



# Somatic Copy Number Alterations in Human Cancers: An Analysis of Publicly Available Data From The Cancer Genome Atlas

Luuk Harbers<sup>1,2</sup>, Federico Agostini<sup>1,2</sup>, Marcin Nicos<sup>3</sup>, Dimitri Poddighe<sup>4,5</sup>, Magda Bienko<sup>1,2</sup> and Nicola Crosetto<sup>1,2\*</sup>

<sup>1</sup> Division of Genome Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden, <sup>2</sup> Bienko-Crosetto Lab, Science for Life Laboratory, Stockholm, Sweden, <sup>3</sup> Department of Pneumology, Oncology and Allergology, Medical University of Lublin, Lublin, Poland, <sup>4</sup> Department of Medicine, Nazarbayev University School of Medicine, Nur-Sultan, Kazakhstan, <sup>5</sup> Clinical Academic Department of Pediatrics, National Research Center for Maternal and Child Health, University Medical Center, Nur-Sultan, Kazakhstan

## OPEN ACCESS

### Edited by:

Pasquale Simeone,  
University of Studies G. d'Annunzio  
Chieti and Pescara, Italy

### Reviewed by:

Gisela Ceballos,  
Instituto Nacional de Medicina  
Genómica (INMEGEN), Mexico  
Angelo Veronese,  
Università degli Studi di Bari  
"Aldo Moro", Italy

### \*Correspondence:

Nicola Crosetto  
nicola.crosetto@ki.se

### Specialty section:

This article was submitted to  
Cancer Genetics,  
a section of the journal  
Frontiers in Oncology

Received: 26 April 2021

Accepted: 06 July 2021

Published: 28 July 2021

### Citation:

Harbers L, Agostini F, Nicos M, Poddighe D, Bienko M and Crosetto N (2021) Somatic Copy Number Alterations in Human Cancers: An Analysis of Publicly Available Data From The Cancer Genome Atlas. *Front. Oncol.* 11:700568. doi: 10.3389/fonc.2021.700568

Somatic copy number alterations (SCNAs) are a pervasive trait of human cancers that contributes to tumorigenesis by affecting the dosage of multiple genes at the same time. In the past decade, The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC) initiatives have generated and made publicly available SCNA genomic profiles from thousands of tumor samples across multiple cancer types. Here, we present a comprehensive analysis of 853,218 SCNAs across 10,729 tumor samples belonging to 32 cancer types using TCGA data. We then discuss current models for how SCNAs likely arise during carcinogenesis and how genomic SCNA profiles can inform clinical practice. Lastly, we highlight open questions in the field of cancer-associated SCNAs.

**Keywords:** copy number alterations, cancer, TCGA, cosmic genes, 3D genome

## INTRODUCTION

### A Bit of Semantics

Cancer cells almost invariably harbor altered karyotypes. Deviations from the normal karyotype can range from entire or partial gains or losses of chromosomes and large intra-chromosomal inversions or translocations between different chromosomes, to more complex rearrangements, such as "chromothripsis" (where massive genomic rearrangements are generated in a single event and are localized to isolated chromosomal regions) (1). Inversions, translocations, and complex rearrangements are usually referred to as structural variants (SVs). On the other hand, there is considerable confusion in the literature as to the terminology used to describe different deviations from the normal copy number state of diploid cells ( $2n$ ) (2). Strictly speaking, any deviation from the  $2n$  copy number state of a region in the genome represents a copy number alteration (CNA). However, gains or losses of entire chromosomes have historically been referred to as aneuploidy or whole chromosome aneuploidy, while gains or losses of chromosomal arms or smaller chromosomal regions have been variably termed segmental or partial aneuploidy, (focal) CNAs or small insertions or deletions (indels), depending on the size of the genomic region amplified or deleted (2). The mechanisms responsible for whole chromosome CNAs, arm-level CNAs, focal

CNAs and indels are different—as we discuss below—therefore calling for the use of separate terms. However, for simplicity, we adopt the convention recently proposed by Ben-David and Amon (2) and use: (i) the term aneuploidy to describe all CNAs affecting either entire chromosome arms (excluding the short arms of acrocentric chromosomes) or whole chromosomes; (ii) the term CNAs to describe all sub-arm gains or losses larger than 10 kilobases (kb); and (iii) the term indels to describe all other CNAs. Specifically, here we refer to CNAs arising post-zygotically in a somatic cell as somatic CNAs or SCNAs. In contrast, we refer to CNAs that occur in the germline, and are therefore inheritable, as copy number variants or CNVs.

## The Landscape of SCNAs Across Human Cancers

Since the launch of The Cancer Genome Atlas (TCGA) and International Cancer Genome Consortium (ICGC) initiatives in 2005 and 2008, respectively, SCNAs have been profiled in thousands of tumor samples in virtually all cancer types using high density single nucleotide polymorphism (SNP) arrays. Recently, the Pan-Cancer Analysis of Whole Genomes (PCAWG) consortium of TCGA and ICGC further expanded the already huge repertoire of publicly available cancer SCNA datasets, by profiling more than 2,600 cancers by whole genome sequencing (1, 3). A summary of the TCGA and PCAWG studies in which SCNAs have been profiled and/or analyzed is available as **Supplementary Table 1**. All these datasets are derived from primary tumors. Recently, the first pan-cancer whole genome analysis of metastatic tumors became available, which includes high resolution SCNA profiles from 2,520 samples (4). Altogether, these data offer a unique opportunity to explore the prevalence, type, and genomic distribution of SCNAs in different tumor types. In turn, this information can be used to guide hypotheses about the mechanisms of formation and evolution of SCNAs during tumorigenesis.

The first comprehensive pan-cancer analysis of SCNAs was published in 2013 and was based on a still relatively small set ( $n = 4,934$ ) of TCGA samples, which were available at that time (5). This study revealed broad differences in the prevalence of SCNAs across different tumor types, with some cancers such as ovarian carcinomas having a large fraction of their genome either amplified or deleted (5). The same study was also the first to reveal recurrent patterns of focal amplifications and deletions involving known cancer genes, such as amplifications of *CCND1*, *EGFR*, *MYC* and deletions of *CDKN2A*, *CDKN2B* and *STK11* (5). Since then, most pan-cancer analyses have focused on mutations in cancer genes and on mutational patterns, and only a few of them have specifically focused on SCNAs. In one of these analyses (6), 39,568 SCNAs from 3,131 tumors sequenced by TCGA were integrated with genome wide maps of chromosome contact frequencies measured with high-throughput chromosome conformation capture (Hi-C) (7). This analysis was the first to suggest a link between the genomic landscape of cancer SCNAs and the three-dimensional (3D) architecture of the genome, and to report a different distribution of amplifications and deletions within the 3D genome (6). More recently, the association

between the 3D genome and cancer SCNAs was further investigated by the PCAWG consortium, by integrating whole genome SV profiles from 2,658 cancers representing 38 tumor types with available Hi-C data from multiple cell lines (8). This comprehensive analysis revealed that deletions are more frequently found within so-called topologically associating domains (TADs) (9) and lamina associated domains (LADs) (10)—two types of structural domains shaping the 3D genome—whereas duplications tend to span multiple TADs and are more frequently detected within inter-LAD regions, suggesting that these alterations arise through different mechanisms. Integration of SCNA and TAD annotations, together with gene expression profiles, has also been used to identify the presence of rearrangements involving *cis*-regulatory elements, such as enhancers, within cancer genomes (11). This approach led to the discovery of enhancer hijacking events that result in the activation of the *IRS4* gene in a subset of lung cancers and of the *IGF2* gene in a subset of colorectal cancers, by bringing these genes in physical proximity to active enhancers that are normally not contacted by the same genes (11).

Although very insightful, we argue that pan-cancer analyses such as those mentioned above, which intersect various omics data with databases of SCNAs, need to be placed in the right context by a pan-cancer analysis of the SCNAs themselves, including a survey of their type, prevalence, and distribution across the genome. An analysis of this sort was recently performed on the prevalence and type of aneuploidies across 10,249 tumors and 32 cancer types using TCGA data (12), but it did not consider SCNAs as we define them here. In this article, we therefore re-analyze all the available SCNA datasets from TCGA and assess their prevalence, type and genomic distribution across different tumor types, and which cancer associated genes are mostly affected by these SCNAs. We then discuss the (limited) available information about how SCNAs likely emerge and evolve during the process of tumorigenesis and highlight open research questions in the field of cancer-associated SCNAs. We acknowledge that TCGA data do not provide an exhaustive portrait of genomic alterations in cancers and that important differences exist between the TCGA cohort and the U.S. general population of cancer patients (13), not to mention the rest of the world. We therefore warn the Reader against generalizing the conclusions of our analysis outside of the TCGA population.

## METHODS

### TCGA Datasets

We downloaded all the TCGA data and associated clinical information by creating a Genomic Data Commons (GDC) manifest of primary tumors using the R (version 4.0.3) package *GenomicDataCommons* (version 1.12.0) (14). We then filtered the manifest based on primary tumors for which copy number data are available. To obtain the relevant clinical data we translated TCGA barcodes using the R package *TCGAutils* (15) (version 1.8.1).

## Analysis of SCNA Frequency Across Different Tumor Types

We called genomic regions with a log<sub>2</sub> ratio above 0.32 as amplified, whereas we called genomic regions with a log<sub>2</sub> ratio below -0.42 as deleted. We then classified amplified and deleted genomic regions as following: (i) indels if they were smaller than 10 kb in size; (ii) aneuploidies if they spanned more than 75% of a chromosomal arm; (iii) SCNAs for all the other alterations. We computed the total percentage of the genome that was either amplified or deleted and the frequency of each alteration for every tumor sample, separately for all tumor types. When computing the alteration frequency per chromosome, we normalized the values by chromosome length. Finally, we computed the P values for the total percentage and for the total number of amplifications and deletions for each tumor type using the R function *wilcox.test()* and the Spearman correlation using the R function *cor()*, which are both available in the *stats* package (version 4.0.3).

## Analysis of SCNA Lengths

We computed the average length of SCNAs for every tumor sample, separately for all tumor types. We also calculated the average SCNA length for each chromosome, with and without normalization by chromosome length. We calculated the P values for the mean SCNA length of amplifications and deletions using the R function *wilcox.test()* which is available in the *stats* package (version 4.0.3).

## Analysis of COSMIC Genes Affected by SCNAs

To generate a list of genes associated with different tumor types, we downloaded the COSMIC Cancer Gene Census (16), a list of 723 cancer associated genes. Following this, we overlapped the genomic regions of these 723 genes with the previously called SCNAs regions in all the TCGA samples analyzed. Genes are called amplified or deleted if there is at least a 50% overlap of the gene with a SCNA.

## Analysis of Co-Amplified and Co-Deleted COSMIC Genes

To generate the list of genes pairs that are simultaneously affected by SCNAs, we analyzed all pair combinations of the 723 COSMIC genes and reported the frequency of all the possible co-occurring events (*i.e.*, AMP-AMP, DEL-DEL and AMP-DEL) across all experiments for each tumor type (see **Supplementary Table 2**).

## RESULTS

We analyzed the prevalence, type, and genomic distribution of 853,218 SCNAs across 10,729 tumor samples belonging to 32 cancer types (see *Methods*), being guided by the following five key questions.

## How Much of the Genome Is Affected by SCNAs in Different Cancer Types?

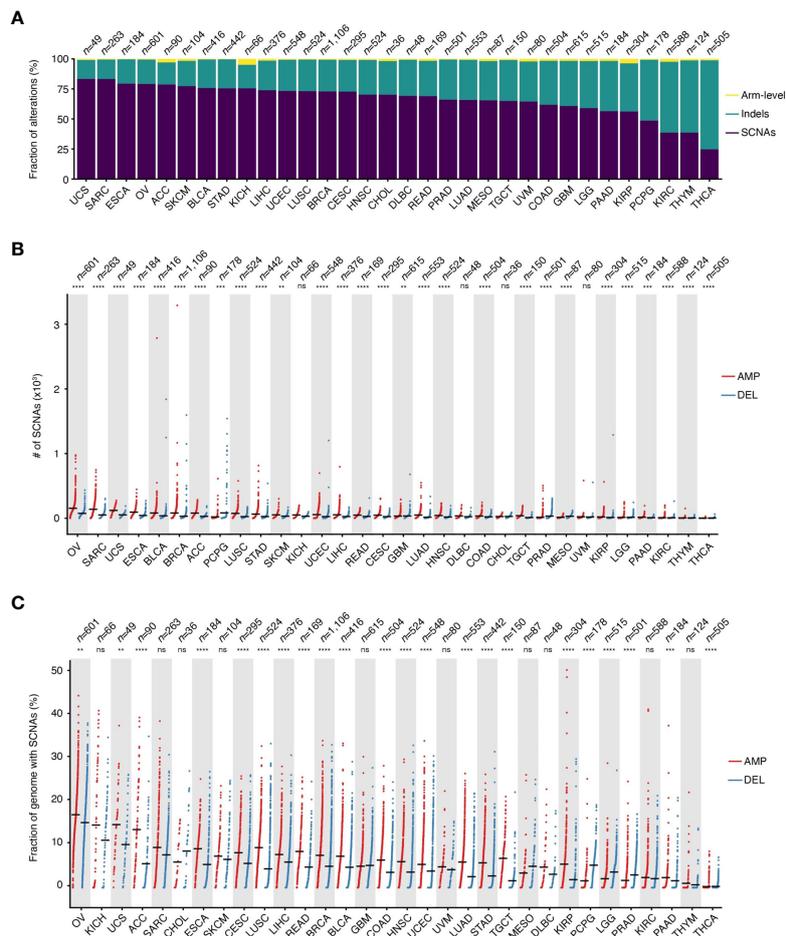
SCNAs defined as amplifications or deletions longer than 10 kb and spanning at most 75% of the corresponding chromosome arm represent the main type of alteration in 28 out of 32 (87.5%) tumor types for which copy number data are available in TCGA (**Figure 1A**). Indels represent the second most frequent type of alteration in these tumors, while they prevail over SCNAs in thyroid carcinomas (THCA), thymomas (THYM), kidney renal clear cell carcinomas (KIRC), and pheochromocytomas and paragangliomas (PCPG). SCNAs account for more than 80% of all copy number changes detected in uterine corpus endometrial carcinomas (UCEC) and in sarcomas (SARC). When examining the total number of SCNAs and the percentage of the genome affected by them, ovarian carcinomas (OV) and SARC carry, on average, the highest burden of SCNAs, followed by uterine carcinosarcoma (UCS), esophageal carcinomas (ESCA), bladder urothelial carcinomas (BLCA), and breast invasive carcinomas (BRCA) (**Figures 1B, C**). There is considerable heterogeneity in the number of SCNA events per sample within the same tumor type, with OV, SARC, BRCA, and PCPG showing the highest variability (**Figure 1B**). Similarly, the percentage of the genome altered varies extensively from tumor to tumor within the same type, with some tumors having more than 30% of the genome either amplified or deleted (**Figure 1C**). In most cancer types, amplifications are significantly more abundant and affect a larger fraction of the genome compared to deletions (**Figures 1B, C**). However, in the same tumor type, the number of amplifications is, on average, correlated with the number of deletions (**Supplementary Figure 1A**). In multiple tumor types, the number of amplifications and deletions is also significantly correlated within the same sample (**Supplementary Figure 1B**), suggesting that in these tumors both SCNA types are driven by common mechanism(s).

## What Is the Average Size of SCNAs?

Across the 32 tumor types examined, the mean SCNA length is  $3.83 \pm 7.89$  megabases (Mb, mean  $\pm$  s.d.), with deletions spanning longer distances on average (mean length: 4.69 Mb vs. 3.38 Mb for deletions and amplifications, respectively). The trend, however, is highly dependent on the tumor type (**Figure 2A**). Some chromosomes tend to harbor longer SCNAs: for example, chromosome (chr) 13 and 15 have the longest SCNAs in uveal melanomas (UVM), whereas chr1 and chr2 carry the longest SCNAs in KIRC and kidney renal papillary cell carcinomas (KIRP) (**Supplementary Figure 2A**). Interestingly, after normalizing by chromosome length, the group including chr13–22 shows proportionally larger SCNAs compared to chr1–12 and chrX (**Supplementary Figure 2B**). Although the basis of these differences in SCNA length is unknown, it might be related to variable chromatin composition and localization of these two groups of chromosomes within the cell nucleus in different cell types.

## Which Chromosomes Are Most Frequently Affected by SCNAs?

In addition to certain chromosomes harboring longer or shorter SCNAs, there is also a clear propensity for SCNAs to occur within certain chromosomes, even when accounting for



**FIGURE 1** | Prevalence of SCNAs across different tumor types. **(A)** Percentage of arm-level aneuploidies, indels and SCNAs by tumor type. **(B)** Number of SCNAs by tumor type, separately for amplifications (AMP) and deletions (DEL). Each dot represents one tumor sample. Black horizontal bars: mean values. **(C)** Same as in **(B)**, asterisks indicate statistical significance levels (Wilcoxon’s test, two-sided) at 0.05 (\*), 0.01 (\*\*), 0.001 (\*\*\*) and 0.0001 (\*\*\*\*). ns, not statistically significant. See **Supplementary Table 4** for the list of tumor type abbreviations used.

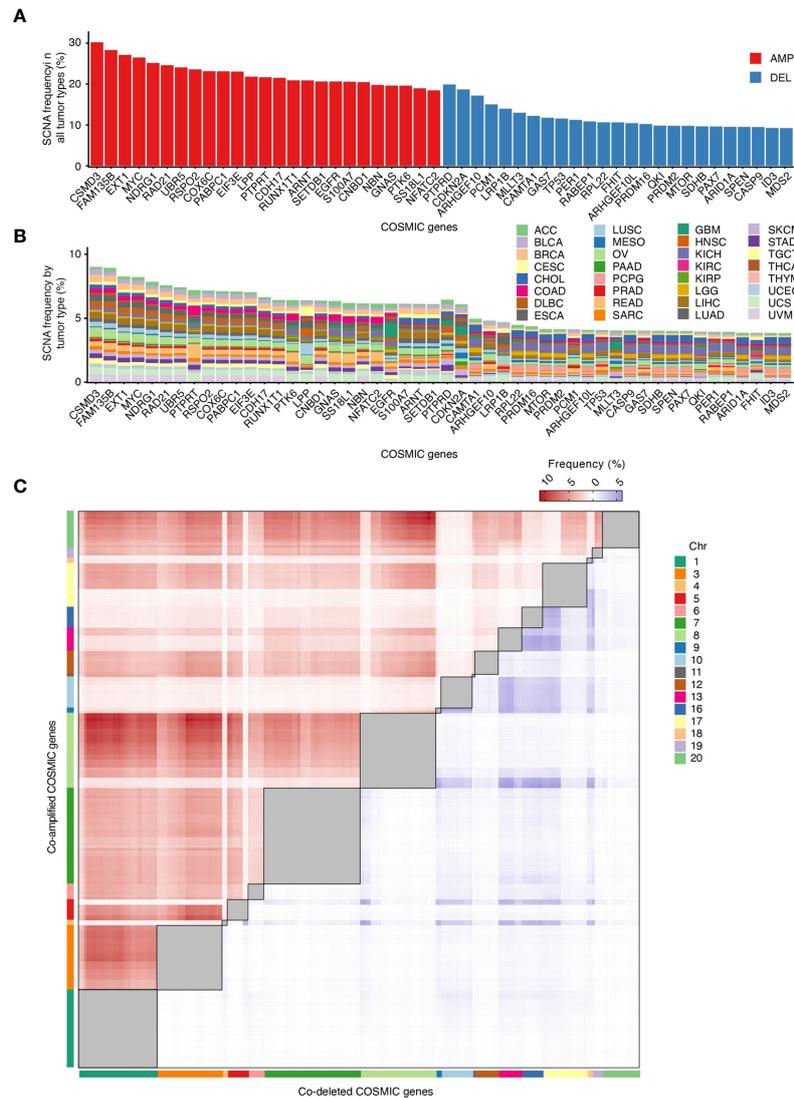
differences in chromosome length. For instance, a relatively large proportion of SCNA events are found within chr8 across many tumor types, whereas most of the SCNAs detected in adenocarcinomas of the colon (COAD) and rectum (READ) fall within chr13 and chr20, and most of the SCNAs in BRCA and cholangiocarcinomas (CHOL) affect chr17 while other chromosomes, such as chr15, are largely unaffected by SCNAs (**Figure 2B**). Notably, chr8 harbors the *MYC* gene, which is one of the most frequently amplified genes in human cancers (17), whereas chr17 contains the *ERBB2* gene, which is amplified in about 20% of all invasive breast cancers (18). However, the fact that some chromosomes or chromosomal regions appear to be more susceptible to SCNAs in human cancers might be independent of the fitness advantage imparted by the amplification or deletion of specific genes. For example, this could depend on differences in the amount of DNA damage and type or efficiency of DNA repair along individual chromosomes.

Indeed, using methods that probe the frequency of endogenous DNA double-strand breaks (DSBs) at high resolution along the genome, we and others have demonstrated that these lesions—which likely underpin the formation of many SCNAs—form non-randomly along the genome as a result of the interplay between DNA replication, transcription and 3D genome dynamics (19–25).

### Which Cancer Genes Are Most Frequently Amplified or Deleted?

In tumors, SCNAs are thought to confer a fitness advantage to the cells that harbor them, by leading to increased levels of expression of certain genes and decreased expression of others (26). We therefore examined which among the genes listed in the Catalogue Of Somatic Mutations in Cancer (COSMIC) (16) are most frequently altered by SCNAs. *CSMD3* and *FAM135B*,





**FIGURE 3 |** COSMIC genes most frequently affected by SCNAs. **(A)** Percentage of all tumor types in which the 25 most frequently amplified (red) and deleted (blue) COSMIC genes are altered. **(B)** Percentage of tumors in which each COSMIC gene shown in **(A)** is altered in each of the 32 tumor types analyzed. **(C)** Frequency of COSMIC gene pairs that are either co-amplified (red) or co-deleted (blue) in at least 5% of all the 10,729 tumors analyzed. Intra-chromosomal pairs, which are the majority, are masked (grey squares) to make the inter-chromosomal pairs visible. The full list of inter-chromosomal pairs (including AMP-DEL events) sorted by tumor type is available in **Supplementary Table 2**.

located on chr13, chr16, chr17 and chr18. Thirdly, a small number of cancer genes accounts for most co-amplification or co-deletion events. The *MYC* gene is involved in 31 out of the 100 most frequent AMP-AMP pairs across different chromosomes in the 10,729 tumors analyzed, followed by *FAM135B* (19%) and *NDRG1* (15%) (**Supplementary Tables 2, 3**). Among the top 100 DEL-DEL pairs, *ARHGEF10* is engaged in 31% of the co-deletion events, followed by *PCMI* (20%) and *CDKN2A* (13%) (**Supplementary Tables 2, 3**). These preferred co-amplification and co-deletion events likely depend on both the effect of the selection forces acting during tumor evolution—which might favor certain SCNA combinations over others—as well as on

the genomic context in which SCNAs arise. For example, the fact that AMP-AMP are more frequent on larger chromosomes and DEL-DEL on smaller ones might be related to differences in the radial placement and/or frequency of reciprocal contacts between these chromosomes inside the cell nucleus.

## DISCUSSION

Our analysis of publicly available TCGA data clearly shows that SCNAs are pervasive across human cancers and that certain tumor types have a considerable portion of their genome affected

by SCNAs. What remains unclear is how cancer associated SCNAs form in the first place, and how (if) they evolve during tumor progression. The molecular mechanisms of SCNA formation have been extensively studied in yeast and reviewed elsewhere (27). In contrast, very little is known about how SCNAs arise in mammalian cells and during tumorigenesis. This is not surprising, considering that the etiology of SCNAs in tumors remains unclear and it is very challenging to recapitulate them experimentally. In the following paragraphs we discuss the available scientific evidence that informs our current view of how SCNAs might form and evolve during tumorigenesis, by distinguishing three main sources of evidence: 1) integrative genomic analyses; 2) *in vitro* experiments; and 3) phylogenetic analyses based on multi-region or single-cell sequencing of pre-malignant lesions, primary tumors, and metastases. Importantly, the mechanisms of SCNA formation are different from those that cause aneuploidy, which have been reviewed elsewhere (2, 12) and will not be discussed here.

### Integrative Genomic Analyses

Integrative analyses of SCNAs and (epi)genomic features associated with DNA fragility and damage can suggest potential mechanisms by which SCNAs arise during tumorigenesis or help formulate hypotheses that can be tested experimentally. By intersecting genome-wide DSB maps with SCNA data available from the early stage of TCGA, we previously showed that cancer-associated deletions and, to a lesser extent, amplifications significantly overlap with genomic regions where DSBs accumulate upon exposure of cells to mild doses of the DNA polymerase inhibitor aphidicolin—a treatment that causes replication stress (19). Subsequently, a pan-cancer analysis intersected somatic focal ( $\leq 5$  Mb) deletions identified by TCGA in 10,221 tumor specimens spanning 30 cancer types with known common fragile sites (CFSs), to assess whether deletions occur more frequently at these sites than elsewhere in the genome (28). CFSs are megabase-sized genomic regions that tend to break frequently upon induction of replication stress and are cytogenetically detected as chromosome gaps or breaks in metaphase spread preparations (28). This analysis revealed that CFSs significantly overlap with the boundaries of focal deletions detected in cancer, suggesting that replication stress might represent the driving force of (at least some of) these alterations in human cancers (28). Although insightful, these analyses only provide correlative evidence and cannot pinpoint the exact causes and mechanisms of SCNA formation in cancer.

### In Vitro Experiments

One way to experimentally investigate the etiopathogenesis of SCNAs would be to expose normal or cancer cells to a variety of DNA damaging agents or genetic perturbations that favor the accumulation of DNA damage, followed by monitoring the emergence of SCNAs at the single-cell level at different timepoints. This approach has long been hindered by the lack of suitable methods for detecting SCNAs in single cells. However, in the past decade, several groups have begun exploring the formation of SCNAs *in vitro* using single-cell readouts. As mentioned above, replication stress causes the accumulation of DSBs at common fragile sites, which are enriched for breakpoints

of cancer-associated deletions (28). Prompted by this observation, pioneering experiments based on single-cell cloning and array comparative genomic hybridization (aCGH) showed that acute exposure of cells to DNA replication stressors, such as aphidicolin or hydroxyurea, results in rapid formation of SCNAs at CFSs *in vitro* (29, 30). Analysis of breakpoint junctions revealed that these SCNAs were characterized by short ( $< 6$  bp) microhomologies, suggesting that they formed through non-homologous end-joining (NHEJ) or another replication-coupled process, such as template switching (30). Subsequent integration of maps of SCNAs induced *in vitro* with gene expression profiles from the same cells revealed that long ( $> 1$  Mb) transcribed genes overlapping with CFSs represent hotspots of SCNA formation following replication stress (31). More recently, the development of single-cell whole genome sequencing (WGS) methods has enabled more explorations of the emergence of SCNAs upon exposure of cells to DNA damaging agents *in vitro*. By applying single-cell WGS to diploid cells exposed to aphidicolin throughout one cell division cycle, different types of replication stress-induced SCNAs were identified, including a group of large ( $> 20$  Mb) SCNAs associated with late DNA replication timing, low expression of nearby genes, and proximity to large genes, as well as a group of short ( $\leq 20$  Mb) amplifications with contrasting associations with various genomic features (32). Recently, another single-cell genome sequencing method named direct nuclear tagmentation and RNA sequencing (DNTR-seq) was used to profile SCNAs induced by exposing HT116 colon carcinoma cells to high doses of the topoisomerase II inhibitor, etoposide, or to ionizing radiation *in vitro* (33). This study revealed that single copy deletions were the most common type of SCNAs caused by etoposide treatment and ionizing radiation, and that the probability of a copy number change within a given genomic region increased with the distance of the region from the centromere (33). Unfortunately, the same study did not compare the landscape of these SCNAs with those emerging following replication stress. Therefore, it is not possible at this point to conclude whether different types of DNA damage result in different genomic landscapes of SCNAs, and whether the type of cell in which the damage occurs influences the type and genomic distribution of the SCNAs formed.

### Phylogenetic Analyses

In addition to understanding the causes and mechanisms of SCNA formation in tumors, a fundamental question is when SCNAs arise in tumorigenesis and if/how the genomic landscape of SCNAs changes during tumor evolution. The SCNAs that are detected in a tumor at any given timepoint during the history of the disease represent the result of the continuous interplay between the mutagenic processes that trigger their formation (genome instability) and the selection forces (either positive or negative) that act upon them based on their impact on cellular fitness. A glimpse into these complex evolutionary dynamics is offered by comparative genomic studies that have analyzed SCNAs in the same tumor type at different disease stages. Comparison of precursor lesions and invasive melanomas revealed that SCNAs were prevalent only in the latter (34), suggesting that SCNAs are not an early DNA damage event during melanoma formation.

In contrast, multi-region SCNA profiling in biopsies of Barret's esophagus—the precursor lesion of esophageal adenocarcinoma—revealed the presence of multiple clonal SCNAs already at this early stage (35). Notably, the SCNA formation rate inferred using a Bayesian phylogenetic analysis was low in these precursor lesions (approx. 0.005 events per year per locus per allele), suggesting that, once they have been acquired in a rapid 'burst', SCNAs remain relatively stable throughout tumor evolution (35). This 'punctuated evolution' model has also been used to explain the evolution of SCNAs in triple-negative breast cancers (36), colorectal cancers (37) and pancreatic cancers (38). On the other hand, a recent single-cell WGS study conducted on 8 triple-negative breast cancers and 4 cell lines concluded that SCNAs continue to form during the evolution of these tumors, even though the majority of SCNAs are acquired in a single 'burst' in the early stages of tumorigenesis (39). Continuous SCNA evolution has also been suggested by numerous studies comparing primary (P) and metastatic (M) tumors or different regions within P and M lesions. For example, a comparison of P and M samples from patients with castration resistant prostate cancer found that the SCNA burden was significantly higher in M compared to P samples (40). Another study, in which SCNAs were compared between paired P-M samples from 10 patients with advanced prostate cancer, revealed that the SCNA landscape likely continued to evolve both in the primary tumor and in the metastases after dissemination (41). Variable P-M divergence in SCNA patterns has also been reported for colon (42–44), breast (45–48), lung (49) and renal cancers (50). Similarly, numerous studies based on multi-region tumor sequencing have consistently shown that the genomic pattern of SCNAs often differs between spatially distinct regions of the same tumor (51–54). In line with this, using a novel method that allows profiling SCNAs in multiple regions of individual formalin-fixed paraffin embedded tumor sections, we were able to show that SCNA genomic profiles can vary substantially between different regions of matched P-M breast cancer samples (55). Collectively, the results of these studies indicate that SCNAs likely originate in various ways in different tumor types and suggest that the genomic landscape of SCNAs is rather dynamic throughout tumor evolution, presumably due to the continuous interplay between genome instability and the selection forces at play. In this context, it is worth mentioning that many genes involved in the DNA damage response (DDR) and different types of DNA damage repair are frequently mutated in human cancers (56, 57), which likely contributes to fuel genomic instability and SCNA accumulation during tumor progression.

## CONCLUSION

It is by now clear that SCNAs constitute a pervasive and distinctive trait of human cancers, which carries independent prognostic information and likely reflects the level and type of genomic instability characteristic of each tumor. However, even though SCNAs have been profiled across thousands of tumor samples, we still know remarkably little about the mechanisms and processes that lead to their emergence during tumorigenesis, and how (if) dynamic these alterations are in tumor cells. This is mainly because, for ethical and practical reasons, it is

not possible to take multiple biopsies of the same tumor at different stages of its growth—which would allow monitoring how SCNAs form and evolve *in vivo*. However, the availability of tractable *in vitro* systems, such as tumor organoids (58), alongside with single-cell genome sequencing methods (59) and CRISPR technologies for genome editing (60) should enable such investigations in the near future. We propose that frequent sampling and single-cell genome sequencing could be used to trace the formation and evolution of SCNAs in cells grown for prolonged periods of time in the presence of genetic or environmental cues that might cause or facilitate the formation of SCNAs (*e.g.*, DNA replication or transcription stressors). Patient-derived tumor xenografts (PDX) (61) represent another powerful tool for studying if and how the genomic landscape of SCNAs changes during the course of tumor evolution. Indeed, such models are already providing precious insights into the evolutionary dynamics of actionable and resistance-conferring point mutations (62).

A largely unexplored question is how SCNAs affect the spatial arrangement of the genome and whether this contributes to changes in gene expression and ultimately cell fitness. It is plausible that large amplifications or deletions, as well as other types of genomic rearrangements, will result in repositioning of genes and regulatory regions in their immediate neighborhood or even in changes to the overall 3D architecture of the genome. For example, using a novel method that we recently developed to infer radial genomic positions in the cell nucleus, we were able to show that translocations lead to substantial changes in the radial placement of the two parts of the same chromosome which get separated by the translocation, that in turn leads to large groups of consecutive genes being placed in either more active or more repressed chromatin neighborhoods (63). In the future, it will be important to develop methods that can probe SCNAs together with the 3D genome conformation and gene expression in single cells, to test their interplay and gain a deeper understanding of the mechanisms by which SCNAs convey a fitness advantage to the tumor cells that carry them.

Finally, SCNAs might represent an unexplored 'Achille's heel' of tumors, which could serve as a point of attack for novel therapeutic strategies. We hypothesize that, in cells with highly rearranged genomes and a high burden of SCNAs, the 3D genome might be in an unstable equilibrium, which would render them vulnerable to pharmacological perturbation of factors responsible for maintaining a certain arrangement of chromatin in 3D. Additionally, cells carrying many SCNAs might be more susceptible to disruption of key DNA repair pathways or to further accumulation of SCNAs, for example by treatment with DNA replication and/or transcription stressors. Exploring these possibilities will likely open new therapeutic avenues and ideally lead to more effective treatments for metastatic cancers, which, tragically, remain largely incurable.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

*Conceptualization*, NC. *Literature summary*, MN, DP. *TCGA data analysis*, LH, FA. *Funding acquisition*, NC, MB. *Supervision*, NC. *Visualization*, LH, FA, MB, NC. *Writing*, NC with contributions from all the authors. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by funding from the H2020-MSCA-ITN-2018 Marie Skłodowska-Curie Actions Innovative Training Networks ('aDDRess', grant no. 812829) to LH; the Polish National Science Center (grant no. UMO- 2016/23/D/NZ2/02890) to MN; the Swedish Cancer Research Foundation (grant no. 19 0130 Pj 03 H) to MB; and the Swedish Cancer Research Foundation (grant no. CAN 2018/728), the Strategic Research

## REFERENCES

- Cortés-Ciriano I, Lee JJ-K, Xi R, Jain D, Jung YL, Yang L, et al. Comprehensive Analysis of Chromothripsis in 2,658 Human Cancers Using Whole-Genome Sequencing. *Nat Genet* (2020) 52(3):331–41. doi: 10.1038/s41588-019-0576-7
- Ben-David U, Amon A. Context Is Everything: Aneuploidy in Cancer. *Nat Rev Genet* (2020) 21(1):44–62. doi: 10.1038/s41576-019-0171-x
- Li Y, Roberts ND, Wala JA, Shapira O, Schumacher SE, Kumar K, et al. Patterns of Somatic Structural Variation in Human Cancer Genomes. *Nature* (2020) 578(7793):112–21. doi: 10.1038/s41586-019-1913-9
- Priestley P, Baber J, Lolkema MP, Steeghs N, de Bruijn E, Shale C, et al. Pan-Cancer Whole-Genome Analyses of Metastatic Solid Tumours. *Nature* (2019) 575(7781):210–6. doi: 10.1038/s41586-019-1689-y
- Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, et al. Pan-Cancer Patterns of Somatic Copy Number Alteration. *Nat Genet* (2013) 45(10):1134–40. doi: 10.1038/ng.2760
- Fudenberg G, Getz G, Meyerson M, Mirny LA. High Order Chromatin Architecture Shapes the Landscape of Chromosomal Alterations in Cancer. *Nat Biotechnol* (2011) 29(12):1109–13. doi: 10.1038/nbt.2049
- Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive Mapping of Long Range Interactions Reveals Folding Principles of the Human Genome. *Science* (2009) 326(5950):289–93. doi: 10.1126/science.1181369
- Akdemir KC, Le VT, Chandran S, Li Y, Verhaak RG, Beroukhim R, et al. Disruption of Chromatin Folding Domains by Somatic Genomic Rearrangements in Human Cancer. *Nat Genet* (2020) 52(3):294–305. doi: 10.1038/s41588-019-0564-y
- Szabo Q, Bantignies F, Cavalli G. Principles of Genome Folding Into Topologically Associating Domains. *Sci Adv* (2019) 5(4):eaaw1668. doi: 10.1126/sciadv.aaw1668
- van Steensel B, Belmont AS. Lamina-Associated Domains: Links With Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* (2017) 169(5):780–91. doi: 10.1016/j.cell.2017.04.022
- Weischenfeldt J, Dubash T, Drinas AP, Mardin BR, Chen Y, Stütz AM, et al. Pan-Cancer Analysis of Somatic Copy-Number Alterations Implicates IRS4 and IGF2 in Enhancer Hijacking. *Nat Genet* (2017) 49(1):65–74. doi: 10.1038/ng.3722
- Knouse KA, Davoli T, Elledge SJ, Amon A. Aneuploidy in Cancer: Seq-Ing Answers to Old Questions. *Annu Rev Cancer Biol* (2017) 1(1):335–54. doi: 10.1146/annurev-cancerbio-042616-072231
- Wang X, Steensma JT, Bailey MH, Feng Q, Padda H, Johnson KJ. Characteristics of The Cancer Genome Atlas Cases Relative to U.S. General Population Cancer Cases. *Br J Cancer* (2018) 119(7):885–92. doi: 10.1038/s41416-018-0140-8

Programme in Cancer (StratCan) at Karolinska Institutet (grant no. 2201) and the Swedish Foundation for Strategic Research (grant no. BD15\_0095) to NC.

## ACKNOWLEDGMENTS

We thank Britta Bouwman (Bienko-Crosetto lab) for critically reading the manuscript and for her help during the final revision.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2021.700568/full#supplementary-material>

- Wilson S, Fitzsimons M, Ferguson M, Heath A, Jensen M, Miller J, et al. Developing Cancer Informatics Applications and Tools Using the NCI Genomic Data Commons API. *Cancer Res* (2017) 77(21):e15–8. doi: 10.1158/0008-5472.CAN-17-0598
- Ramos M, Schiffer L, Davis S, Waldron L. TCGAutils: TCGA Utility Functions for Data Management. (2021) *R package version 1.12.0*. Available at: <https://bioconductor.org/packages/release/bioc/html/TCGAutils.html>.
- Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC Cancer Gene Census: Describing Genetic Dysfunction Across All Human Cancers. *Nat Rev Cancer* (2018) 18(11):696–705. doi: 10.1038/s41568-018-0060-1
- Schaub FX, Dhankani V, Berger AC, Trivedi M, Richardson AB, Shaw R, et al. Pan-Cancer Alterations of the MYC Oncogene and Its Proximal Network Across the Cancer Genome Atlas. *Cell Syst* (2018) 6(3):282–300.e2. doi: 10.1016/j.cels.2018.03.003
- Arteaga CL, Sliwkowski MX, Osborne CK, Perez EA, Puglisi F, Gianni L. Treatment of HER2-Positive Breast Cancer: Current Status and Future Perspectives. *Nat Rev Clin Oncol* (2011) 9(1):16–32. doi: 10.1038/nrclinonc.2011.177
- Crosetto N, Mitra A, Silva MJ, Bienko M, Dojer N, Wang Q, et al. Nucleotide-Resolution DNA Double-Strand Break Mapping by Next-Generation Sequencing. *Nat Methods* (2013) 10(4):361–5. doi: 10.1038/nmeth.2408
- Canela A, Sridharan S, Sciascia N, Tubbs A, Meltzer P, Sleckman BP, et al. DNA Breaks and End Resection Measured Genome-Wide by End Sequencing. *Mol Cell* (2016) 63(5):898–911. doi: 10.1016/j.molcel.2016.06.034
- Yan WX, Mirzazadeh R, Garnerone S, Scott D, Schneider MW, Kallas T, et al. BLISS Is a Versatile and Quantitative Method for Genome-Wide Profiling of DNA Double-Strand Breaks. *Nat Commun* (2017) 12:8:15058. doi: 10.1038/ncomms15058
- Bouwman BAM, Crosetto N. Endogenous DNA Double-Strand Breaks During DNA Transactions: Emerging Insights and Methods for Genome-Wide Profiling. *Genes (Basel)* (2018) 9(12):632–63. doi: 10.3390/genes9120632
- Gothe HJ, Bouwman BAM, Gusmao EG, Piccinno R, Petrosino G, Sayols S, et al. Spatial Chromosome Folding and Active Transcription Drive DNA Fragility and Formation of Oncogenic MLL Translocations. *Mol Cell* (2019) 75(2):267–83. doi: 10.1016/j.molcel.2019.05.015
- Dellino GI, Palluzzi F, Chiariello AM, Piccioni R, Bianco S, Furia L, et al. Release of Paused RNA Polymerase II at Specific Loci Favors DNA Double-Strand-Break Formation and Promotes Cancer Translocations. *Nat Genet* (2019) 51(6):1011–23. doi: 10.1038/s41588-019-0421-z
- Bouwman BAM, Agostini F, Garnerone S, Petrosino G, Gothe HJ, Sayols S, et al. Genome-Wide Detection of DNA Double-Strand Breaks by in-suspension BLISS. *Nat Protoc* (2020) 15(12):3894–941. doi: 10.1038/s41596-020-0397-2

26. Santaguida S, Amon A. Short- and Long-Term Effects of Chromosome Mis-Segregation and Aneuploidy. *Nat Rev Mol Cell Biol* (2015) 16(8):473–85. doi: 10.1038/nrm4025
27. Hastings PJ, Lupski JR, Rosenberg SM, Ira G. Mechanisms of Change in Gene Copy Number. *Nat Rev Genet* (2009) 10(8):551–64. doi: 10.1038/nrg2593
28. Glover TW, Wilson TE, Arlt MF. Fragile Sites in Cancer: More Than Meets the Eye. *Nat Rev Cancer* (2017) 17(8):489–501. doi: 10.1038/nrc.2017.52
29. Arlt MF, Mulle JG, Schaibley VM, Ragland RL, Durkin SG, Warren ST, et al. Replication Stress Induces Genome-Wide Copy Number Changes in Human Cells That Resemble Polymorphic and Pathogenic Variants. *Am J Hum Genet* (2009) 84(3):339–50. doi: 10.1016/j.ajhg.2009.01.024
30. Arlt MF, Ozdemir AC, Birkeland SR, Wilson TE, Glover TW. Hydroxyurea Induces De Novo Copy Number Variants in Human Cells. *Proc Natl Acad Sci USA* (2011) 108(42):17360–5. doi: 10.1073/pnas.1109272108
31. Wilson TE, Arlt MF, Park SH, Rajendran S, Paulsen M, Ljungman M, et al. Large Transcription Units Unify Copy Number Variants and Common Fragile Sites Arising Under Replication Stress. *Genome Res* (2015) 25(2):189–200. doi: 10.1101/gr.177121.114
32. Mazzagatti A, Shaikh N, Bakker B, Spierings DCJ, Wardenaar R, Maniati E, et al. DNA Replication Stress Generates Distinctive Landscapes of DNA Copy Number Alterations and Chromosome Scale Losses. *bioRxiv* (2020) 743658. doi: 10.1101/743658
33. Zachariadis V, Cheng H, Andrews N, Enge M. A Highly Scalable Method for Joint Whole-Genome Sequencing and Gene-Expression Profiling of Single Cells. *Mol Cell* (2020) 80:541–53. doi: 10.1101/2020.03.04.976530
34. Shain AH, Yeh I, Kovalyshyn I, Sriharan A, Talevich E, Gagnon A, et al. The Genetic Evolution of Melanoma From Precursor Lesions. *N Engl J Med* (2015) 373(20):1926–36. doi: 10.1056/NEJMoa1502583
35. Martinez P, Mallo D, Paulson TG, Li X, Sanchez CA, Reid BJ, et al. Evolution of Barrett's Esophagus Through Space and Time at Single-Crypt and Whole-Biopsy Levels. *Nat Commun* (2018) 9(1):794. doi: 10.1038/s41467-017-02621-x
36. Gao R, Davis A, McDonald TO, Sei E, Shi X, Wang Y, et al. Punctuated Copy Number Evolution and Clonal Stasis in Triple-Negative Breast Cancer. *Nat Genet* (2016) 48(10):1119–30. doi: 10.1038/ng.3641
37. Sottoriva A, Kang H, Ma Z, Graham TA, Salomon MP, Zhao J, et al. A Big Bang Model of Human Colorectal Tumor Growth. *Nat Genet* (2015) 47(3):209–16. doi: 10.1038/ng.3214
38. Notta F, Chan-Seng-Yue M, Lemire M, Li Y, Wilson GW, Connor AA, et al. A Renewed Model of Pancreatic Cancer Evolution Based on Genomic Rearrangement Patterns. *Nature* (2016) 538(7625):378–82. doi: 10.1038/nature19823
39. Minussi DC, Nicholson MD, Ye H, Davis A, Wang K, Baker T, et al. Breast Tumours Maintain a Reservoir of Subclonal Diversity During Expansion. *Nature* (2021) 592(7853):302–8. doi: 10.1038/s41586-021-03357-x
40. van Dessel LF, van Riet J, Smits M, Zhu Y, Hamberg P, van der Heijden MS, et al. The Genomic Landscape of Metastatic Castration-Resistant Prostate Cancers Reveals Multiple Distinct Genotypes With Potential Clinical Impact. *Nat Commun* (2019) 10(1):5251. doi: 10.1038/s41467-019-13084-7
41. Gudem G, Van Loo P, Kremeyer B, Alexandrov LB, Tubio JMC, Papaemmanuil E, et al. The Evolutionary History of Lethal Metastatic Prostate Cancer. *Nature* (2015) 520(7547):353–7. doi: 10.1038/nature14347
42. Lee SY, Haq F, Kim D, Jun C, Jo H-J, Ahn S-M, et al. Comparative Genomic Analysis of Primary and Synchronous Metastatic Colorectal Cancers. *PLoS One* (2014) 9(3):e90459. doi: 10.1371/journal.pone.0090459
43. Mamlouk S, Childs LH, Aust D, Heim D, Melching F, Oliveira C, et al. DNA Copy Number Changes Define Spatial Patterns of Heterogeneity in Colorectal Cancer. *Nat Commun* (2017) 8:14093. doi: 10.1038/ncomms14093
44. Ma Y-S, Huang T, Zhong X-M, Zhang H-W, Cong X-L, Xu H, et al. Proteogenomic Characterization and Comprehensive Integrative Genomic Analysis of Human Colorectal Cancer Liver Metastasis. *Mol Cancer* (2018) 17(1):139. doi: 10.1186/s12943-018-0890-1
45. Kjällquist U, Erlandsson R, Tobin NP, Alkodsai A, Ullah I, Stålhammar G, et al. Exome Sequencing of Primary Breast Cancers With Paired Metastatic Lesions Reveals Metastasis-Enriched Mutations in the A-Kinase Anchoring Protein Family (AKAPs). *BMC Cancer* (2018) 18(1):174. doi: 10.1186/s12885-018-4021-6
46. Tang M-HE, Dahlgren M, Brueffer C, Tjitrowirjo T, Winter C, Chen Y, et al. Remarkable Similarities of Chromosomal Rearrangements Between Primary Human Breast Cancers and Matched Distant Metastases as Revealed by Whole-Genome Sequencing. *Oncotarget* (2015) 6(35):37169–84. doi: 10.18632/oncotarget.5951
47. Schrijver WAME, Selenica P, Lee JY, Ng CKY, Burke KA, Piscuoglio S, et al. Mutation Profiling of Key Cancer Genes in Primary Breast Cancers and Their Distant Metastases. *Cancer Res* (2018) 78(12):3112–21. doi: 10.1158/0008-5472.CAN-17-2310
48. Ng CKY, Bidard F-C, Piscuoglio S, Geyer FC, Lim RS, de Bruijn I, et al. Genetic Heterogeneity in Therapy-Naïve Synchronous Primary Breast Cancers and Their Metastases. *Clin Cancer Res* (2017) 23(15):4402–15. doi: 10.1158/1078-0432.CCR-16-3115
49. Shih DJH, Nayyar N, Bihun I, Dagogo-Jack I, Gill CM, Aquilanti E, et al. Genomic Characterization of Human Brain Metastases Identifies Drivers of Metastatic Lung Adenocarcinoma. *Nat Genet* (2020) 52(4):371–7. doi: 10.1038/s41588-020-0592-7
50. Becerra MF, Reznik E, Redzematovic A, Tennenbaum DM, Kashan M, Ghanaat M, et al. Comparative Genomic Profiling of Matched Primary and Metastatic Tumors in Renal Cell Carcinoma. *Eur Urol Focus* (2018) 4(6):986–94. doi: 10.1016/j.juro.2017.02.1116
51. Gerlinger M, Horswell S, Larkin J, Rowan AJ, Salm MP, Varela I, et al. Genomic Architecture and Evolution of Clear Cell Renal Cell Carcinomas Defined by Multiregion Sequencing. *Nat Genet* (2014) 46(3):225–33. doi: 10.1038/ng.2891
52. de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, Yates L, et al. Spatial and Temporal Diversity in Genomic Instability Processes Defines Lung Cancer Evolution. *Science* (2014) 346(6206):251–6. doi: 10.1126/science.1253462
53. Sottoriva A, Spiteri I, Piccirillo SGM, Touloumis A, Collins VP, Marioni JC, et al. Intratumor Heterogeneity in Human Glioblastoma Reflects Cancer Evolutionary Dynamics. *Proc Natl Acad Sci USA* (2013) 110(10):4009–14. doi: 10.1073/pnas.1219747110
54. Yates LR, Gerstung M, Knappskog S, Desmedt C, Gundem G, Van Loo P, et al. Subclonal Diversification of Primary Breast Cancer Revealed by Multiregion Sequencing. *Nat Med* (2015) 21(7):751–9. doi: 10.1038/nm.3886
55. Zhang X, Garnerone S, Simonetti M, Harbers L, Nicos M, Mirzazadeh R, et al. CUTseq Is a Versatile Method for Preparing Multiplexed DNA Sequencing Libraries From Low-Input Samples. *Nat Commun* (2019) 10(1):4732. doi: 10.1038/s41467-019-12570-2
56. Chae YK, Anker JF, Carneiro BA, Chandra S, Kaplan J, Kalyan A, et al. Genomic Landscape of DNA Repair Genes in Cancer. *Oncotarget* (2016) 7(17):23312–21. doi: 10.18632/oncotarget.8196
57. Knijnenburg TA, Wang L, Zimmermann MT, Chambwe N, Gao GF, Cherniack AD, et al. Genomic and Molecular Landscape of DNA Damage Repair Deficiency Across The Cancer Genome Atlas. *Cell Rep* (2018) 23(1):239–54.e6. doi: 10.1016/j.celrep.2018.03.076
58. Drost J, Clevers H. Organoids in Cancer Research. *Nat Rev Cancer* (2018) 18(7):407–18. doi: 10.1038/s41568-018-0007-6
59. Gawad C, Koh W, Quake SR. Single-Cell Genome Sequencing: Current State of the Science. *Nat Rev Genet* (2016) 17(3):175–88. doi: 10.1038/nrg.2015.16
60. Nakamura M, Gao Y, Dominguez AA, Qi LS. CRISPR Technologies for Precise Epigenome Editing. *Nat Cell Biol* (2021) 23(1):11–22. doi: 10.1038/s41556-020-00620-7
61. Hidalgo M, Amant F, Biankin AV, Budinská E, Byrne AT, Caldas C, et al. Patient-Derived Xenograft Models: An Emerging Platform for Translational Cancer Research. *Cancer Discov* (2014) 4(9):998–1013. doi: 10.1158/2159-8290.CD-14-0001
62. Eirew P, Steif A, Khattra J, Ha G, Yap D, Farahani H, et al. Dynamics of Genomic Clones in Breast Cancer Patient Xenografts at Single-Cell Resolution. *Nature* (2015) 518(7539):422–6. doi: 10.1038/nature13952
63. Girelli G, Custodio J, Kallas T, Agostini F, Wernersson E, Spanjaard B, et al. GPSeq Reveals the Radial Organization of Chromatin in the Cell Nucleus. *Nat Biotechnol* (2020) 38:1184–93. doi: 10.21203/rs.3.pex-570/v1

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in

this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Harbers, Agostini, Nicos, Poddighe, Bienko and Crosetto. This is an open-access article distributed under the terms of the Creative Commons Attribution

License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.