Current Genetics
https://doi.org/10.1007/s00294-019-00980-0

MINI-REVIEW

An evolving view of copy number variants

Stephanie Lauer[1] · **David Gresham**[2]

Received: 1 March 2019 / Revised: 17 April 2019 / Accepted: 20 April 2019

Abstract
Copy number variants (CNVs) are regions of the genome that vary in integer copy number. CNVs, which comprise both amplifications and deletions of DNA sequence, have been identified across all domains of life, from bacteria and archaea to plants and animals. CNVs are an important source of genetic diversity, and can drive rapid adaptive evolution and progression of heritable and somatic human diseases, such as cancer. However, despite their evolutionary importance and clinical relevance, CNVs remain understudied compared to single-nucleotide variants (SNVs). This is a consequence of the inherent difficulties in detecting CNVs at low-to-intermediate frequencies in heterogeneous populations of cells. Here, we discuss molecular methods used to detect CNVs, the limitations associated with using these techniques, and the application of new and emerging technologies that present solutions to these challenges. The goal of this short review and perspective is to highlight aspects of CNV biology that are understudied and define avenues for further research that address specific gaps in our knowledge of these complex alleles. We describe our recently developed method for CNV detection in which a fluorescent gene functions as a single-cell CNV reporter and present key findings from our evolution experiments in *Saccharomyces cerevisiae*. Using a CNV reporter, we found that CNVs are generated at a high rate and undergo selection with predictable dynamics across independently evolving replicate populations. Many CNVs appear to be generated through DNA replication-based processes that are mediated by the presence of short, interrupted, inverted-repeat sequences. Our results have important implications for the role of CNVs in evolutionary processes and the molecular mechanisms that underlie CNV formation. We discuss the possible extension of our method to other applications, including tracking the dynamics of CNVs in models of human tumors.

Keywords Copy number variation · Evolution · Chemostat · Cancer · Gene amplification · Gene deletion

Introduction

Variation in DNA copy number has been appreciated since the earliest days of molecular genetics. The first gene duplications and deletions were characterized early in the 1900s (Sturtevant 1925; Bridges 1936; Taylor and Raes 2004) and later studies revealed the extent of gene family expansions, whole-genome duplications, and polyploidy in natural populations (Avise and Kitto 1973; Hopkinson et al. 1976; Ferris and Whitt 1979; Stuber and Goodman 1983; Schughart et al. 1989). Susumo Ohno and others (Ohno 1970; Anderson and Roth 1977; Hughes 1994) demonstrated that gene duplication is important for generating evolutionary novelty and diversity over long time scales (Walsh 1995, 2003; Lynch and Force 2000; Lynch and Conery 2000; Conant and Wolfe 2008), but CNV formation can also drive rapid adaptation in response to stress and changes in the environment (Arlt et al. 2011, 2014; Hong and Gresham 2014; Greenblum et al. 2015; Dulmage et al. 2018; Bussotti et al. 2018). CNVs provide a substrate for natural selection in diverse scenarios ranging from niche adaptation to domestication and speciation (Stratton et al. 2009; Geiger et al. 2010; Clop et al. 2012; Ramirez et al. 2014; Żmieńko et al. 2014; Dhami et al. 2016; Turner et al. 2017; Zuellig and Sweigart 2018). Despite their clear importance, we do not fully understand the causes and consequences of CNV formation at the molecular level. In this short review and perspective

Communicated by M. Kupiec.

✉ David Gresham
dgresham@nyu.edu

[1] Institute for Systems Genetics, New York University Langone Health, New York, NY, USA

[2] Center for Genomics and System Biology, Department of Biology, New York University, New York, NY, USA

Published online: 10 May 2019

piece, we discuss existing CNV detection methods and the limitations of these methods for studying large, complex CNVs compared to single-nucleotide variants (SNVs). New technologies have the potential to further our understanding of the dynamics with which CNVs are generated, selected, and maintained, as well as the mechanistic processes that underlie CNV formation. We also summarise the findings from our recent study using a phenotypic reporter to detect and isolate single cells with CNVs. Our results have broad implications for the significant role of CNVs in driving both adaptive evolution and disease.

Challenges to studying the role of CNVs in adaptive evolution

Determining the functional consequences of a CNV is challenging. While SNVs encompass a single base pair, and indels can include several base pairs of DNA, CNVs range in size from 10^2 to 10^6 base pairs. As a result, CNVs are frequently alleles of large effect that can simultaneously impact multiple protein-coding genes and regulatory regions. CNVs can lead to changes in organismal fitness through a variety of mechanisms including: increases or decreases in dosage for genes within the CNV allele as well as neighboring genes (Molina et al. 2008; Merla et al. 2006; Gamazon et al. 2011), global changes to the transcriptome (Henrichsen et al. 2009), re-organization of chromatin domains (Lupiáñez et al. 2015; Lupiáñez et al. 2016; Franke et al. 2016; Spielmann et al. 2018), promoter capture and other modifications to the regulatory landscape (Koszul et al. 2004; Blount et al. 2008; Chan et al. 2010; Blount et al. 2012), and formation of chimeric genes (Arguello et al. 2006; Rippey et al. 2013; Aigner et al. 2013; Schrider et al. 2013; Mayo et al. 2017). Because CNVs can impact cell physiology at multiple levels, predicting how a CNV will affect the properties of a cell can be difficult. Because CNVs can encompass such large regions of the genome, they are likely to have pleiotropic effects (Gamazon and Stranger 2015). Disentangling the role of CNVs in evolution can be particularly challenging, especially in natural populations in which genetic manipulations cannot readily be performed.

CNVs can be detected using a variety of molecular and imaging methods. Cytogenetics, microscopy, and gel electrophoresis were among the first methods used to study CNVs (reviewed in Taylor and Raes 2004). Southern blotting and quantitative PCR (qPCR) can also be used to detect increases and deletions of specific DNA sequences (Gonzalez et al. 2005; Aitman et al. 2006). Global comparative genomics techniques such as array comparative genomic hybridization (aCGH) vastly improved genome-wide CNV detection (Sebat et al. 2004; LeCaignec et al. 2006; Michels et al. 2007; Konings et al. 2012; Hong and Gresham 2014; Payen et al. 2014). The development of next-generation sequencing technologies launched a new era of CNV discovery, enabling the detection of smaller and more complex CNVs in the genomes of diverse organisms (Greenblum et al. 2015; Hartmann and Croll 2017; Pham et al. 2017; Zhou et al. 2018; Yang et al. 2018; Wang et al. 2019; Rigau et al. 2019). One of the first comprehensive CNV maps of the human genome, generated using aCGH, uncovered 1447 copy number variable regions across 360 megabase pairs of the genome (Redon et al. 2006). Nearly 10 years later, a data set that collated 23 studies (the majority of which used next-generation sequencing) identified 202,431 copy number variable regions (Zarrei et al. 2015). Improved detection by genome sequencing has revealed how prevalent CNVs are in nature, and that de novo CNVs affect more base pairs of human DNA each generation than point mutations (Itsara et al. 2010).

A variety of approaches have been developed to detect CNVs from short-read DNA sequencing data. The number of reads that map to each nucleotide position (often referred to as the read depth) can be compared across regions of the genome. Those regions with increased or decreased read depth are used to infer the presence of duplications and deletions, respectively. Other methods use the existence of unique classes of sequence reads to map novel junctions created at the breakpoints of CNVs. Split reads are single reads that map to two distinct locations in the reference genome, whereas discordant reads are paired-end reads that map in the improper orientation or with an atypical distance in the reference genome. The accuracy of these methods is limited by both the quality and depth of sequencing reads. Furthermore, duplications are repetitive by nature and CNV breakpoints are often located within repetitive sequences or transposable elements, such as *Alu* elements (Chen et al. 2014; Gu et al. 2015; Kim et al. 2016). The repetitive nature of these features can make detection of CNVs using reference-based alignment extremely challenging. Detection of breakpoints that correspond to the novel sequence generated by CNVs is frequently impossible in these regions. Long-read sequencing technologies are poised to overcome the inherent challenges in using short reads, especially when mapping to repetitive regions (Huddleston et al. 2017; Chakraborty et al. 2018). However, there are still limitations with these technologies (Couldrey et al. 2017) and the low throughput and comparatively high cost of long-read sequencing technologies remains prohibitive for high-resolution, real-time analysis of CNV dynamics.

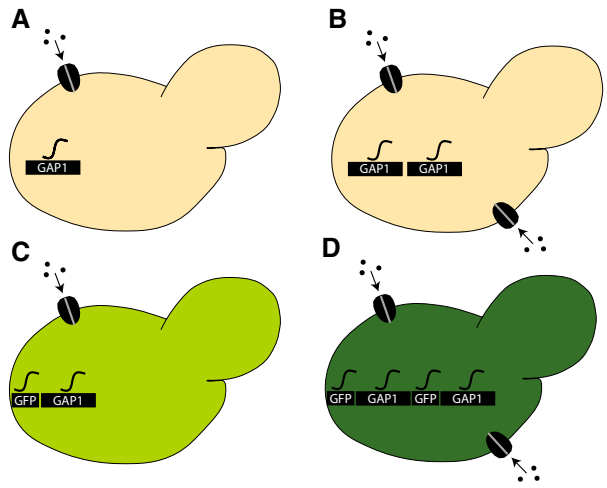

Fig. 1 Use of a fluorescent protein gene as a CNV reporter. **a** *GAP1* encodes the high-affinity general amino acid permease, which is responsible for uptake of amino acids from the environment. **b** *GAP1* duplications lead to an increased abundance of transporters on the cell membrane, providing a selective advantage to cells growing in glutamine-limited environments (Gresham et al. 2010; Hong and Gresham 2014). **c** Insertion of a green fluorescent protein (GFP) gene immediately adjacent to *GAP1* provides a phenotypic reporter for *GAP1* CNVs. **d** Co-duplication of *GAP1* and GFP results in an increase in cellular fluorescence that can be measured using flow cytometry

experiments performed using microbes. Studies, since the 1960s, have reported the occurrence of CNVs in populations of microbes evolving in nutrient-limited chemostats (Horiuchi et al. 1963; Hansche 1975; Sonti and Roth 1989; Brown et al. 1998; Gresham et al. 2008). However, the limitations of existing methods for detecting CNVs in heterogeneous populations have impeded progress in understanding the diversity and dynamics of de novo CNVs in the context of these experiments. To overcome the challenges in detecting and isolating cells with CNVs in evolving populations, we developed a phenotypic reporter that relies on duplication or deletion of a constitutively expressed green fluorescent protein (GFP) gene (Fig. 1). The GFP reporter is integrated adjacent to a gene of interest and simultaneous duplication of the two genes results in a quantitative increase in cellular fluorescence (Fig. 1). Simultaneous deletion of the target gene and the reporter results in a decrease in fluorescent signal. Cellular fluorescence can be measured every few generations during experimental evolution using flow cytometry, providing a high-resolution and real-time view of CNV dynamics. An additional advantage of this approach is that single cells with CNVs can be efficiently isolated from heterogeneous populations using fluorescence-activated cell sorting (FACS) (Fig. 2).

We used the CNV reporter to investigate the role of CNVs in mediating adaptation to nitrogen poor conditions

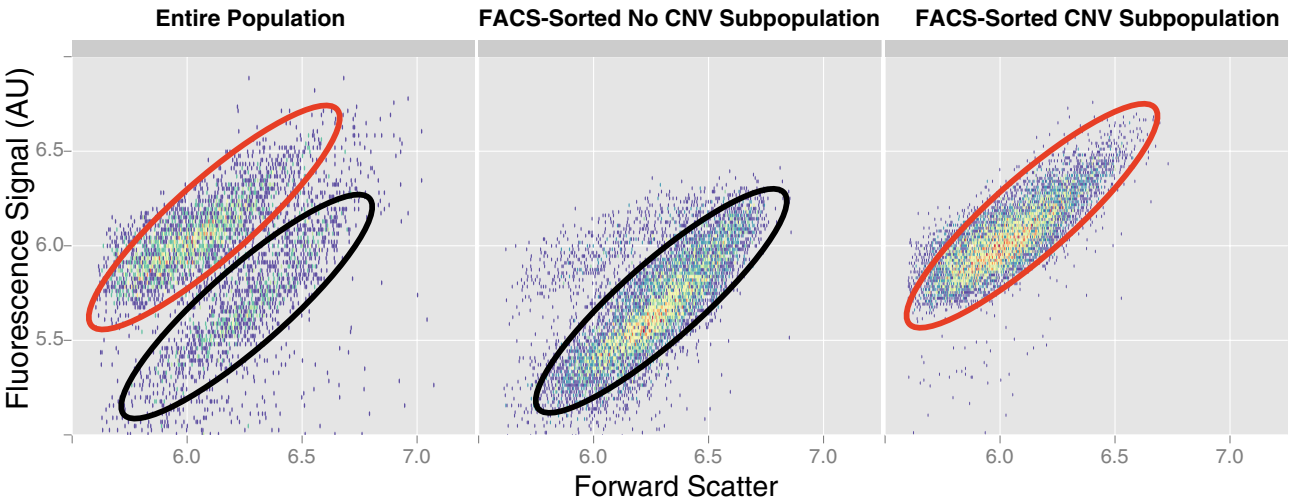

Fig. 2 Fluorescence-activated cell sorting (FACS) can be used to isolate cells with CNVs. Single cells from a heterogeneous population are sorted after defining one (black) and two copy (red) gates based on fluorescent signals from control populations. Whole-genome sequencing and other downstream applications can be performed using the fractionated CNV subpopulation. Forward scatter is proportional to measurements of cell size

A phenotypic reporter for CNV detection

We were motivated to develop a CNV reporter by the observation that CNVs are frequently selected during evolution in *Saccharomyces cerevisiae* under two selective conditions: glutamine limitation and urea limitation (Lauer et al. 2018). We performed long-term experimental evolution in chemostats and tracked gene duplications and deletions at

the *GAP1* locus. *GAP1* encodes the general amino acid permease, which is responsible for transporting amino acids such as glutamine into the cell (Fig. 1). Previous studies indicated that CNVs occur at the *GAP1* locus under these conditions, but these experiments relied on sampling clonal isolates from late in the history of the population, and thus, the dynamics were unknown (Gresham et al. 2008, 2010; Hong and Gresham 2014). For all nine experimental populations evolving in the presence of glutamine limitation, we observed increases in fluorescence, indicating that *GAP1* duplications occurred predictably and repeatedly. These results demonstrate that *GAP1* duplications are beneficial and strongly selected in this condition. We also detected a complete loss of fluorescent signal in one of nine urea-limited populations, consistent with the occurrence and selection of a *GAP1* deletion. *GAP1* CNVs identified using flow cytometry were independently confirmed by whole-genome sequencing as well as pulsed-field gel electrophoresis and Southern blot analysis. Our results demonstrate that the CNV reporter successfully and accurately measures a variety of copy number and structural variants including duplications, deletions, translocations, aneuploidies, and large segmental aneuploidies. Furthermore, the fluorescent signal measured by flow cytometry correlates with the number of copies of *GAP1*, providing comparable resolution to molecular methods such as qPCR or genome sequencing.

Although our CNV reporter was informative only about CNVs at the *GAP1* locus, we also identified structural variants and CNVs at other genomic loci using whole-genome sequencing data. Interestingly, the urea permease, *DUR3*, was duplicated in more than half of the urea-limited populations. Further characterization of CNV alleles revealed that *DUR3* was always present in five copies (Fig. 3a). Compared to *GAP1*, which was present at 2–4 copies and had a wider variety of CNV sizes and structures, *DUR3* CNVs were only ~ 25 kilobase pairs in length (Fig. 3b).

We further characterized CNVs at both the *GAP1* and *DUR3* loci using short-read sequencing data to define the breakpoints. We found that pairs of short, interrupted, inverted-repeat sequences were present at the breakpoints for all identified *DUR3* CNVs. These short inverted repeats, which have a median length of 8 nucleotides, are indicative of DNA replication-based mechanisms of CNV formation such as origin-dependent inverted-repeat amplification (ODIRA) (Brewer et al. 2011, 2015) or microhomology-mediated break-induced replication (MMBIR) (Lee et al. 2007; Hastings et al. 2009). Approximately half of the resolvable *GAP1* CNV breakpoints also contained pairs of inverted-repeat sequences, indicating that replication-mediated CNV formation may be an underappreciated source of de novo CNVs in *S. cerevisiae*.

The discrepancy in copy number and CNV allele size between *GAP1* and *DUR3* is surprising given that

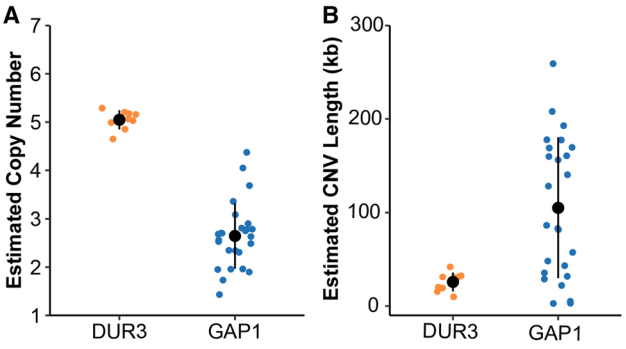

Fig. 3 *GAP1* and *DUR3* CNVs differ in copy number and length. *DUR3* CNVs were characterized from a total of 9 clones, isolated from 6 independently evolving populations. *GAP1* CNVs were characterized from a total of 25 clones, isolated from 11 independently evolving populations (9 experimental populations with the *GAP1* CNV reporter and 2 additional control populations with the CNV reporter at a neutral locus). Clones were isolated from generation 150 and generation 250 of the evolution experiment. **a** *DUR3* CNVs are present at significantly higher copy number than *GAP1* CNVs (*t* test, p value < 0.01). **b** *DUR3* CNVs are significantly smaller than *GAP1* CNVs (*t* test, p value < 0.01)

replication-based mechanisms are likely to be involved in generating both types of CNVs. This suggests that there are fundamental differences in the generation and selection of CNV alleles at these two loci and more broadly across other regions of the genome. *GAP1* is located on the right arm of chromosome XI, positioned midway between the centromere and telomere, where it is flanked by long terminal repeats (LTRs) and adjacent to an origin of replication. *DUR3* is positioned on the left arm of chromosome VIII, approximately 10 kilobase pairs from a long repetitive track that includes LTRs and retrotransposon genes. Determining how the intrinsic properties of a gene and its neighboring features (including DNA replication origins, centromeres, telomeres, tRNA genes, and repetitive elements) contribute to the molecular and mechanistic processes that underlie CNV formation is an important next step that can now be efficiently tested using our CNV reporter (Qian and Adhya 2017; Tosato et al. 2017; Mason and McEachern 2018).

Potential applications of a CNV reporter

The use of a phenotypic CNV reporter can be readily extended to additional applications to address heretofore intractable problems. Follow-up experiments using fluctuating or complex environments in chemostats are particularly interesting. For example, analyzing *GAP1* CNV dynamics in a nitrogen-limited chemostat where glutamine and urea are both present in limiting concentrations would allow us to simultaneously track divergent ecological niches and

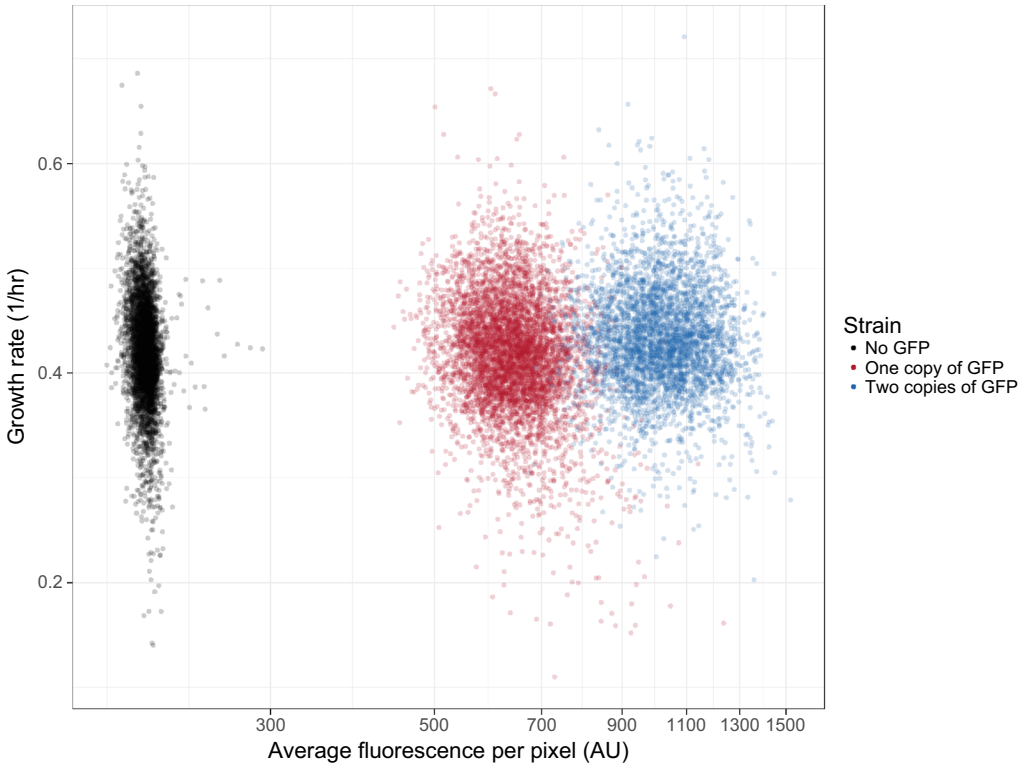

Fig. 4 Cells with two copies of GFP can be distinguished from cells with one copy of GFP using live-cell imaging. A high-throughput microcolony growth-rate assay (Ziv et al. 2013, 2017) was performed using live-cell imaging. Cells with two copies of GFP cluster together and have higher fluorescent signal than cells with one copy of GFP. Image courtesy of Naomi Ziv

adaptive strategies (i.e., the presence of *GAP1* amplifications vs. *GAP1* deletions in independent replicates). Extension of the approach to a two-color system (for example, using GFP at the *GAP1* locus and mCherry at the *DUR3* locus) would enable direct comparisons of CNV dynamics at these two loci in this type of complex environment. A two-color system could also be useful for differentiating between diploidization or aneuploidy and amplification of a specific gene of interest (Steinrueck and Guet 2017; Harari et al. 2018).

A prevailing view in evolutionary biology is that mutations are randomly generated, and then, selection acts on them to dictate their ultimate fate in a population. However, recent studies suggest that CNV formation may be linked to replication, transcription, and environmental stress (Chen et al. 2015; Wilson et al. 2015; Thomas and Rothstein 1989; Skourti-Stathaki and Proudfoot 2014; Mason and McEachern 2018). Induction of replicative stress leads to increased formation of CNVs (Foster 2007; Galhardo et al. 2007; Shor et al. 2013). Active transcription units may be "hotspots" for CNV formation, as collisions between the replisome and RNA polymerase (as well as other transcription-mediated events) can lead to DNA damage and a higher probability of improper DNA repair (Wilson et al. 2015). It has been proposed that gene duplication may occur in response to environmental stimuli, with compelling evidence from studies of the rDNA array and the *CUP1* locus in yeast (Jack et al. 2015; Hull et al. 2017; Mansisidor et al. 2018). An intriguing possibility is that the high level of *GAP1* transcription in nitrogen-limited chemostats (Airoldi et al. 2016) may enhance *GAP1* CNV formation rates. To dissect the relationship between environmental stimulation and CNV formation, *GAP1* CNVs could be tracked and compared under activating and repressive conditions. To determine if transcription is directly responsible for mediating environmental stimulation, *GAP1* transcription could be manipulated using an estradiol-inducible system (McIsaac et al. 2014), a Tet-on/Tet-off system (Das et al. 2016), or by blocking *GAP1* transcription with a nuclease-deficient Cas9 (Jensen 2018). Findings from these types of studies would have broad implications and could challenge our current belief that natural selection is driven by genotypic variation that arises solely as a result of random mutation.

In addition to the types of experiments described above, the CNV reporter could be potentially integrated into animal and plant cell lines at specific loci of interest. In newer work from our lab, we tested whether cells containing a gene duplication could be distinguished from cells containing a single gene copy using a microcolony growth-rate assay (Ziv et al. 2013, 2017). In these experiments, cells with two copies of GFP have distinctly higher levels of fluorescence

when measured using live-cell imaging (Fig. 4). This suggests that the CNV reporter could be used to track CNV formation within developing tissues or cancer cell lines using imaging methods. Moreover, as our study demonstrated that subpopulations of cells with de novo CNVs can be isolated by FACS (Fig. 2), this approach would be ideal for fractionating heterogeneous tumor populations. RNA sequencing or chromatin conformation capture of fractionated populations could then be performed to investigate how CNV alleles alter gene expression and chromatin organization, respectively.

Using a CNV reporter to track and isolate cancer cells with amplification of known oncogenes, such as those in the *MYC* and *RAS* families, would be an interesting avenue to pursue. Depending on the level of *MYC* gene dosage, cells can undergo apoptosis, proliferative arrest, or cellular senescence (Gabay et al. 2014; Bagci and Kurtgöz 2015; Chen et al. 2018). While *MYC* amplification certainly plays a role in tumorigenesis, tumor progression is typically coupled with other oncogenic events (Gabay et al. 2014). Use of a CNV reporter could, therefore, be advantageous for determining the temporal order or genetic context of oncogene amplification, especially if additional fluorescent reporters were used to track CNVs at multiple loci. This would enable further dissection of the causal links between CNVs and tumorigenicity.

Conclusions

In summary, we have developed a phenotypic reporter that overcomes many of the inherent challenges in studying CNVs in heterogeneous populations of cells, which we have discussed here in a short review of existing and emerging technologies used for CNV detection. In our recent study using the CNV reporter, we showed that CNVs arise early and predictably during experimental evolution in nutrient-limited chemostats. *GAP1* CNVs are generated at a high rate and can be variable in both size and copy number. The extent of CNV diversity is surprising given that the majority of CNVs are formed through replication-mediated processes involving short, inverted-repeat sequences. These findings suggest that much remains to be learned about the mechanisms and molecular processes underlying CNV formation, especially across different genomic contexts. We believe that our phenotypic CNV reporter opens the door to numerous questions that were previously intractable, including an investigation of the role of environmental stimulation in generating CNVs. Although our initial study was performed in yeast cells, we are enthused about the prospects for applying this approach broadly to study the role of duplications and deletions in disease and evolution.

References

Aigner J, Villatoro S, Rabionet R, Roquer J, Jiménez-Conde J, Martí E, Estivill X (2013) A common 56-kilobase deletion in a primate-specific segmental duplication creates a novel butyrophilin-like protein. BMC Genet 14(July):61

Airoldi EM, Miller D, Athanasiadou R, Brandt N, Abdul-Rahman F, Neymotin B, Hashimoto T, Bahmani T, Gresham D (2016) Steady-State and dynamic gene expression programs in *Saccharomyces cerevisiae* in response to variation in environmental nitrogen. Mol Biol Cell 27(8):1383–1396

Aitman TJ, Dong R, Vyse TJ, Norsworthy PJ, Johnson MD, Smith J, Mangion J et al (2006) Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. Nature 439(7078):851–855

Anderson RP, Roth JR (1977) Tandem genetic duplications in phage and bacteria. Annu Rev Microbiol 31(1):473–505

Arguello J Roman, Chen Y, Yang S, Wang W, Long M (2006) Origination of an X-linked testes chimeric gene by illegitimate recombination in drosophila. PLoS Genet 2(5):e77

Arlt MF, Ozdemir AC, Birkeland SR, Wilson TE, Glover TW (2011) Hydroxyurea induces de novo copy number variants in human cells. Proc Natl Acad Sci USA 108(42):17360–17365

Arlt MF, Rajendran S, Birkeland SR, Wilson TE, Glover TW (2014) Copy number variants are produced in response to low-dose ionizing radiation in cultured cells. Environ Mol Mutagen 55(2):103–113

Avise JC, Kitto GB (1973) Phosphoglucose isomerase gene duplication in the bony fishes: an evolutionary history. Biochem Genet 8(2):113–132

Bagci O, Kurtgöz S (2015) Amplification of cellular oncogenes in solid tumors. N Am J Med Sci 7(8):341–346

Blount ZD, Borland CZ, Lenski RE (2008) Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. Proc Natl Acad Sci USA 105(23):7899–7906

Blount ZD, Barrick JE, Davidson CJ, Lenski RE (2012) Genomic analysis of a key innovation in an experimental *Escherichia coli* population. Nature 489(7417):513–518

Brewer BJ, Payen C, Raghuraman MK, Dunham MJ (2011) Origin-dependent inverted-repeat amplification: a replication-based model for generating palindromic amplicons. PLoS Genet 7(3):e1002016

Brewer BJ, Payen C, Di Rienzi SC, Higgins MM, Ong G, Dunham MJ, Raghuraman MK (2015) Origin-dependent inverted-repeat amplification: tests of a model for inverted DNA amplification. PLoS Genet 11(12):e1005699

Bridges CB (1936) The Bar 'GENE' a duplication. Science 83(2148):210–211

Brown CJ, Todd KM, Rosenzweig RF (1998) Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. Mol Biol Evol 15(8):931–942

Bussotti G, Gouzelou E, Boite MC, Kherachi I, Spath GF (2018) Leishmania genome dynamics during environmental adaptation reveals strain-specific differences in gene copy number variation, karyotype instability, and telomeric amplification. mBio 9(6):e01399-18

Caignec L, Cedric CS, Sermon K, De Rycke M, Thienpont B, Debrock S, Staessen C et al (2006) Single-cell chromosomal imbalances detection by array CGH. Nucl Acids Res 34(9):e68

Chakraborty M, VanKuren NW, Zhao R, Zhang X, Kalsow S, Emerson JJ (2018) Hidden genetic variation shapes the structure of functional elements in *Drosophila*. Nat Genet 50(1):20–25

Chan YF, Marks ME, Jones FC, Jr GV, Shapiro MD, Brady SD, Southwick AM et al (2010) Adaptive evolution of pelvic reduction in

sticklebacks by recurrent deletion of a Pitx1 enhancer. Science 327(5963):302–305
Chen L, Zhou W, Zhang L, Zhang F (2014) Genome architecture and its roles in human copy number variation. Genom Inf 12(4):136–144
Chen L, Zhou W, Zhang C, Lupski JR, Jin L, Zhang F (2015) CNV instability associated with DNA replication dynamics: evidence for replicative mechanisms in CNV mutagenesis. Hum Mol Genet 24(6):1574–1583
Chen H, Liu H, Qing G (2018) Targeting oncogenic Myc as a strategy for cancer treatment. Signal Trans Target Therap 3(February):5
Clop A, Vidal O, Amills M (2012) Copy number variation in the genomes of domestic animals. Anim Genet 43(5):503–517
Conant GC, Wolfe KH (2008) Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet 9(12):938–950
Couldrey C, Keehan M, Johnson T, Tiplady K, Winkelman A, Littlejohn MD, Scott A et al (2017) Detection and assessment of copy number variation using PacBio long-read and illumina sequencing in New Zealand dairy cattle. J Dairy Sci 100(7):5472–5478
Das AT, Tenenbaum L, Berkhout B (2016) Tet-on systems for doxycycline-inducible gene expression. Curr Gene Ther 16(3):156–167
Dhami MK, Hartwig T, Fukami T (2016) Genetic basis of priority effects: insights from nectar yeast. Proc Biol Sci R Soc 283(1840):20161455. https://doi.org/10.1098/rspb.2016.1455
Dulmage KA, Darnell CL, Vreugdenhil A, Schmid AK (2018) Copy number variation is associated with gene expression change in archaea. Microbial Genom. https://doi.org/10.1099/mgen.0.000210
Ferris SD, Whitt GS (1979) Evolution of the differential regulation of duplicate genes after polyploidization. J Mol Evol 12(4):267–317
Foster PL (2007) Stress-induced mutagenesis in bacteria. Crit Rev Biochem Mol Biol 42(5):373–397
Franke M, Ibrahim DM, Andrey G, Schwarzer W, Heinrich V, Schöpflin R, Kraft K et al (2016) Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature 538(7624):265–269
Gabay M, Li Y, Felsher DW (2014) MYC activation is a hallmark of cancer initiation and maintenance. Cold Spring Harbor Perspect Med 4(6):a014241. https://doi.org/10.1101/cshperspect.a014241
Galhardo RS, Hastings PJ, Rosenberg SM (2007) Mutation as a stress response and the regulation of evolvability. Crit Rev Biochem Mol Biol 42(5):399–435
Gamazon ER, Stranger BE (2015) The impact of human copy number variation on gene expression. Brief Funct Genom 14(5):352–357
Gamazon ER, Nicolae DL, Cox NJ (2011) A study of CNVs as trait-associated polymorphisms and as expression quantitative trait loci. PLoS Genet 7(2):e1001292
Geiger T, Cox J, Mann M (2010) Proteomic changes resulting from gene copy number variations in cancer cells. PLoS Genet 6(9):e1001090
Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ et al (2005) The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science 307(5714):1434–1440
Greenblum S, Carr R, Borenstein E (2015) Extensive strain-level copy-number variation across human gut microbiome species. Cell 160(4):583–594
Gresham D, Desai MM, Tucker CM, Jenq HT, Pai DA, Ward A, DeSevo CG, Botstein D, Dunham MJ (2008) The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. PLoS Genet 4(12):e1000303
Gresham D, Usaite R, Germann SM, Lisby M, Botstein D, Regenberg B (2010) Adaptation to diverse nitrogen-limited environments by deletion or extrachromosomal element formation of the GAP1 locus. Proc Natl Acad Sci USA 107(43):18551–18556
Gu S, Yuan B, Campbell IM, Beck CR, Carvalho Claudia M B, Nagamani Sandesh C S, Erez A et al (2015) Alu-mediated diverse and complex pathogenic copy-number variants within human chromosome 17 at p13.3. Hum Mol Genet 24(14):4061–4077
Hansche PE (1975) Gene duplication as a mechanism of genetic adaptation in *Saccharomyces cerevisiae*. Genetics 79(4):661–674
Harari Y, Ram Y, Kupiec M (2018) Frequent ploidy changes in growing yeast cultures. Curr Genet 64(5):1001–1004
Hartmann FE, Croll D (2017) Distinct trajectories of massive recent gene gains and losses in populations of a microbial eukaryotic pathogen. Mol Biol Evol 34(11):2808–2822
Hastings PJ, Ira G, Lupski JR, Iafrate AJ, Feuk L, Rivera MN, Listewnik ML et al (2009) A microhomology-mediated break-induced replication model for the origin of human copy number variation. PLoS Genet 5(1):e1000327
Henrichsen CN, Vinckenbosch N, Zöllner S, Chaignat E, Pradervand S, Schütz F, Ruedi M, Kaessmann H, Reymond A (2009) Segmental copy number variation shapes tissue transcriptomes. Nat Genet 41(4):424–429
Hong J, Gresham D (2014) Molecular specificity, convergence and constraint shape adaptive evolution in nutrient-poor environments. PLoS Genet 10(1):e1004041
Hopkinson DA, Edwards YH, Harris H (1976) The distributions of subunit numbers and subunit sizes of enzymes: a study of the products of 100 human gene loci. Ann Hum Genet 39(4):383–411
Horiuchi T, Horiuchi S, Novick A (1963) The genetic basis of hypersynthesis of beta-galactosidase. Genetics 48:157–169
Huddleston J, Chaisson MJP, Steinberg KM, Warren W, Hoekzema K, Gordon D, Graves-Lindsay TA et al (2017) Discovery and genotyping of structural variation from long-read haploid genome sequence data. Genome Res 27(5):677–685
Hughes AL (1994) The evolution of functionally novel proteins after gene duplication. Proc R Soc Lond B Biol Sci 256(1346):119–124
Hull RM, Cruz C, Jack CV, Houseley J (2017) Environmental change drives accelerated adaptation through stimulated copy number variation. PLoS Biol 15(6):e2001333
Itsara A, Hao W, Smith JD, Nickerson DA, Romieu I, London SJ, Eichler EE (2010) De Novo rates and selection of large copy number variation. Genome Res 20(11):1469–1481
Jack CV, Cruz C, Hull RM, Keller MA, Ralser M, Houseley J (2015) Regulation of ribosomal DNA amplification by the TOR pathway. Proc Natl Acad Sci USA 112(31):9674–9679
Jensen MK (2018) Design principles for nuclease-deficient CRISPR-based transcriptional regulators. FEMS Yeast Res 18(4):foy039. https://doi.org/10.1093/femsyr/foy039
Kim S, Cho C-S, Han K, Lee J (2016) Structural variation of alu element and human disease. Genom Inf 14(3):70–77
Konings P, Vanneste E, Jackmaert S, Ampe M, Verbeke G, Moreau Y, Vermeesch JR, Voet T (2012) Microarray analysis of copy number variation in single cells. Nat Protoc 7(2):281–310
Koszul R, Caburet S, Dujon B, Fischer G (2004) Eukaryotic genome evolution through the spontaneous duplication of large chromosomal segments. EMBO J 23(1):234–243
Lauer S, Avecilla G, Spealman P, Sethia G, Brandt N, Levy SF, Gresham D (2018) Single-cell copy number variant detection reveals the dynamics and diversity of adaptation. PLoS Biol 16(12):e3000069
Lee JA, Carvalho Claudia M B, Lupski JR (2007) A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell 131:1235–1247
Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D et al (2015) Disruptions of topological chromatin

domains cause pathogenic rewiring of gene-enhancer interactions. Cell 161(5):1012–1025
Lupiáñez DG, Spielmann M, Mundlos S (2016) Breaking TADs: how alterations of chromatin domains result in disease. Trends Genet TIG 32(4):225–237
Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290(5494):1151–1155
Lynch M, Force A (2000) The probability of duplicate gene preservation by subfunctionalization. Genetics 154(1):459–473
Mansisidor A, Jr TM, Srivastava P, Dartis DD, Delgado AP, Blitzblau HG, Klein H, Hochwagen A (2018) Genomic copy-number loss is rescued by self-limiting production of DNA circles. Mol Cell 72(3):583-93.e4
Mason Jennifer M O, McEachern MJ (2018) Chromosome ends as adaptive beginnings: the potential role of dysfunctional telomeres in subtelomeric evolvability. Curr Genet 64(5):997–1000
Mayo S, Monfort S, Roselló M, Orellana C, Oltra S, Caro-Llopis A, Martínez F (2017) Chimeric genes in deletions and duplications associated with intellectual disability. Int J Genom Proteom 2017(May):4798474
McIsaac RS, Gibney PA, Chandran SS, Benjamin KR, Botstein D (2014) Synthetic biology tools for programming gene expression without nutritional perturbations in *Saccharomyces cerevisiae*. Nucl Acids Res 42(6):e48
Merla G, Howald C, Henrichsen CN, Lyle R, Wyss C, Zabot M-T, Antonarakis SE, Reymond A (2006) Submicroscopic deletion in patients with Williams-Beuren syndrome influences expression levels of the nonhemizygous flanking genes. Am J Hum Genet 79(2):332–341
Michels E, De Preter K, Van Roy N, Speleman F (2007) Detection of DNA copy number alterations in cancer by array comparative genomic hybridization. Genet Med 9(September):574
Molina J, Carmona-Mora P, Chrast J, Krall PM, Canales CP, Lupski JR, Reymond A, Walz K (2008) Abnormal social behaviors and altered gene expression rates in a mouse model for Potocki-Lupski syndrome. Hum Mol Genet 17(16):2486–2495
Ohno S (1970) Evolution by gene duplication. Springer, Berlin
Payen C, Di Rienzi SC, Ong GT, Pogachar JL, Sanchez JC, Sunshine AB, Raghuraman MK, Brewer BJ, Dunham MJ (2014) The dynamics of diverse segmental amplifications in populations of *Saccharomyces cerevisiae* adapting to strong selection. Gene Genomes Genet 4(3):399–409
Pham GM, Newton L, Wiegert-Rininger K, Vaillancourt B, Douches DS, Robin Buell C (2017) Extensive genome heterogeneity leads to preferential allele expression and copy number-dependent expression in cultivated potato. Plant J Cell Mol Biol 92(4):624–637
Qian Z, Adhya S (2017) DNA repeat sequences: diversity and versatility of functions. Curr Genet 63(3):411–416
Ramirez O, Olalde I, Berglund J, Lorente-Galdos B, Hernandez-Rodriguez J, Quilez J, Webster MT et al (2014) analysis of structural diversity in wolf-like canids reveals post-domestication variants. BMC Genom 15(1):465
Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Daniel Andrews T, Fiegler H et al (2006) Global variation in copy number in the human genome. Nature 444(7118):444–454
Rigau M, Juan D, Valencia A, Rico D (2019) Intronic CNVs and gene expression variation in human populations. PLoS Genet 15(1):e1007902
Rippey C, Walsh T, Gulsuner S, Brodsky M, Nord AS, Gasperini M, Pierce S et al (2013) Formation of chimeric genes by copy-number variation as a mutational mechanism in *Schizophrenia*. Am J Hum Genet 93(4):697–710
Schrider DR, Navarro Fabio C P, Galante Pedro A F, Parmigiani RB, Camargo AA, Hahn MW, de Souza SJ (2013) Gene copy-number polymorphism caused by retrotransposition in humans. PLoS Genet 9(1):e1003242
Schughart K, Kappen C, Ruddle FH (1989) Duplication of large genomic regions during the evolution of vertebrate homeobox genes. Proc Natl Acad Sci USA 86(18):7067–7071
Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Månér S et al (2004) Large-scale copy number polymorphism in the human genome. Science 305(5683):525–528
Shor E, Fox CA, Broach JR (2013) The yeast environmental stress response regulates mutagenesis induced by proteotoxic stress. PLoS Genet 9(8):e1003680
Skourti-Stathaki K, Proudfoot NJ (2014) A double-edged sword: r loops as threats to genome integrity and powerful regulators of gene expression. Genes Dev 28(13):1384–1396
Sonti RV, Roth JR (1989) Role of gene duplications in the adaptation of *Salmonella typhimurium* to growth on limiting carbon sources. Genetics 123(1):19–28
Spielmann M, Lupiáñez DG, Mundlos S (2018) Structural variation in the 3D genome. Nat Rev Genet 19(7):453–467
Steinrueck M, Guet CC (2017) Complex chromosomal neighborhood effects determine the adaptive potential of a gene under selection. eLife. https://doi.org/10.7554/eLife.25100
Stratton MR, Campbell PJ, Andrew Futreal P (2009) The cancer genome. Nature 458(7239):719–724
Stuber CW, Goodman MM (1983) Inheritance, intracellular localization, and genetic variation of phosphoglucomutase isozymes in maize (*Zea mays* L.). Biochem Genet 21(7–8):667–689
Sturtevant AH (1925) The effects of unequal crossing over at the bar locus in *Drosophila*. Genetics 10(2):117–147
Taylor JS, Raes J (2004) Duplication and divergence: the evolution of new genes and old ideas. Annu Rev Genet 38:615–643
Thomas BJ, Rothstein R (1989) Elevated recombination rates in transcriptionally active DNA. Cell 56(4):619–630
Tosato V, Sims J, West N, Colombin M, Bruschi CV (2017) Post-translocational adaptation drives evolution through genetic selection and transcriptional shift in *Saccharomyces cerevisiae*. Curr Genet 63(2):281–292
Turner KM, Deshpande V, Beyter D, Koga T, Rusert J, Lee C, Li B et al (2017) Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. Nature 543(7643):122–125
Walsh JB (1995) How often do duplicated genes evolve new functions? Genetics 139(1):421–428
Walsh B (2003) Population-genetic models of the fates of duplicate genes. Genetica 118(2–3):279–294
Wang H, Chai Z, Dan H, Ji Q, Xin J, Zhang C, Zhong J (2019) A global analysis of CNVs in diverse yak populations using whole-genome resequencing. BMC Genom 20(1):61
Wilson TE, Arlt MF, Park SH, Rajendran S, Paulsen M, Ljungman M, Glover TW (2015) Large transcription units unify copy number variants and common fragile sites arising under replication stress. Genome Res 25(2):189–200
Yang L, Lingyang X, Zhou Y, Liu M, Wang L, Kijas JW, Zhang H, Li L, Liu GE (2018) Diversity of copy number variation in a worldwide population of sheep. Genomics 110(3):143–148
Zarrei M, MacDonald JR, Merico D, Scherer SW (2015) A copy number variation map of the human genome. Nat Rev Genet 16(3):172–183
Zhou B, Ho SS, Zhang X, Pattni R, Haraksingh RR, Urban AE (2018) Whole-genome sequencing analysis of CNV using low-coverage and paired-end strategies is efficient and outperforms array-based CNV analysis. J Med Genet 55(11):735–743
Ziv N, Siegal ML, Gresham D (2013) Genetic and nongenetic determinants of cell growth variation assessed by high-throughput microscopy. Mol Biol Evol 30(12):2568–2578

Ziv N, Shuster BM, Siegal ML, Gresham D (2017) Resolving the complex genetic basis of phenotypic variation and variability of cellular growth. Genetics 206(3):1645–1657

Żmieńko A, Samelak A, Kozłowski P, Figlerowicz M (2014) Copy number polymorphism in plant genomes. TAG 127(1):1–18

Zuellig MP, Sweigart AL (2018) Gene duplicates cause hybrid lethality between sympatric species of mimulus. PLoS Genet 14(4):e1007130