Into the blue...Using mouse models to uncover genes driving tumorigenesis and therapy resistance in human breast cancer

Ruiter, J.R. de

8

General discussion

8.1 Introduction

In this work, we have made extensive use of several mouse models for studying tumor development and therapy resistance in two different types of breast cancer: invasive lobular carcinoma (ILC) and BRCA1-deficient triple-negative breast cancer (TNBC). This allowed us to identify several known and novel drivers of ILC (Chapters 3 and 4) and BRCA1-deficient TNBC (Chapter 5), which were also shown to be frequently aberrated in human tumors. Besides this, we also demonstrated how mouse models can be used to identify resistance mechanisms in both ILC (Chapter 6) and BRCA-deficient TNBC (Chapter 7), by treating mice with transplanted tumor fragments with targeted treatments and comparing therapy-sensitive and -resistant tumors.

In the following sections, we focus on some of the concepts that underlie the work presented in this thesis. Specifically, we will first elaborate on our approaches used to identify cancer genes and therapy resistance mechanisms and reflect on several advantages and shortcomings of the presented approaches. Next, we touch on several recent advances in sequencing technologies and challenges in increasing the accessibility of sequencing data for future studies. Finally, we reflect on the current role(s) of mouse models in cancer research and how this is impacted by increased human sequencing efforts and the developments of organoid technologies.

8.2 Strengths and weaknesses of insertional mutagenesis strategies

One of the major goals/challenges in cancer research involves identifying which genes are driving tumor development, based on the idea that these genes can be targeted specifically in new therapies. In this work, we have shown how transposon-based insertional mutagenesis (IM) can be used to screen for candidate cancer genes, allowing us to identify driver genes in a mouse model of invasive lobular carcinoma (ILC) (Chapter 3). Moreover, we also have demonstrated how transposon insertions can be identified from RNA-sequencing data, and how this approach provides an important biological filter that focuses on transposon insertions that are actually expressed in the tumors, and therefore most likely to be relevant for tumor formation (4)

One of the main advantages of IM strategies, compared to other mutagenesis approaches, is that they provide a relatively unbiased genome-wide approach to forward genetic screening, whilst allowing efficient retrieval of transposon insertions via targeted sequencing. Moreover, depending on the used transposon, IM-based

screens can be used to simultaneously identify both candidate oncogenes and tumor suppressors in a single screen. As such, IM-based screening approaches have been successfully used to identify cancer driver genes in a wide variety of cancer types[1–7]. In these settings, IM has been particularly effective in identifying cancer drivers that are not readily apparent in human sequencing data, due to complex amplicons or more subtle events such as alternative splicing. A clear example is given by our transposon mutagenesis screen, which identified several ILC driver genes that were located in an amplicon on human chromosome 1q. Due to the size of this amplicon, these genes would have been difficult to identify without our screening results.

An important limitation of IM strategies is that different transposon systems have specific biases[8], which can lead to different results depending on the used transposon. Part of these biases are inherent to the used transposon systems (e.g., *piggyBac*[9,10], *Sleeping Beauty*[11,12]), which have been shown to have integration biases towards specific genetic sequences and/or certain gene features (e.g., integration within or upstream of genes)[13]. Other biases stem from the sequences used to construct the transposon, which may affect the mutagenicity of the transposon in certain types of cells. For example, the *T2/Onc2* transposon used in our screen contains an MSCV promoter, which is likely to drive oncogene expression in hematopoietic cells at a higher rate than in other cell types. For this reason, other transposons such as the *T2/Onc3*[14] transposon have been developed for targeting different cell types by including different promoter sequences.

These biases may explain to some extent why our screen identified a relatively concise set of candidate genes, compared to other recent IM screens[1,4,7]. For instance, SB transposons generally have a lower integration rate than other, more aggressive transposon systems such as *piggyBac*[15]. This might explain why we identify fewer integration sites than studies with other transposons[7], but it does not explain differences with SB-based studies[1,4]. Similarly, the used transposon (*T2Onc/2*) may not be particularly active in mammary gland epithelium, due to limited activity of the MSCV promoter in this tissue. Alternatively, biological constraints might be limiting the number of potential ILC driver genes, as is suggested by the limited number of driver mutations associated with ILC[16]. Finally, sequencing issues (e.g. limited sequencing depth) might have reduced the sensitivity of our insertion detection. However, orthogonal analyses using different sequencing approaches (IM-Fusion[17], Tagmap[18]) identified similar patterns of insertions, indicating that this is unlikely to be the case.

Another limitation of IM strategies is that targeted sequencing approaches aimed at identifying transposon insertions are unlikely to identify other types of mutations generated by competing, spontaneous mutational processes occurring in tumors. As such, naive IM screening approaches may miss additional candidate genes that are,

for example, aberrated by spontaneous point-mutations or copy number changes. The risk of additional mutations depends highly on the used mouse model, with more genomically unstable tumors being more likely to acquire mutations via alternative mechanisms. This issue can be mitigated by either applying IM in genomically stable models, or by combining targeted sequencing strategies with other sequencing modalities such as exome-sequencing[19]. Alternatively, RNA-sequencing-based approaches[17,20] such as IM-Fusion can be used for simultaneous identification of transposon insertions and mutations, as a single RNA-sequencing dataset can be used for insertion detection, variant calling, gene-fusion detection and differential expression analyses.

Altogether, IM strategies are effective approaches for identifying candidate cancer driver genes and potential mechanisms of therapy resistance. Moreover, when combined with somatic engineering approaches, IM provides an especially powerful method for identifying candidate cancer genes in specific genetic backgrounds and rapidly validating these candidates in the same *in vivo* context. Besides this, deeper sequencing and single-cell sequencing approaches[21] may provide more insight into the heterogeneity of IM-induced tumors and allow researchers to study clonal evolution within these tumors. This type of approach – optionally combined with multiplexed *in vivo* CRISPR approaches that permit studying interactions between multiple cancer driver genes[22,23] – may help determine how heterogeneous patient tumors are likely to evolve over time and in response to specific treatments, enabling the development of better treatment strategies.

8.3 Trade-offs in uncovering therapy resistance mechanisms

Acquired therapy resistance is currently a major challenge in the clinic that limits the effectiveness of many chemotherapy drugs and targeted therapies. In this work, we aimed to identify resistance mechanisms to two different targeted therapies, FGFR inhibition (FGFRi) in ILC (Chapter 6) and PARP inhibition (PARPi) in BRCA1-deficient TNBC (Chapter 7). To do so, we transplanted donor tumors into multiple syngeneic mice, which were subjected to treatment. By subsequently contrasting therapy-resistant and -sensitive (i.e. vehicle-treated) tumors, this approach allowed us to identify potential resistance mechanisms to both targeted treatments.

In Chapter 6, we focused on identifying FGFRi resistance mechanisms using insertional mutagenesis, based on the premise that SB-induced tumors can acquire new transposon insertions during treatment and in doing so become resistant to the applied therapy[24,25]. One advantage of using IM for this purpose, is that potential

resistance mechanisms are relatively easy to identify by using targeted sequencing to reveal the locations of new insertions in resistant tumors. However, a limitation of the approach is that the range of identifiable resistance mechanisms may be limited by integration biases of the used transposon and by the types of mutations that can be generated by transposon insertions (e.g. gene overexpression and gene truncation, but not point mutations). Moreover, competing mutagenic processes may result in additional mutations that may contribute to resistance but cannot be detected by targeted insertion sequencing, as evidenced by the FGFR2 mutations that we identified in RNA-sequencing data from several FGFRi-resistant ILC tumors.

In Chapter 7, we used genomic (copy number variation (CNV)-sequencing) and transcriptomic (RNA-sequencing) approaches to identify potential PARPi resistance mechanisms in BRCA2-deficient mammary tumors from KB2P mice. One of the main challenges of this type of approach is that genomically unstable tumors (such as KB2P tumors) are likely to contain many additional aberrations after treatment, of which only a very small fraction is likely to be causally involved in treatment resistance. As such, identifying potential resistance mechanisms requires additional prioritization of mutations, either using computational algorithms or by integrating data from additional sequencing approaches, other experiments or external databases. In the KB2P model, we addressed this issue by requiring genes to be frequently amplified/deleted and to show a strong effect on expression, providing additional evidence that these mutations have a functional effect. Additional evidence from a separate shRNA screen was used to further validate our results, by providing orthogonal evidence that candidate genes were indeed involved in resistance.

An alternative approach for identifying acquired resistance mechanisms involves correction of copy number profiles of resistant tumors for events that were present before treatment, and subsequent application of algorithms such as RUBIC[26] or GISTIC[27] to identify recurrent copy number events that are acquired after treatment. This would potentially allow us to identify DNA copy number alterations (CNAs) that arose in response to treatment. The main advantage of this approach is that it may restrict the number of potential candidate genes by focusing on the minimal regions of recurrent CNAs. However, this approach requires a snapshot of the pre-treatment copy number state for each tumor to avoid calling events that were already present in the pre-treatment tumor samples. In our case, these samples were not available. Moreover, it may be challenging to obtain representative pre-treatment samples for each donor, due to differences in pre-treatment tumor fragments stemming from intra-tumor heterogeneity. In addition, this approach assumes that the same resistance mechanism occurs in each tumor, and it may therefore not have enough statistical power in a population displaying heterogeneous resistance mechanisms.

In essence, the main difference between our two resistance screens lies in their focus. In the IM-based approach, we first used RNA-sequencing data to identify if any known mechanisms could explain the observed resistance, allowing us to perform a relatively focused analysis of this genome-wide dataset. We then used insertion-site sequencing to identify novel resistance mechanisms in an unbiased fashion. Due to the selective nature of the used transposon, this analysis identified a relatively short list of novel mechanisms, which might not have been identified through genome-wide analyses. In contrast, our analysis of the KB2P tumors was performed in a genome-wide fashion and relied on additional data sources (such as the curated list of DNA damage response (DDR) genes and the secondary shRNA screen) to prioritize candidate genes. In this respect, IM-based approaches have the advantage of being cleaner than a genome-wide analysis, at the risk of missing novel resistance mechanisms that are not captured by the transposon system. On the other hand, the sequencing approach has the advantage of potentially being more comprehensive but runs the risk of missing mechanisms in the analysis stage, due to the added complexity of the analysis, biases introduced by prior knowledge, and/or lack of statistical power resulting from small sample size and heterogeneous resistance mechanisms. As such, the choice of which approach to use likely depends on various aspects of the used mouse model (such as its degree of genomic instability) and pre-existing knowledge of potential resistance mechanisms.

Treatment resistance will most likely remain a substantial challenge for many anti-cancer therapies in the foreseeable future. One particularly challenging aspect of addressing therapy resistance is that it is essentially a race against evolution, in which tumors are likely to develop new resistance mechanisms in response to any applied therapies. As such, effective therapeutic strategies will need to be designed to minimize or prevent the emergence of resistance, for example by targeting convergent signaling pathways[28] or changing treatment strategies (e.g. by employing drug holidays[29] or combination therapies[30]). However, testing different (combination) treatments in different genetic backgrounds is likely to be prohibitively expensive and time consuming, due to the sheer amount of possible combinations involved. As such, we expect that computational models trained on existing experimental data will become crucial in modeling how tumor cells and cell populations respond to specific (combination) treatments, allowing us to predict which (combinations of) anti-cancer drugs are most likely to be most effective in treating a given tumor and preventing resistance.

8.4 Advances in tumor sequencing technologies

For the studies described in this thesis, we have made substantial use of second-generation sequencing technologies, which have greatly improved our ability to perform in-depth genomic and transcriptomic characterization of cell populations. However, one limitation of these technologies, is that they essentially reduce (cancer) genomes into a large collection of very short sequence reads, which must be pieced back together using a predefined reference genome to extract anything meaningful from the data. The fact that these short sequencing reads do not allow *de novo* assembly of cancer genomes inherently limits the power of our analyses. Moreover, this fragmented snapshot is not extracted from individual cells, but from bulk populations of thousands to millions of tumor cells, meaning that any derivation is going to be an average blur of a population of cells, rather than a detailed picture of all individual cells that captures all of their differences.

In this respect, one of the most exciting recent developments is the emergence of single-cell sequencing technologies, which can be used to characterize the genomic or transcriptomic state of individual cells[31,32]. This makes single-cell approaches especially suitable for studying intra-tumor heterogeneity, which is considered to be one of the main drivers of tumor progression, metastasis and escape from therapy[33]. As such, single-cell RNA-sequencing approaches have frequently been used to identify distinct populations of cells within tumors[34–36] and to explore how these populations may be involved in tumor progression and therapy resistance[37–39]. Similarly, single-cell CNV-sequencing and exome-sequencing approaches have been used to track cancer evolution by deriving how distinct tumor subclones evolved during tumor progression[40,41]. Limiting aspects of current single-cell technologies are low sequencing depths and high costs due to the large numbers of cells that need to be sequenced for a comprehensive analysis. However, with dropping sequencing costs and the development of more cost-effective approaches, we expect that single-cell sequencing will become an increasingly important pillar in studying tumor heterogeneity and cancer evolution.

Besides this, third-generation sequencing approaches (e.g., Nanopore, PacBio) are poised to further revolutionize sequencing analyses by providing much longer reads (10,000-100,000 bp) than current (second-generation) technologies, which are generally limited to reads of hundreds of base pairs. Increased read lengths provide substantial benefits for many sequencing analyses, as longer reads are much easier to stitch back together than short sequences, as demonstrated by *de novo* assemblies based on long-read sequencing[42]. Similarly, identification of complex structural variations in patient samples have been shown to benefit greatly from longer read lengths, allowing Nanopore-based approaches to achieve substantial accuracy in

identifying variants, even with relatively low sequencing depths[43]. Furthermore, in RNA-sequencing based approaches, long-read sequencing has been used to identify novel gene transcripts[44] and quantify isoform expression[45] more accurately than current sequencing approaches.

Together, we expect that advances in single-cell and third-generation sequencing will substantially increase our ability to detect more complex mutational events, such as structural variants and alternative splicing, whilst simultaneously providing detailed insight into the heterogeneous tumor environment. However, properly leveraging these technologies will also require the development of new computational approaches for analyzing the produced data, as is evidenced by the widespread development of software packages for the analysis of single-cell and long-read sequencing data.

8.5 Increasing the reproducibility and accessibility of sequencing data

To optimally exploit the full compendium of sequencing data that is currently being generated, it will be important to generate and share sequencing datasets in a way that they can be easily queried, integrated and visualized by fellow researchers. One of the main challenges in this respect is the heterogeneity of many analyses, which is fueled by lack of consensus on best practices for analyzing different types of sequencing data. This has led to the development of many different variants of sequencing analysis pipelines, which generally differ in their choice of software for different steps of the analysis. As a result, these pipelines can produce considerably different results, even when applied to the same datasets[46–48], thus complicating the integration of datasets from different research groups.

To establish a set of best practices, several efforts have been made to address this issue by benchmarking different approaches on gold-standard datasets[46–50]. Ideally, the results of these benchmarks will be used to develop templates for many standard analyses in community projects (e.g. bcbio-nextgen*, Snakemake workflows†) and to facilitate the sharing and democratization of sequencing pipelines. An important part of these projects is the development of open-source workflow management systems (e.g. Snakemake[51], Nextflow[52]), which allow users to define abstract workflows that can be shared and reused to analyze new datasets. Together with tools for managing software installations in virtual environments (e.g., Bioconda[53], Docker‡,

*http://bcbio-nextgen.readthedocs.io/en/latest
†https://bitbucket.org/johanneskoester/snakemake-workflows
‡https://www.docker.com

Singularity[54]), these developments make it easier to develop workflows that can be easily shared with, and applied by others, thus increasing the reproducibility and uniformity of sequencing analyses.

Additionally, with the large amounts of data that are being generated, it is becoming increasingly important to facilitate the deposition of these data into large centralized repositories, where they can be easily be integrated and analyzed by other researchers. Unfortunately, processed datasets are rarely shared in this fashion, as existing repositories (e.g. GEO[55], ENA[56]) are generally restricted to specific data types and therefore provide limited capabilities for integrating and visualizing datasets from different sequencing modalities. Portals such as cBioportal[57] are a great step forward in this respect, as they provide basic functionality for querying and visualizing different datasets without requiring specific computational expertise. In future work, it will be important to expand these existing initiatives to include many more public datasets, ideally upon publication of the corresponding manuscripts. Besides this, we expect that integration with similar portals containing data from non-human organisms (such as mouse models) will play an important role in disseminating insights from these model systems and improve cross-pollination with human sequencing efforts.

8.6 A future for mouse model systems?

Model systems have long been important cornerstones of cancer research, as they provide renewable and manipulatable biological systems that can be used to formulate and test hypotheses. Compared to other systems, mouse models have been particularly useful as they provide a controlled and experimentally tractable *in vivo* setting that remains relatively close to human biology. As a result, many different types of mouse models of human cancer have been developed over the years, of which we provided an overview in Chapter 1. These developments have given rise to a vast collection of mouse models with different genetic backgrounds, mimicking important aspects of many human cancer types[58,59].

An important advantage of model systems such as mouse models, compared to human sequencing projects, is that they can provide unambiguous, causative evidence of whether a given mutation is involved in tumorigenesis or therapy resistance. For example, by introducing a mutation into a mouse model that does by itself not develop tumors, we can unequivocally establish whether the added mutation results in increased tumor formation[60]. Human sequencing projects are much more limited in this respect, as without further validation these projects can generally only provide correlative evidence[61]. As such, model systems play an important role in

the follow-up of driver genes identified in these projects, by allowing us to validate their role(s) in tumor formation and determine how they may affect other tumor characteristics such as therapy response.

The utility of mouse models in cancer research will likely be enhanced by the development of lentiviral and CRISPR-based somatic engineering approaches[62,63], which allow the rapid introduction of mutations into existing baseline mouse models (Chapter 5). As such, these approaches can be used to speed-up the creation of new mouse models containing specific mutations, greatly expanding the capacity to further validate and characterize candidate driver genes. Moreover, by combining vectors containing/targeting different driver genes, somatic engineering can also be used to generate complex mouse models with different combinations of mutations, enabling more detailed dissection of interactions between different driver genes and how these influence tumor development and therapy response.

However, mouse models are not alone and face competition from human-based models. For instance, human tumor-derived organoid systems have recently gained much popularity in cancer research by enabling efficient derivation of three-dimensional (3D) cell culture models that maintain the genomic complexity of tumors more faithfully than conventional cell line models[64,65]. For this reason, organoids have been touted as an effective approach for generating *in vitro* models of patient tumors, which can be sequenced and screened with anti-cancer drugs to identify mutations and correlate these with therapy response[66]. Besides this, organoids can also be manipulated using lentiviral and CRISPR-based somatic technologies, allowing validation of driver genes or mutations that are predicted to influence therapy response. Combined with falling sequencing costs and increasing patient sequencing efforts, these efforts in organoid models are poised to provide a wealth of data correlating human tumor genotypes with drug response. Given enough data, this may substantially improve our ability to build computational models predicting a patient's therapy response, without requiring detailed modelling in (mouse) model systems.

Organoid-based approaches are however limited by the fact that they are grown in a rather artificial *in vitro* environment, and as such do not faithfully reflect the complexity of the tumor microenvironment and/or interactions with an active immune system. As a result, (humanized) mouse models are much better suited for studying the effects of immune-based therapies, which have gained much popularity due to their unparalleled success in treating specific cancer types[67–70]. Besides this, mouse models of *de novo* tumorigenesis provide opportunities for studying all stages of tumor initiation, progression and metastasis, which is challenging to do with models based on end-stage patient tumors. Finally, detailed *in vivo* characterization is likely to remain important to gain insight into the role of specific drivers in different stages of tumorigenesis, which can be used to design and develop novel

therapies targeting driver-specific vulnerabilities. As such, we expect that mouse model systems will continue to play an important role in cancer research. However, to maximize their impact, studies should focus on exploiting the unique properties of mouse models to complement patient- and organoid-based approaches, rather than mimicking studies that are better performed in more human-like (*in vitro*) models.

8.7 References

[1] Roberto Rangel, Song-Choon Lee, Kenneth Hon-Kim Ban, et al. “Transposon mutagenesis identifies genes that cooperate with mutant Pten in breast cancer progression”. In: *Proceedings of the National Academy of Sciences* 113.48 (2016), E7749–E7758 (cit. on p. 287).

[2] Roberto Rangel, Liliana Guzman-Rojas, Takahiro Kodama, et al. “Identification of new tumor suppressor genes in triple-negative breast cancer”. In: *Cancer Research* 77.15 (2017), pp. 4089–4101 (cit. on p. 287).

[3] Sjors M Kas, Julian R de Ruiter, Koen Schipper, et al. “Insertional mutagenesis identifies drivers of a novel oncogenic pathway in invasive lobular breast carcinoma”. In: *Nature Genetics* 49.8 (2017), pp. 1219–1230 (cit. on p. 287).

[4] Michael B Mann, Michael A Black, Devin J Jones, et al. “Transposon mutagenesis identifies genetic drivers of Braf V600E melanoma”. In: *Nature Genetics* 47.5 (2015), p. 486 (cit. on p. 287).

[5] Takahiro Kodama, Justin Y Newberg, Michiko Kodama, et al. “Transposon mutagenesis identifies genes and cellular processes driving epithelial-mesenchymal transition in hepatocellular carcinoma”. In: *Proceedings of the National Academy of Sciences* 113.24 (2016), E3384–E3393 (cit. on p. 287).

[6] Haruna Takeda, Alistair G Rust, Jerrold M Ward, et al. “Sleeping Beauty transposon mutagenesis identifies genes that cooperate with mutant Smad4 in gastric cancer development”. In: *Proceedings of the National Academy of Sciences* 113.14 (2016), E2057–E2065 (cit. on p. 287).

[7] Roland Rad, Lena Rad, Wei Wang, et al. “A conditional piggyBac transposition system for genetic screening in mice identifies oncogenic networks in pancreatic cancer”. In: *Nature Genetics* 47.1 (2015), p. 47 (cit. on p. 287).

[8] Johann de Jong, Waseem Akhtar, Jitendra Badhai, et al. “Chromatin landscapes of retroviral and transposon integration profiles”. In: *PLoS Genetics* 10.4 (2014), e1004250 (cit. on p. 287).

[9] Sheng Ding, Xiaohui Wu, Gang Li, et al. “Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice”. In: *Cell* 122.3 (2005), pp. 473–483 (cit. on p. 287).

[10] Roland Rad, Lena Rad, Wei Wang, et al. “PiggyBac transposon mutagenesis: a tool for cancer gene discovery in mice”. In: *Science* 330.6007 (2010), pp. 1104–1107 (cit. on p. 287).

[11] Adam J Dupuy, Keiko Akagi, David A Largaespada, Neal G Copeland, and Nancy A Jenkins. “Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system”. In: *Nature* 436.7048 (2005), p. 221 (cit. on p. 287).

[12] Lara S Collier, Corey M Carlson, Shruthi Ravimohan, Adam J Dupuy, and David A Largaespada. “Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse”. In: *Nature* 436.7048 (2005), pp. 272–276 (cit. on p. 287).

[13] Johann de Jong, Jeroen de Ridder, Louise van der Weyden, et al. “Computational identification of insertional mutagenesis targets for cancer gene discovery”. In: *Nucleic Acids Research* 39.15 (2011), e105–e105 (cit. on p. 287).

[14] Adam J Dupuy, Laura M Rogers, Jinsil Kim, et al. “A modified sleeping beauty transposon system that can be used to model a wide variety of human cancers in mice”. In: *Cancer Research* 69.20 (2009), pp. 8150–8156 (cit. on p. 287).

[15] Sareina Chiung-Yuan Wu, Yaa-Jyuhn James Meir, Craig J Coates, et al. “piggyBac is a flexible and highly active transposon as compared to sleeping beauty, Tol2, and Mos1 in mammalian cells”. In: *Proceedings of the National Academy of Sciences* 103.41 (2006), pp. 15008–15013 (cit. on p. 287).

[16] Giovanni Ciriello, Michael L Gatza, Andrew H Beck, et al. “Comprehensive molecular portraits of invasive lobular breast cancer”. In: *Cell* 163.2 (2015), pp. 506–519 (cit. on p. 287).

[17] Julian R de Ruiter, Sjors M Kas, Eva Schut, et al. “Identifying transposon insertions and their effects from RNA-sequencing data”. In: *Nucleic Acids Research* 45.12 (2017), pp. 7064–7077 (cit. on pp. 287, 288).

[18] David L Stern. “Tagmentation-based mapping (TagMap) of mobile DNA genomic insertion sites”. In: *BioRxiv* (2017), p. 037762 (cit. on p. 287).

[19] Louise van der Weyden, George Giotopoulos, Kim Wong, et al. “Somatic drivers of B-ALL in a model of ETV6-RUNX1; Pax5+/- leukemia”. In: *BMC Cancer* 15.1 (2015), p. 585 (cit. on p. 288).

[20] Nuri A Temiz, Branden S Moriarity, Natalie K Wolf, et al. “RNA sequencing of Sleeping Beauty transposon-induced tumors detects transposon-RNA fusions in forward genetic cancer screens”. In: *Genome Research* 26.1 (2016), pp. 119–129 (cit. on p. 288).

[21] Karen M Mann, Justin Y Newberg, Michael A Black, et al. “Analyzing tumor heterogeneity and driver genes in single myeloid leukemia cells with SBCapSeq”. In: *Nature Biotechnology* 34.9 (2016), pp. 962–972 (cit. on p. 288).

[22] Roman Maresch, Sebastian Mueller, Christian Veltkamp, et al. “Multiplexed pancreatic genome engineering and cancer induction by transfection-based CRISPR/Cas9 delivery in mice”. In: *Nature Communications* 7 (2016), p. 10770 (cit. on p. 288).

[23] Zoë N Rogers, Christopher D McFarland, Ian P Winters, et al. “A quantitative and multiplexed approach to uncover the fitness landscape of tumor suppression in vivo”. In: *Nature Methods* 14.7 (2017), p. 737 (cit. on p. 288).

[24] Daniele Perna, Florian A Karreth, Alistair G Rust, et al. “BRAF inhibitor resistance mediated by the AKT pathway in an oncogenic BRAF mouse melanoma model”. In: *Proceedings of the National Academy of Sciences* 112.6 (2015), E536–E545 (cit. on p. 288).

[25] Emilie A Chapeau, Agnieszka Gembarska, Eric Y Durand, et al. “Resistance mechanisms to TP53-MDM2 inhibition identified by in vivo piggyBac transposon mutagenesis screen in an Arf-/- mouse model”. In: *Proceedings of the National Academy of Sciences* 114.12 (2017), pp. 3151–3156 (cit. on p. 288).

[26] Ewald van Dyk, Marlous Hoogstraat, Jelle ten Hoeve, Marcel JT Reinders, and Lodewyk FA Wessels. "RUBIC identifies driver genes by detecting recurrent DNA copy number breaks". In: *Nature Communications* 7 (2016), p. 12159 (cit. on p. 289).

[27] Craig H Mermel, Steven E Schumacher, Barbara Hill, et al. "GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy number alteration in human cancers". In: *Genome Biology* 12.4 (2011), R41 (cit. on p. 289).

[28] Kirstine Jacobsen, Jordi Bertran-Alamillo, Miguel Angel Molina, et al. "Convergent Akt activation drives acquired EGFR inhibitor resistance in lung cancer". In: *Nature Communications* 8.1 (2017), p. 410 (cit. on p. 290).

[29] Chong Sun, Liqin Wang, Sidong Huang, et al. "Reversible and adaptive resistance to BRAF (V600E) inhibition in melanoma". In: *Nature* 508.7494 (2014), p. 118 (cit. on p. 290).

[30] Adam C Palmer and Peter K Sorger. "Combination cancer therapy can confer benefit via patient-to-patient variability without drug additivity or synergy". In: *Cell* 171.7 (2017), pp. 1678–1691 (cit. on p. 290).

[31] Charles Gawad, Winston Koh, and Stephen R Quake. "Single-cell genome sequencing: current state of the science". In: *Nature Reviews Genetics* 17.3 (2016), p. 175 (cit. on p. 291).

[32] Serena Liu and Cole Trapnell. "Single-cell transcriptome sequencing: recent advances and remaining challenges". In: *F1000 Research* 5 (2016) (cit. on p. 291).

[33] Nicholas McGranahan and Charles Swanton. "Clonal heterogeneity and tumor evolution: past, present, and the future". In: *Cell* 168.4 (2017), pp. 613–628 (cit. on p. 291).

[34] Woosung Chung, Hye Hyeon Eum, Hae-Ock Lee, et al. "Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer". In: *Nature Communications* 8 (2017), p. 15081 (cit. on p. 291).

[35] Chunhong Zheng, Liangtao Zheng, Jae-Kwang Yoo, et al. "Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing". In: *Cell* 169.7 (2017), pp. 1342–1356 (cit. on p. 291).

[36] Boris J Winterhoff, Makayla Maile, Amit Kumar Mitra, et al. "Single cell sequencing reveals heterogeneity within ovarian cancer epithelium and cancer associated stromal cells". In: *Gynecologic Oncology* 144.3 (2017), pp. 598–606 (cit. on p. 291).

[37] Alice Giustacchini, Supat Thongjuea, Nikolaos Barkas, et al. "Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia". In: *Nature Medicine* 23.6 (2017), p. 692 (cit. on p. 291).

[38] Nicholas E Navin. "Tumor evolution in response to chemotherapy: phenotype versus genotype". In: *Cell Reports* 6.3 (2014), pp. 417–419 (cit. on p. 291).

[39] Charissa Kim, Ruli Gao, Emi Sei, et al. "Chemoresistance evolution in triple-negative breast cancer delineated by single-cell sequencing". In: *Cell* 173.4 (2018), pp. 879–893 (cit. on p. 291).

[40] Nicholas Navin, Jude Kendall, Jennifer Troge, et al. "Tumour evolution inferred by single-cell sequencing". In: *Nature* 472.7341 (2011), p. 90 (cit. on p. 291).

[41] Yong Wang, Jill Waters, Marco L Leung, et al. “Clonal evolution in breast cancer revealed by single nucleus genome sequencing”. In: *Nature* 512.7513 (2014), p. 155 (cit. on p. 291).

[42] Jeong-Sun Seo, Arang Rhie, Junsoo Kim, et al. “De novo assembly and phasing of a Korean human genome”. In: *Nature* 538.7624 (2016), p. 243 (cit. on p. 291).

[43] Mircea Cretu Stancu, Markus J Van Roosmalen, Ivo Renkens, et al. “Mapping and phasing of structural variation in patient genomes using nanopore sequencing”. In: *Nature Communications* 8.1 (2017), p. 1326 (cit. on p. 292).

[44] Lucy C de Jong, Simone Cree, Vanessa Lattimore, et al. “Nanopore sequencing of full-length BRCA1 mRNA transcripts reveals co-occurrence of known exon skipping events”. In: *Breast Cancer Research* 19.1 (2017), p. 127 (cit. on p. 292).

[45] Mohan T Bolisetty, Gopinath Rajadinakaran, and Brenton R Graveley. “Determining exon connectivity in complex mRNAs by nanopore sequencing”. In: *Genome Biology* 16.1 (2015), p. 204 (cit. on p. 292).

[46] Sohyun Hwang, Eiru Kim, Insuk Lee, and Edward M Marcotte. “Systematic comparison of variant calling pipelines using gold standard personal exome variants”. In: *Scientific Reports* 5 (2015), p. 17875 (cit. on p. 292).

[47] Adam Cornish and Chittibabu Guda. “A comparison of variant calling pipelines using genome in a bottle as a reference”. In: *BioMed Research International* 2015 (2015), p. 456479 (cit. on p. 292).

[48] Tyler S Alioto, Ivo Buchhalter, Sophia Derdak, et al. “A comprehensive assessment of somatic mutation detection in cancer using whole-genome sequencing”. In: *Nature Communications* 6 (2015), p. 10001 (cit. on p. 292).

[49] Xiangtao Liu, Shizhong Han, Zuoheng Wang, Joel Gelernter, and Bao-Zhu Yang. “Variant callers for next-generation sequencing data: a comparison study”. In: *PloS One* 8.9 (2013), e75619 (cit. on p. 292).

[50] Justin M Zook, David Catoe, Jennifer McDaniel, et al. “Extensive sequencing of seven human genomes to characterize benchmark reference materials”. In: *Scientific Data* 3 (2016), p. 160025 (cit. on p. 292).

[51] Johannes Köster and Sven Rahmann. “Snakemake—a scalable bioinformatics workflow engine”. In: *Bioinformatics* 28.19 (2012), pp. 2520–2522 (cit. on p. 292).

[52] Paolo Di Tommaso, Maria Chatzou, Evan W Floden, et al. “Nextflow enables reproducible computational workflows”. In: *Nature Biotechnology* 35.4 (2017), p. 316 (cit. on p. 292).

[53] Björn Grüning, Ryan Dale, Andreas Sjödin, et al. “Bioconda: sustainable and comprehensive software distribution for the life sciences”. In: *Nature Methods* 15.7 (2018), pp. 475–476 (cit. on p. 292).

[54] Gregory M Kurtzer, Vanessa Sochat, and Michael W Bauer. “Singularity: Scientific containers for mobility of compute”. In: *PloS One* 12.5 (2017), e0177459 (cit. on p. 293).

[55] Tanya Barrett, Stephen E Wilhite, Pierre Ledoux, et al. “NCBI GEO: archive for functional genomics data sets—update”. In: *Nucleic Acids Research* 41.D1 (2012), pp. D991–D995 (cit. on p. 293).

[56] Rasko Leinonen, Ruth Akhtar, Ewan Birney, et al. “The European nucleotide archive”. In: *Nucleic Acids Research* 39.suppl_1 (2010), pp. D28–D31 (cit. on p. 293).

[57] Ethan Cerami, Jianjiong Gao, Ugur Dogrusoz, et al. “The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data”. In: *Cancer Discovery* 2.5 (2012), pp. 401–404 (cit. on p. 293).

[58] Jessica C Walrath, Jessica J Hawes, Terry van Dyke, and Karlyne M Reilly. “Genetically engineered mouse models in cancer research”. In: *Advances in Cancer Research*. Vol. 106. Elsevier, 2010, pp. 113–164 (cit. on p. 293).

[59] Kristopher K Frese and David A Tuveson. “Maximizing mouse cancer models”. In: *Nature Reviews Cancer* 7.9 (2007), p. 654 (cit. on p. 293).

[60] Douglas Hanahan, Erwin F Wagner, and Richard D Palmiter. “The origins of oncomice: a history of the first transgenic mice genetically engineered to develop cancer”. In: *Genes & Development* 21.18 (2007), pp. 2258–2270 (cit. on p. 293).

[61] William G Kaelin Jr. “Common pitfalls in preclinical cancer target validation”. In: *Nature Reviews Cancer* 17.7 (2017), p. 425 (cit. on p. 293).

[62] Wen Bu, Li Xin, Michael Toneff, Lei Li, and Yi Li. “Lentivirus vectors for stably introducing genes into mammary epithelial cells in vivo”. In: *Journal of Mammary Gland Biology and Neoplasia* 14.4 (2009), p. 401 (cit. on p. 294).

[63] Stefano Annunziato, Sjors M Kas, Micha Nethe, et al. “Modeling invasive lobular breast carcinoma by CRISPR/Cas9-mediated somatic genome editing of the mammary gland”. In: *Genes & Development* 30 (2016), pp. 1470–1480 (cit. on p. 294).

[64] Jarno Drost, Wouter R Karthaus, Dong Gao, et al. “Organoid culture systems for prostate epithelial and cancer tissue”. In: *Nature Protocols* 11.2 (2016), p. 347 (cit. on p. 294).

[65] Norman Sachs and Hans Clevers. “Organoid cultures for the analysis of cancer phenotypes”. In: *Current opinion in genetics & development* 24 (2014), pp. 68–73 (cit. on p. 294).

[66] Marc van de Wetering, Hayley E Francies, Joshua M Francis, et al. “Prospective derivation of a living organoid biobank of colorectal cancer patients”. In: *Cell* 161.4 (2015), pp. 933–945 (cit. on p. 294).

[67] F Stephen Hodi, Steven J O’day, David F McDermott, et al. “Improved survival with ipilimumab in patients with metastatic melanoma”. In: *New England Journal of Medicine* 363.8 (2010), pp. 711–723 (cit. on p. 294).

[68] Julie R Brahmer, Scott S Tykodi, Laura QM Chow, et al. “Safety and activity of anti–PD-L1 antibody in patients with advanced cancer”. In: *New England Journal of Medicine* 366.26 (2012), pp. 2455–2465 (cit. on p. 294).

[69] Roy S Herbst, Jean-Charles Soria, Marcin Kowanetz, et al. “Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients”. In: *Nature* 515.7528 (2014), p. 563 (cit. on p. 294).

[70] Thomas Powles, Joseph Paul Eder, Gregg D Fine, et al. “MPDL3280A (anti-PD-L1) treatment leads to clinical activity in metastatic bladder cancer”. In: *Nature* 515.7528 (2014), p. 558 (cit. on p. 294).