



Trajectories of genetics, 150 years after Mendel/Trajectoires de la génétique, 150 ans après Mendel  
 Using large-scale genome variation cohorts to decipher the molecular mechanism of cancer



*Étude à grande échelle de variations génétiques pour déchiffrer les mécanismes moléculaires de l'oncogénèse*

Nina Habermann<sup>a</sup>, Balca R. Mardin<sup>a</sup>, Sergei Yakneen<sup>a</sup>, Jan O. Korbel<sup>a,b,\*</sup>

<sup>a</sup> European Molecular Biology Laboratory (EMBL), Genome Biology Unit, Meyerhofstraße 1, 69117 Heidelberg, Germany

<sup>b</sup> European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, CB10 1SD Cambridge, UK

ARTICLE INFO

Article history:

Received 5 April 2016

Accepted after revision 30 April 2016

Available online 21 June 2016

Keywords:

Genetic variation  
 Chromothripsis  
 Cancer genomics  
 Structural variation  
 Big data  
 Hybrid cloud

Mots clés :

Variation génétique  
 Chromothripsie  
 Génomique du cancer  
 Variation structurale  
 Mégadonnées  
 Informatique en nuages hybrides

ABSTRACT

Characterizing genomic structural variations (SVs) in the human genome remains challenging, and there is a growing interest to understand somatic SVs occurring in cancer, a disease of the genome. A havoc-causing SV process known as chromothripsis scars the genome when localized chromosome shattering and repair occur in a one-off catastrophe. Recent efforts led to the development of a set of conceptual criteria for the inference of chromothripsis events in cancer genomes and to the development of experimental model systems for studying this striking DNA alteration process *in vitro*. We discuss these approaches, and additionally touch upon current “Big Data” efforts that employ hybrid cloud computing to enable studies of numerous cancer genomes in an effort to search for commonalities and differences in molecular DNA alteration processes in cancer.

© 2016 Published by Elsevier Masson SAS on behalf of Académie des sciences.

RÉSUMÉ

Caractériser les variations structurelles génomiques (SVs) dans le génome humain reste difficile et comprendre ces variations somatiques survenant dans le cancer, une maladie du génome, fait l'objet d'un intérêt croissant. Un processus de SV ravageur, connu sous le nom de chromothripsie, endommage le génome par un phénomène de fragmentation-réparation localisé en une seule étape, qui devient soudainement catastrophique. Des efforts récents ont conduit à l'élaboration d'un ensemble de critères conceptuels pour révéler des événements de chromothripsie dans les cancers et pour développer des systèmes modèles expérimentaux pour étudier *in vitro* ce processus d'altération de l'ADN. Nous discuterons ces approches et de plus, nous évoquerons les efforts actuels qui, mettant en œuvre des ressources informatiques à

\* Corresponding author. European Molecular Biology Laboratory (EMBL), Genome Biology Unit, Meyerhofstraße 1, 69117 Heidelberg, Germany.  
 E-mail address: korbel@embl.de (J.O. Korbel).

l'échelle mondiale de type « Big Data » sur de nombreux génomes, recherchent des points communs et des différences dans les processus d'altération moléculaire de l'ADN dans le cancer.

© 2016 Publié par Elsevier Masson SAS au nom de Académie des sciences.

## 1. Introduction

Genetic variation is a fundamental reason why humans differ from one another and why some individuals are more susceptible to diseases such as cancer than others, and there is a growing interest in investigating the mechanisms and the phenotypic consequences of genetic variation. A sharp decrease in the costs of DNA sequencing has enabled the sequencing of numerous human genomes and their mining using “Big Data” analytical approaches for unraveling molecular disease processes [1]. One example for such studies is the presently ongoing Pan-Cancer Analysis of Whole Genomes (PCAWG) project (<https://dcc.icgc.org/pcawg>), a forerunner project in Big Data analytics of genomes from patients, which aims to integrate data from somatic and germline whole genomes, DNA methylomes, transcriptomes, and clinical data from more than 2800 cancer patients amounting to nearly a Petabyte of sequencing data [1]. The objective of PCAWG is to unravel commonalities and to distinguish factors between cancer types and subtypes at the molecular level, to facilitate the molecular classification of malignancies with impact on diagnostics and treatment, and to uncover causalities linking genotype, environment, and phenotype. The unprecedented resource developed through PCAWG will enable standardized analysis of cancer genomes and associated datasets including transcriptomes, DNA methylomes, and clinical data to obtain insights into molecular disease processes relevant to cancer.

The objective of this paper is to briefly review methodologies for analyzing disease processes, with a specific focus on complex DNA rearrangements, in cancer genomes. Additionally, we will provide an outlook to coming Big Data efforts – with one example being the PCAWG project – which will facilitate the understanding of basic processes as well as translational research (Box 1).

## 2. Cancer genomes can evolve through catastrophes: the molecular process of chromothripsis

Cancer genome sequencing has enabled new insights into how tumors evolve, and has led to quite remarkable findings relating to the fact that cancer is not merely driven by stepwise alterations but can arise in conjunction with bursts of mutational events [2]. One particularly remarkable example for this is chromothripsis, a molecular process first described by Stephens et al. in 2011 based on cancer genome analysis, which can scar individual chromosome arms or one up to several chromosomes when localized chromosome shattering and repair occurs in a one-off catastrophe [3]. Rearrangement patterns associated with chromothripsis occur in approximately 2–3% of human cancers [3] and *TP53* germline mutations are linked with the occurrence of chromothripsis in pediatric medulloblastoma [4]. Chromothripsis is also

abundant in other cancers, such as bladder [5], breast [6], melanoma [7], and in bone cancers [3]. While the prevalence of chromothripsis in diverse cancer cell lines and cancer genomes [3,4,8] suggest a crucial role of chromothripsis in cancer development, the reproducible inference of this process has remained challenging, requiring that cataclysmic one-off rearrangements can be distinguished from localized genetic lesions that occur in a stepwise fashion. We recently devised a set of conceptual criteria for the inference of complex DNA rearrangements suitable for rigorous statistical analyses, which included previously established [3] as well as novel criteria: clustering of breakpoints, regularity of oscillating copy-number states, interspersed loss and retention of heterozygosity, prevalence of rearrangements affecting a specific haplotype, randomness of DNA segment order and fragment joins, as well as the ability to walk the derivative chromosome [8]. These criteria attempt to reject the alternative hypothesis that DNA rearrangements have occurred in a stepwise (progressive) fashion. Further refinement of these criteria allow inferring chromothripsis events in conjunction with additional stepwise patterns of DNA alteration [9], and collectively, these criteria have begun to be used quite regularly to operationally define chromothripsis events based on cancer genome sequencing data.

### Box 1. Challenges in genomics due to recent increase in data set sizes.

Opportunities emerging from large datasets are not without challenges. When using a typical university Internet connection, for example, it would take more than 15 months to transfer a data set such as the one amassed by the ICGC from its data repository into a researcher's local IT infrastructure [1]. And the hardware needed to store, let alone process the data, would produce costs at a million dollar level each year. So, consider a PhD student pursuing a 3-year thesis project at a European university. Before she/he could even consider to start pursuing analyses, 15 months of download time would need to be overcome – a considerable portion of the overall time frame anticipated for the thesis, and a million USD in storage would need to be spent before any research work could begin.

Given the establishment of new sequencing platforms that can now sequence over 10,000 human genomes a year at several locations around the world, the number of human genomes generated in the context of disease research is going to be ramped up dramatically in the coming 5 to 10 years. Genomics England Ltd, a government-driven initiative in the UK, for example, plans to sequence 100,000 human genomes – including 25,000 from cancer patients, by the end of 2017.

### 3. Methods for mechanistic dissection of catastrophic DNA rearrangement processes

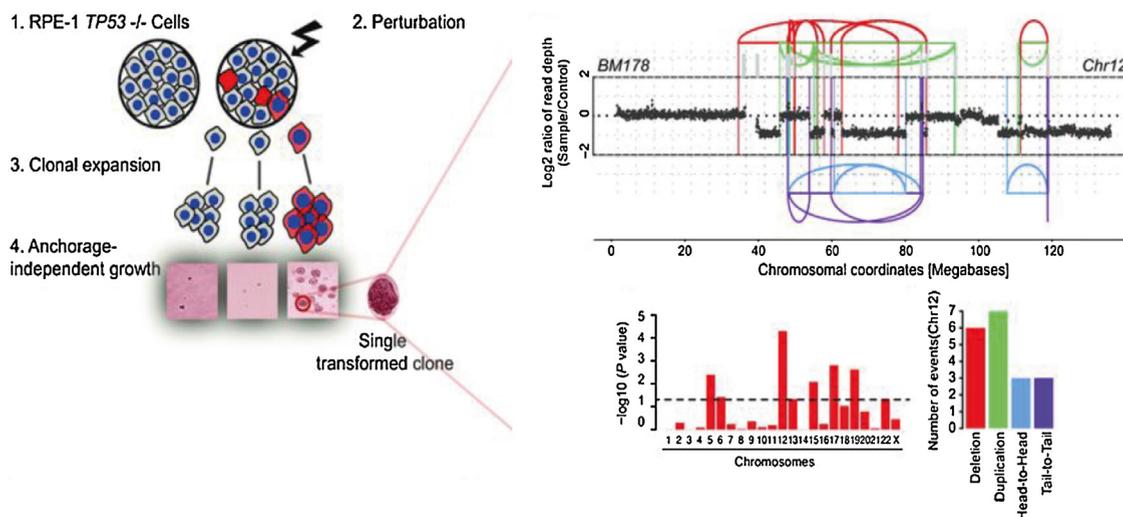
Although cellular catastrophes occurring during key stages of the cell cycle were proposed to play roles in initiating catastrophic DNA rearrangements [10], the mechanistic basis of chromothripsis has not been studied until recently, largely due to lack of suitable cell-based model systems. In order to investigate the underlying mechanism(s) of chromothripsis events, recent studies have described cell-based models for characterizing chromothripsis, which include a methodology termed CAST (*Complex Alterations After Selection And Transformation*) developed by our group [11]. CAST (Fig. 1) is based on an untransformed model cell line (hTERT-RPE-1), application of genetic or chemical perturbations, selection of DNA alterations conferring a growth advantage by soft agar colony formation and screening for extensive copy-number alterations using low-coverage DNA sequencing followed by in-depth characterization of complex rearrangements by long-range paired-end sequencing [11].

Using this *in vitro* system, we were able to reproducibly generate chromothripsis events in RPE-1 cells upon global DNA damage by doxorubicin. In addition, we performed in-depth characterization of cell lines harboring chromothripsis alterations that were generated by CAST, and we applied the above-mentioned criteria to obtain additional insights into the process [11]. By doing so, we were able to demonstrate that telomeric stress can initiate chromothripsis events *in vitro* (when using an siRNA against the shelterin complex component *TRF2*) consistent with previous genomic analyses suggesting that chromothripsis can be initiated from dicentric chromosomes and breakage-fusion-bridge cycles [9]. The link between telomere crisis and complex DNA rearrangements were further

substantiated by recent reports providing compelling evidence for telomeric stress being an initiating factor for complex rearrangements both *in vitro* mediated by TREX-1 endonuclease following telomere crisis as well as [12] *in vivo* in medulloblastomas, ependymomas, glioblastomas, and chronic lymphocytic leukemias [13]. Another important implication derived from this *in vitro* system was the association of an increase in ploidy either in the form of tetraploidy (four chromosome sets) or hyperploidy (presence of extra chromosomes in the form of incomplete sets) with chromothripsis. This is an interesting finding considering previous studies suggesting an increase in genomic instability in response to tetraploidy [14]. In the context of chromothripsis, we were able to uncover that the increase in ploidy predisposes to complex genomic rearrangements also *in vivo* analyzing primary SHH-medulloblastoma tumor genomes. In these tumors, we found that hyperploidy is not only strongly associated with chromothripsis, but is also plausibly the initiating factor for complex SRs [11].

Another attractive model for chromothripsis is based on the presumption that chromosomes confined in micronuclei can undergo a catastrophic shattering process in a one-off event [15], a model which recently was significantly substantiated through single-cell sequencing data providing compelling evidence for chromothripsis [16]. It is important to note in this regard that there is also a plausible connection between telomere crises and micronuclei formation, whereby upon telomere loss sister chromatid fusion events may mediate the formation of dicentric chromosomes, which in turn may result in micronuclei formation as a consequence of chromosome segregation defects [11].

In summary, due to recent advances in chromothripsis research, we are now beginning to understand the cellular processes and factors that might be involved in instigating



**Fig. 1.** The CAST approach, providing insights into determinants of chromothripsis. CAST allows the mechanistic dissection of chromothripsis – a cellular catastrophe resulting in massive simultaneous DNA rearrangements. Using hTERT RPE-1 cells, we reproducibly created chromothripsis *in vitro* by chemical and genetic perturbations. Clonal expansion of cells is followed by their in-depth characterization to obtain additional insights into the process (top right panel). CAST revealed an increased rate of chromothripsis in hyperploidy cells (bottom right panel) and following telomere shortening (through siRNA induced silencing of a shelterin complex component), findings that we verified by genome sequencing in primary brain tumours. This figure has been adapted from Fig. 1 of the article by Mardin et al. [11], ©2015 European Molecular Biology Organization.

such dramatic rearrangements. However, it still remains to be seen to what extent a misregulation of these factors is reflected in real tumors and which cellular process actually operates *in vivo* during tumorigenesis generating chromothripsis events.

#### 4. Big data: opportunity for genomics and cancer research

Investigations of molecular disease processes such as chromothripsis events can be significantly leveraged by access to large datasets that facilitate correlative studies to pursue hypothesis- and data-driven research. Thanks to improvements in sequencing technology, the volume of genomic data submitted to public archives is now well into the multipetabyte range (1 petabyte is  $10^{15}$  bytes) [1], and is beginning to match scientific dataset sizes that we previously only knew from physics. A large portion of the genome sequencing data recently generated is from human patients, with cancer genomics playing a forerunner role thanks to recent large-scale funding initiatives (including consortia projects by the International Cancer Genome Consortium [ICGC] for example; see [www.icgc.org](http://www.icgc.org)). Within the ICGC, groups from 17 countries have amassed a data set in excess of two petabytes – equivalent to the capacity of roughly 20,000 smartphones – in just 5 years [1] (Box 2).

Analysis of this treasure trove of data has the potential to enable a range of analyses of relevance to systems biology, computational biology, and biomedicine, which so far could not be undertaken. Cancer genome analyses benefit from large sample sizes, since common cancers are typically classified into numerous subtypes which characteristically show different patterns of mutation. In order to learn about cancer biology and treatment options for patients, and also to uncover relationships between molecular data and clinical data, large sets of patient genomes as well as associated datasets (e.g. epigenomes and transcriptomes) are needed. Furthermore, since major initiatives such as the Cancer Genome Atlas (TCGA) project based in the USA used mainly exome sequencing (i.e., sequencing of those 1–2% of the human genome encoding for protein-coding genes), much less is currently known about mutations in non-coding regions that contain most gene regulatory information – in spite of initial success stories showing that these regions can indeed be very relevant to cancer [17–19]. Additionally, since previous cancer studies focused largely on uncovering somatic DNA variation, relatively little is known about how germline DNA variants may influence somatic mutations, although ~10% of all cancers likely have a hereditary cause [20].

There is by now wide agreement in the human genomics research community that the challenges of accessing Big Data sets are now limiting the scientists' ability to do research, and especially to replicate and build on previous work. One main objective of a novel major international initiative, the PCAWG study, is to identify common patterns of mutation in more than 2800 cancer whole genomes from the International Cancer Genome Consortium. PCAWG is exploring the nature and consequences of somatic and germline variations in both coding

#### Box 2. Big data analysis within PCAWG.

The key technical challenge addressed by the PCAWG project is the collection and uniform analysis of whole genome sequencing data for > 2800 cancer patients, as well as additional data such as RNA and DNA methylation, where available. The total data set size under analysis now exceeds 700 TB (Table 1). To address this challenge, seven project data repositories have been established within high-performance computing centers in Barcelona, Chicago, Heidelberg (DKFZ), London (EMBL-EBI), Santa Cruz, Seoul, and Tokyo. These seven centers have kindly donated computing capacity to perform data processing, so far totaling 8824 computing cores, 24 TB of RAM and 6 PB of hard disk storage. Four out of these seven centers act as academic community cloud computing centers for the purposes of this project, whereas the remaining three contribute their data processing capabilities as traditional High performance computing centers. A commercial cloud-computing provider – Amazon AWS, was also used to analyze about 20% of the samples [1].

In order to facilitate the comparison among diverse tumour types, all tumour and matched normal genomes are subjected to a uniform set of alignment and variant calling algorithms, and a rigorous set of quality control tests. The research activities are coordinated by the steering committee of the project, which stewards a series of working groups comprising more than 700 scientists and covering a range of themes. A set of computer centres around the world share the work load of processing all the data available to PCAWG using a standardized set of analysis tools developed in the context of three major pipelines: the Sanger pipeline, the Broad Institute pipeline and the Heidelberg pipeline developed at the DKFZ and the European Molecular Biology Laboratory (EMBL), and also includes dedicated germline genome analysis pipelines developed at Annai/Stanford as well as at EMBL and EMBL-EBI.

and non-coding regions, with particular emphasis on *cis*-regulatory sites, non-coding RNAs, and large-scale structural alterations [1].

#### 5. Challenges of Big Data analyses

A key determinant for, and perhaps one of the biggest benefits of, cloud computing is the rapid scalability of computational analyses (also often referred to as elasticity) – i.e. the ability to scale up vastly when need increases or down if resources are not being used. Several researchers can work in parallel, sharing their data and methods with ease by performing their analyses within cloud-based virtual computers that can be accessed from desktop computers. Thus, the analysis of a big genome data set that might have previously taken months can be executed in days or in weeks. Especially in public/commercial clouds computing capabilities available for provisioning often appear to be unlimited (e.g. with exceptionally high peak time requirements), which enables applications in the life sciences that may not be feasible/practical at all within a

**Table 1**

Data types and data set size of the Pan-cancer analysis of whole genomes (PCAWG) project (before filtering low-quality samples).

Data types	Number of samples (number of repositories holding at least 50%)
Whole genome data	2834 (2)
RNA-Seq	1367 (1)
DNA Methyl-Seq	467 (1)

single or a few localized data centers. We have recently developed recommendations for the development of National clouds [21]. In our view, an optimal model for the European life science research community is what is termed a hybrid cloud, which combines academic localized data centers provisioning cloud computing to the community and keeping some of the most highly valuable life science datasets locally, as well as public/commercial cloud computing (e.g. using European commercial partners through the Helix Nebula public – private partnership for cloud computing; [www.helix-nebula.eu](http://www.helix-nebula.eu)) providing vast scalability on demand to enable new research applications. Another key advantage for a hybrid cloud is that such a cloud model does not require centralized planning, and could (as long as this is based on agreed standards and frameworks) be built around different funding sources, governance structures and organizational models, to facilitate the (data-heavy) future of collaborative basic and translational research in Europe.

## 6. Conclusion and future

The field of genomics is rapidly changing with the accessibility of new technologies that make human genome sequencing a regular tool employed in molecular biology and genetics laboratories, in biomedical research and in the clinics. The data emerging from these efforts will be of tremendous value not only for human health, but also for molecular research to uncover the basic underpinnings of disease processes such as chromothripsis and to make advances in translational cancer research. Projects such as PCAWG will show us some of the challenges while working with such sets of “Big Data”. More work will be needed to realize collaborative processing of such datasets in different areas of the life sciences, and we propose a hybrid cloud model of collaborative and shared data processing – e.g. to compare crucial data across projects/cohorts to advance the life sciences.

## Disclosure of interest

The authors declare that they have no competing interest.

## Acknowledgements

We thank the technical working group of the Pan-Cancer Analysis of Whole Genomes (PCAWG) initiative, especially Lincoln Stein, Christina Yung, and Brian

O'Connor, for their crucial efforts in compiling the PCAWG donor samples (summarized in Table 1), and thank the members of the Steering committee of PCAWG for facilitating this international initiative.

## References

- [1] L.D. Stein, B.M. Knoppers, P. Campbell, G. Getz, J.O. Korbel, Data analysis: create a cloud commons, *Nature* 523 (2015) 149–151.
- [2] L.R. Yates, P.J. Campbell, Evolution of the cancer genome, *Nat. Rev. Genet.* 13 (2012) 795–806.
- [3] P.J. Stephens, C.D. Greenman, B. Fu, F. Yang, G.R. Bignell, L.J. Mudie, E.D. Pleasance, K.W. Lau, D. Beare, L.A. Stebbings, S. McLaren, M.L. Lin, D.J. McBride, I. Varela, S. Nik-Zainal, C. Leroy, M. Jia, A. Menzies, A.P. Butler, J.W. Teague, M.A. Quail, J. Burton, H. Swerdlow, N.P. Carter, L.A. Morsberger, C. Iacobuzio-Donahue, G.A. Follows, A.R. Green, A.M. Flanagan, M.R. Stratton, P.A. Futreal, P.J. Campbell, Massive genomic rearrangement acquired in a single catastrophic event during cancer development, *Cell* 144 (2011) 27–40.
- [4] T. Rausch, D.T.W. Jones, M. Zapatka, A.M. Stutz, T. Zichner, J. Weischenfeldt, N. Jager, M. Remke, D. Shih, P.A. Northcott, E. Pfaff, J. Tica, Q. Wang, L. Massimi, H. Witt, S. Bender, S. Pleier, H. Cin, C. Hawkins, C. Beck, A. von Deimling, V. Hans, B. Brors, R. Eils, W. Scheurle, J. Blake, V. Benes, A.E. Kulozik, O. Witt, D. Martin, C. Zhang, R. Porat, D.M. Merino, J. Wasserman, N. Jabado, A. Fontebasso, L. Bullinger, F.G. Rucker, K. Dohner, H. Dohner, J. Koster, J.J. Molenaar, R. Versteeg, M. Kool, U. Tabori, D. Malkin, A. Korshunov, M.D. Taylor, P. Lichter, S.M. Pfister, J.O. Korbel, Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations, *Cell* 148 (2012) 59–71.
- [5] C.D. Morrison, P. Liu, A. Woloszynska-Read, J. Zhang, W. Luo, M. Qin, W. Bshara, J.M. Conroy, L. Sabatini, P. Vedell, D. Xiong, S. Liu, J. Wang, H. Shen, Y. Li, A.R. Omilian, A. Hill, K. Head, K. Guru, D. Kunnev, R. Leach, K.H. Eng, C. Darlak, C. Hoefflich, S. Veeranki, S. Glenn, M. You, S.C. Pruitt, C.S. Johnson, D.L. Trump, Whole-genome sequencing identifies genomic heterogeneity at a nucleotide and chromosomal level in bladder cancer, *Proc. Natl. Acad. Sci. U. S. A.* 111 (2014) E672–E681.
- [6] E. Przybytkowski, E. Lenkiewicz, M.T. Barrett, K. Klein, S. Nabavi, C.M. Greenwood, M. Basik, Chromosome-breakage genomic instability and chromothripsis in breast cancer, *BMC Genom.* 15 (2014) 579.
- [7] A.C. van Engen-van Grunsven, M.P. Baar, R. Pfundt, J. Rijntjes, H.V. Kusters-Vandevelde, A.L. Delbecq, J.E. Keunen, J.B. Klevering, P. Wesseling, W.A. Blokx, P.J. Groenen, Whole-genome copy-number analysis identifies new leads for chromosomal aberrations involved in the oncogenesis and metastatic behavior of uveal melanomas, *Melanoma Res.* 25 (2015) 200–209.
- [8] J. Weischenfeldt, R. Simon, L. Feuerbach, K. Schlangen, D. Weichenhan, S. Minner, D. Wuttig, H.J. Warnatz, H. Stehr, T. Rausch, N. Jager, L. Gu, O. Bogatyrova, A.M. Stutz, R. Claus, J. Eils, R. Eils, C. Gerhauser, P.H. Huang, B. Hutter, R. Kabbe, C. Lawrenz, S. Radomski, C.C. Bartholomae, M. Falth, S. Gade, M. Schmidt, N. Amschler, T. Hass, R. Galal, J. Gjonj, R. Kuner, C. Baer, S. Maser, C. von Kalle, T. Zichner, V. Benes, B. Raeder, M. Mader, V. Amstislavskiy, M. Avci, H. Lehrach, D. Parkhomchuk, M. Sultan, L. Burkhardt, M. Graefen, H. Huland, M. Kluth, A. Krohn, H. Sirma, L. Stumm, S. Steurer, K. Grupp, H. Sultmann, G. Sauter, C. Plass, B. Brors, M.L. Yaspo, J.O. Korbel, T. Schlomm, Integrative genomic analyses reveal an androgen-driven somatic alteration landscape in early-onset prostate cancer, *Cancer Cell* 23 (2013) 159–170.
- [9] Y. Li, C. Schwab, S.L. Ryan, E. Papaemmanuil, H.M. Robinson, P. Jacobs, A.V. Moorman, S. Dyer, J. Borrow, M. Griffiths, N.A. Heerema, A.J. Carroll, P. Talley, N. Bown, N. Telford, F.M. Ross, L. Gaunt, R.J. McNally, B.D. Young, P. Sinclair, V. Rand, M.R. Teixeira, O. Joseph, B. Robinson, M. Maddison, N. Dastugue, P. Vandenbergh, C. Haferlach, P.J. Stephens, J. Cheng, P. Van Loo, M.R. Stratton, P.J. Campbell, C.J. Harrison, Constitutional and somatic rearrangement of chromosome 21 in acute lymphoblastic leukaemia, *Nature* 508 (2014) 98–102.
- [10] M.J. Jones, P.V. Jallepalli, Chromothripsis: chromosomes in crisis, *Dev Cell* 23 (2012) 908–917.
- [11] B.R. Mardin, A.P. Drainas, S.M. Waszak, J. Weischenfeldt, M. Isokane, A.M. Stutz, B. Raeder, T. Efthymiopoulos, C. Buccitelli, M. Segura-Wang, P. Northcott, S.M. Pfister, P. Lichter, J. Ellenberg, J.O. Korbel, A cell-based model system links chromothripsis with hyperploidy, *Mol. Syst. Biol.* 11 (2015) 828.
- [12] J. Maciejowski, Y. Li, N. Bosco, P.J. Campbell, T. de Lange, Chromothripsis and kataegis induced by telomere crisis, *Cell* 163 (2015) 1641–1654.
- [13] A. Ernst, D.T. Jones, K.K. Maass, A. Rode, K.I. Deeg, B. Michael Chelliah Jebaraj, A. Korshunov, V. Hovestadt, M.A. Tainsky, K.W. Pajtlar, S. Bender, S. Brabetz, S. Grobner, M. Kool, F. Devens, J. Edelmann,

- C. Zhang, P. Castelo-Branco, U. Tabori, D. Malkin, K. Rippe, S. Stilgenbauer, S.M. Pfister, M. Zapatka, P. Lichter, Telomere dysfunction and chromothripsis, *Int. J. Cancer* (2016).
- [14] T. Fujiwara, M. Bandi, M. Nitta, E.V. Ivanova, R.T. Bronson, D. Pellman, Cytokinesis failure generating tetraploids promotes tumorigenesis in p53-null cells, *Nature* 437 (2005) 1043–1047.
- [15] M. Meyerson, D. Pellman, Cancer genomes evolve by pulverizing single chromosomes, *Cell* 144 (2011) 9–10.
- [16] C.Z. Zhang, A. Spektor, H. Cornils, J.M. Francis, E.K. Jackson, S.W. Liu, M. Meyerson, D. Pellman, Chromothripsis from DNA damage in micronuclei, *Nature* 522 (2015) 179.
- [17] S. Horn, A. Figl, P.S. Rachakonda, C. Fischer, A. Sucker, A. Gast, S. Kadel, I. Moll, E. Nagore, K. Hemminki, D. Schadendorf, R. Kumar, TERT promoter mutations in familial and sporadic melanoma, *Science* 339 (2013) 959–961.
- [18] F.W. Huang, E. Hodis, M.J. Xu, G.V. Kryukov, L. Chin, L.A. Garraway, Highly recurrent TERT promoter mutations in human melanoma, *Science* 339 (2013) 957–959.
- [19] P.A. Northcott, C. Lee, T. Zichner, A.M. Stutz, S. Erkek, D. Kawachi, D.J. Shih, V. Hovestadt, M. Zapatka, D. Sturm, D.T. Jones, M. Kool, M. Remke, F.M. Cavalli, S. Zuyderduyn, G.D. Bader, S. VandenBerg, L.A. Esparza, M. Ryzhova, W. Wang, A. Wittmann, S. Stark, L. Sieber, H. Seker-Cin, L. Linke, F. Kratochwil, N. Jager, I. Buchhalter, C.D. Imbusch, G. Zipprich, B. Raeder, S. Schmidt, N. Diessl, S. Wolf, S. Wiemann, B. Brors, C. Lawerenz, J. Eils, H.J. Warnatz, T. Risch, M.L. Yaspo, U.D. Weber, C.C. Bartholomae, C. von Kalle, E. Turanyi, P. Hauser, E. Sanden, A. Darabi, P. Siesjo, J. Sterba, K. Zitterbart, D. Sumerauer, P. van Sluis, R. Versteeg, R. Volckmann, J. Koster, M.U. Schuhmann, M. Ebinger, H.L. Grimes, G.W. Robinson, A. Gajjar, M. Mynarek, K. von Hoff, S. Rutkowski, T. Pietsch, W. Scheurlen, J. Felsberg, G. Reifenberger, A.E. Kulozik, A. von Deimling, O. Witt, R. Eils, R.J. Gilbertson, A. Korshunov, M.D. Taylor, P. Lichter, J.O. Korbel, R.J. Wechsler-Reya, S.M. Pfister, Enhancer hijacking activates GF11 family oncogenes in medulloblastoma, *Nature* 511 (2014) 428–434.
- [20] J.E. Garber, K. Offit, Hereditary cancer predisposition syndromes, *J. Clin. Oncol.* 23 (2005) 276–292.
- [21] B. Brors, W. Eberhardt, R. Eils, N. Habermann, S. Iakhnin, J.O. Korbel, C. Lawerenz, P. Lichter, R. Lück, F. Molnár-Gábor, T. Rausch, K. Sachs, M. Schlesner, C. von Kalle, S.M. Waszak, J. Weischenfeld, *Appl. Transl. Genom. Cloud White Paper* (2016), [www.genome-cloud.de](http://www.genome-cloud.de).