bioline) with *Met* and β-catenin (*Ctnnb1*) specific primers.

Met FW: 5'-TCTCGGAGCCACAAACTACA-3'
Met RV: 5'-GCAGTCCCGACAAGGTAAAC-3'
Ctnnb1 FW: 5'-TCAGGGCAGGTGAAACTGTA-3'
Ctnnb1 RV: 5'-GACTCCCAGCACACTGAACTTA-3'

The relative copy number levels of *Met* and *Ctnnb1* were quantified using a five-point standard curve. The *Met* relative copy number was normalized to the *Ctnnb1* relative copy number for each sample and subsequently normalized to the normalized relative abundance of wild-type FVB spleen.

Statistical analysis

The effect of AZD4547 treatment on tumor growth of WESB-*Fgfr2*-EV and WESB-*Fgfr2*-ABCG2 established tumors was tested using mixed linear models. Prior to this analysis, the tumor size measurements were scaled so that each tumor's size at the first time point was equal to 1. Inspection of the individual tumor growth curves suggested an approximately linear increase in tumor size over time. The exact growth rate, i.e. the slope of the growth curve, showed some inter-tumor variability. Therefore, we modeled growth rate using a fixed effect population-level slope β, a random effect tumor-level slope $\mathbf{b}_i$ to account for inter-tumor variability, and a fixed effect term γ for the interaction of time and treatment to model the effect of treatment on tumor growth. Additional inter-tumor variability is allowed by random intercepts $\mathbf{a}_i$, which complement the fixed effect intercept α. This leads to the following model formulation for the size of tumor i as a function of time and treatment:

$$\text{Tumor size}_i = \alpha + \mathbf{a}_i + (\beta + \mathbf{b}_i) \times \text{time} + \gamma \times \text{time} \times \text{treatment}_i$$

The significance of the treatment effect was established using an ANOVA comparing the models with and without the interaction term γ.

Supplementary references

1. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013;339:819–23.

2. Harmsen T, Klaasen S, van de Vrugt H, Te Riele H. DNA mismatch repair and oligonucleotide end-protection promote base-pair substitution distal from a CRISPR/Cas9-induced DNA break. Nucleic Acids Res. 2018;46:2945–55.

3. Brinkman EK, Chen T, Amendola M, van Steensel B. Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Res. 2014;42:e168.

4. van de Wetering K, Sapthu S. ABCG2 functions as a general phytoestrogen sulfate transporter in vivo. FASEB J. 2012;26:4014–24.

5. van de Wetering K, Burkon A, Feddema W, Bot A, de Jonge H, Somoza V, et al. Intestinal breast cancer resistance protein (BCRP)/Bcrp1 and multidrug resistance protein 3 (MRP3)/Mrp3 are involved in the pharmacokinetics of resveratrol. Mol Pharmacol. 2009;75:876–85.

6. Doornebal CW, Klarenbeek S, Braumuller TM, Klijn CN, Ciampricotti M, Hau C-S, et al. A preclinical mouse model of invasive lobular breast cancer metastasis. Cancer Res. AACR; 2013;73:353–63.

7. Henneman L, van Miltenburg MH, Michalak EM, Braumuller TM, Jaspers JE, Drenth AP, et al. Selective resistance to the PARP inhibitor olaparib in a mouse model for BRCA1-deficient metaplastic breast cancer. Proc Natl Acad Sci. National Acad Sciences; 2015;112:8409–14.

8. Boelens MC, Nethe M, Klarenbeek S, de Ruiter JR, Schut E, Bonzanni N, et al. PTEN Loss in E-Cadherin-Deficient Mouse Mammary Epithelial Cells Rescues Apoptosis and Results in Development of Classical Invasive Lobular Carcinoma. Cell Rep. 2016;16:2087–101.

9. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011;17:10.

10. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinforma Oxf Engl. 2013;29:15–21.

11. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinforma Oxf Engl. 2016;32:3047–8.

12. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinforma Oxf Engl. 2014;30:923–30.

13. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11:R106.

14. de Ronde JJ, Rigaill G, Rottenberg S, Rodenhuis S, Wessels LFA. Identifying subgroup markers in heterogeneous populations. Nucleic Acids Res. 2013;41:e200.

15. Lai Z, Markovets A, Ahdesmaki M, Chapman B, Hofmann O, McEwen R, et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. Nucleic Acids Res. 2016;44:e108.

16. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17:122.

17. Köster J, Rahmann S. Snakemake--a scalable bioinformatics workflow engine. Bioinforma Oxf Engl. 2012;28:2520–2.

18. Stern DL. Tagmentation-Based Mapping (TagMap) of Mobile DNA Genomic Insertion Sites. 2017;

19. Picelli S, Björklund ÅK, Reinius B, Sagasser S, Winberg G, Sandberg R. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res. 2014;24:2033–40.

20. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.

21. de Jong J, de Ridder J, van der Weyden L, Sun N, van Uitert M, Berns A, et al. Computational identification of insertional mutagenesis targets for cancer gene discovery. Nucleic Acids Res. 2011;39:e105.

22. de Ruiter JR, Kas SM, Schut E, Adams DJ, Koudijs MJ, Wessels LFA, et al. Identifying transposon insertions and their effects from RNA-sequencing data. Nucleic Acids Res. 2017;45:7064–77.