

Overlooked: Extrachromosomal DNA and Their Possible Impact on Whole Genome Sequencing

Reinhard H DENNIN

Submitted: 12 Jul 2017
Accepted: 5 Jan 2018
Online: 27 Apr 2018

Department of Infectious Diseases and Microbiology, University of Luebeck, UKSH, Campus Luebeck, D-23538 Luebeck, Germany

To cite this article: Dennin RH. Overlooked: extrachromosomal DNA and their possible impact on whole genome sequencing. *Malays J Med Sci.* 2018;**25**(2):20–26. <https://doi.org/10.21315/mjms2018.25.2.3>

To link to this article: <https://doi.org/10.21315/mjms2018.25.2.3>

Abstract

Extrachromosomal (ec) DNA in eukaryotic cells has been known for decades. The structures described range from linear double stranded (ds) DNA to circular dsDNA, distinct from mitochondrial (mt) DNA. The sizes of circular forms are described from some hundred base pairs (bp) up to more than 150 kbp. The number of molecules per cell ranges from several hundred to a thousand. Semi-quantitative determinations of circular dsDNA show proportions as high as several percentages of the total DNA per cell. These ecDNA fractions harbor sequences that are known to be present in chromosomal DNA (chrDNA) too. Sequencing projects on, for example the human genome, have to take into account the ecDNA sequences which are simultaneously ascertained; corrections cannot be performed retrospectively. Concerning the results of sequencings derived from extracted whole DNA: if the ecDNA fractions contained therein are not taken into account, erroneous conclusions at the chromosomal level may result.

Keywords: extrachromosomal, episomal, genome, sequencing, chromosome

Introduction

Extrachromosomal dsDNA (ecDNA)—apart from the mtDNA—in eukaryotic cells of human, animal, and plant origin has been known for more than three decades at least (1–5). These ecDNA molecules are, according to the structure, both linear and circular dsDNA; the circular ones are mostly termed small polydispersed circular DNA (spcDNA); occasionally synonymously termed extrachromosomal circular DNA (eccDNA) just to classify the DNA structure (6, 7). They differ in size from about several hundred base pairs (bp) (8) up to ≥ 25 kbp, and up to more than 150 kbp (9) for the large circular sized DNA. The number of spc/eccDNA molecules ranges from several hundreds up to one thousand per cell. The term “polydispersed” already hints at a great diversity of these DNA molecules; it may also indicate different types of functions. The spc/ecc-DNA fraction can constitute a substantial part of the whole DNA

extracted, for example from human peripheral blood mononuclear cells (PBMCs). Semi-quantitative experimental assessments show the spc-dsDNA fraction constitutes several percentages of the whole DNA extracted from PBMCs of healthy human subjects (10). However, it may depend on, for example, the metabolic status of the cell’s activation. Regarding certain sequences, there are indications for “rolling circle replication” (11). Therefore, the composition of ecDNA might change with time and the sequences could be different upon replication. Circular DNA forms derived from exogenous viral infections such as Hepatitis B virus (HBV) (12) or retroviral E-DNA (13) are not considered here.

There is broad agreement that ecDNA originates through diverse mechanisms, for example mobile genetic elements (MGE) (14) and “Mismatch Repair Pathways” (15), that can rearrange diverse DNA sequences from chrDNA (16). In the context of mobile DNA, different

kinds of retro-/transposons (17, 18), long interspersed nucleotide elements (LINE) (19), short interspersed repetitive DNA sequences such as Alu elements (20), and telomeric repeats (21) are known to be essential parts of chrDNA. They are also detected in the ecDNA fraction of eukaryotic cells. This means that chromosomal derived sequences are present in the ecDNA but rearranged in some way; how this works in detail remains speculative (22). Checking these mobile activities revealed aspects of both randomly and non-randomly caused instabilities within the genome (23). It may be based on intrinsic, genetically ingrained structures that are activated on demand by environmental impacts. This might reflect the plasticity of the human genome (24), including the ecDNA. Apart from general aspects, functions of ecDNA in different eukaryotic cells are not known. Yet, certain hints concerning cancer cells are given as described in section '4' below.

It is not known for sure: (i) where the formation of ecDNA takes place, in the cell nucleus and/or the cytoplasm, and (ii) whether it is a short transitory formation or a long time stable status of certain ecDNA regarding possible metabolic functions. Protocols for the preparation of ecDNA, in particular ecc-dsDNA, are outlined by a few studies (22, 25, 26). The reported findings of sequences homologous to chrDNA in ecDNA may reflect the particular interests of the authors and what they were looking for; their results may represent only a tiny part of the real existing sequences of chromosomal origin contained in ecDNA with yet unknown impact. It seems realistic to assume that the extensive repertoire of ecDNA sequences contains even more different chromosomal sequences than previously known.

Possible Impacts

Four selected situations should draw attention to possible discrepancies when "whole DNA from eukaryotic cells" was used for genome sequencing (whole genome sequencing, WGS) to ascertain the sequences of the respective chromosomes. They are intended to point to a paramount importance of both general as well as specific aspects of genetics.

1. General aspects: The formation of ecDNA is obviously a complex process. However, it is known that ecDNA are composed by means of chromosomal mobile sequences, such as transposons, long terminal repeats (LTR), and Alu elements, but they themselves

contain these sequences in different degrees. Therefore, in addition to the existence of ecDNA itself, a potential for changes in the composition of their sequence, including an active function, cannot be ruled out. This concerns species of ecDNA, for example those showing sequences in coincidence with Human Endogenous Retro-Viruses H (HERV-H) (27). HERVs have the potential to replicate and to transpose themselves. Therefore, when discussing the issues of HERVs based only on results from sequenced human genomes based on whole cellular DNA, the possible "contaminating influence" of ecDNA with integrated HERV sequences remains unmentioned (28). Studies on, for example endogenous viruses (29), LINE, *et cetera*, performed with whole DNA extracts from human cells or tissues may need to be reevaluated because the ecDNA fractions were not taken into account. This uncertainty is because sections of the LTR sequences of HERVs have been depicted to be present in the spc/ecDNA fraction. Studies on retrotransposons (30, 31) should be reconsidered regarding their allocation upon alignments of their ecDNA and/or chrDNA if sequencing data is based on whole DNA extracts only. These considerations apply to plants too (32). Therefore, vagueness remains as these facts have not yet been considered in otherwise exceptionally structured review articles (33).

2. The human genome project: Basically, the same holds true for sequencing projects regarding the human genome (34). Here, the term "genome" usually pertains to chrDNA in eukaryotic cells. When the extracted whole DNA from the respective human cells was used for fragmentation, cloning, and sequencing (35, 36), it means: the whole DNA was not separated into (i) chrDNA and (ii) ecDNA before fragmentation. That is, the fragments generated for sequencing consist of a mixture of short DNA sequence stretches derived from both chrDNA and ecDNA, apart from the well-known mtDNA. Furthermore, if the fragmentation of the whole DNA was performed with restriction endonucleases, the remaining single stranded ends of their cutting sequences from fragments derived both from chrDNA and ecDNA would be identical. An incorrect assignment of fragments originating from the ecDNA into the final chrDNA is, therefore, more

likely. In addition, the possible different patterns of methylation of the cutting sequences of the applied restriction enzymes on both chrDNA and ecDNA sequences might result in uncertainties on alignments of chrDNA. Similarly, findings indicate that using mechanical shear forces on whole DNA to get fragments for next-generation sequencing (NGS) results in DNA sequences that are non-randomly fragmented (37). This effect has not been considered for ecDNA with respect to their broad range in molecular sizes. The various ecDNA contain rearranged DNA or DNA composed of shorter chromosomal sequences, for example by incremental acquisition by MGEs. Depending on the proportion of ecDNA in the subject's whole cell DNA, a greater or lesser extent of sequences of ecDNA was fragmented, cloned, and subjected to NGS. Sequences derived from ecDNA do not carry tags that exactly predict their source, such as derived from ecDNA, and might have been handled as sequences of chromosomes in the final alignments. NGS of the whole DNA of eukaryotic cells allows no discrimination between chrDNA and ecDNA sequences. Nothing is known about how many, and where, ecDNA derived sequences are wrongly placed into chrDNA; this may have led to wrong conclusions in final analysis in various fields. Therefore, critical aspects arise: possible uncertainties with the allocation and alignment of the final chromosomal sequences have to be taken into account if the ecDNA fractions have not been considered.

3. The ENCODE project: Furthermore, the issues addressed also apply to discussions on epigenetics of the human genome (38). The ENCODE project is designed to look for epigenomes in the human genome, for example (39). However, ecDNA could exhibit patterns of individual methylation too (10). Therefore, they might be co-precipitated, for example, according to the immunoprecipitation protocols for methylated DNA. These possible situations have not been considered in the respective results.
4. Medical aspects: "...as next-generation sequencing begins to break down the barriers between research and the clinic, as genomic

and clinical data are responsibly integrated into the "cloud" to look for patterns of health and disease..." (33). However, it is imperative to note that results showing increased numbers of ec/spcDNA in aged and malignant cells containing indicator sequences (40), in particular the double-minute ecDNAs (41, 42), must receive attention. In cancer cell lines, the amount of spcDNA can be as high as up to 17.8% (14). The points in question here are the issues: when sequencing whole DNA from such aged cells, how do these indicator sequences harbored in ecDNA get identified for the final chromosomal sequence alignments? How are they treated if they contain mutations, single ones or multiple copy number variations? Not considering the fraction of ecDNA in the case of the individualised sequencing of whole DNA for multigenic analysis may entail incorrect association in disease assessments. This has also to be seen in the context of Alu sequences in "germline genetic diseases" (43), which also applies to "normal lymphocytes" (14). Therefore, activities to integrate patients' genetic data gained from NGS of whole DNA into this "cloud" to use them as diagnostic tools should not be enforced until any uncertainties possibly deriving from ecDNAs are eliminated.

Discussion

Studies on ecDNA from human cells and tissues have shown that they contain genetic elements which are known to belong to chrDNA. Protocols are available for the correction of possible "sequence errors" after sequencing (44–48). However, genomic DNA sequencing, that is WGS, in which ecDNA has not been separated from chrDNA prior to fragmentation can cause problems. This applies if sequencings are performed and the results are analysed without knowledge of the sequences of the ecDNA as parts of the whole cellular DNA. It is, therefore, questionable whether, after NGS of the whole DNA of eukaryotic cells, corrections to chrDNA with bioinformatics lacking data on ecDNA can be adequately achieved.

Therefore, evidence-based studies are required to demonstrate whether the disregard of ecDNA can or may not cause any uncertainties in the assessment of chromosomal sequences in general, and also of diseases on the basis of individualised sequencing. For example,

sequences of ecDNA have to be subtracted by correctional algorithms from sequences concurrently obtained from whole DNA of cells to keep chromosomal sequences free from contaminating ecDNA sequences. Further strategies for that reason are outlined by a few studies (22, 25, 26).

The demand for evaluating NGS results using “professional standards” is correct (49), but it requires additional aspects due to possible imprecisions by ecDNA. This also applies to the latest NGS (50, 51), especially when cancer neochromosomes have to be considered (52). Based on such results, it should be possible (i) to clarify the potential impact of hitherto not considered ecDNAs on whole DNA sequencing, (ii) to exclude imponderabilities as far as possible, and (iii) to develop standards that make the results of different NGS protocols undoubtedly comparable.

Conclusions

This review raises critical questions about the reliance on so-called genome sequencing data of whole DNA, in particular of human genomes. EcDNA in eukaryotic cells is a fact. The evaluation of their possible impact in NGS of the human genome is necessary. Independent of the sequencing protocols applied, if the ecDNA has not been separated from chrDNA prior to fragmentation, uncertainties may remain regarding the final chromosomal sequences depending on the fraction of admixtures from ecDNA. Certain aspects have to be addressed: (i) the self-replicating potential of ecDNA with the consequence of mutations and rearrangements that can result, (ii) possible interchange in both directions between ecDNA and chrDNA. Careful assessments of the risks are necessary, especially when it comes to medical applications.

Abbreviations: Alu, sequences named according to the restriction endonuclease AluI; bp(s), base pair(s); chrDNA, chromosomal DNA; ecDNA, extrachromosomal DNA; eccDNA, extrachromosomal circular DNA; HERV-H, Human Endogenous Retroviruses type H; mtDNA, mitochondrial DNA; LINE, long interspersed nucleotide elements; LTR, long terminal repeats; MGE, mobile genetic element(s); NGS, next-generation sequencing; spcDNA, small polydispersed circular DNA; PBMC, peripheral blood mononuclear cells; WGS, whole genome sequencing.

Correspondence

Prof Dr Reinhard H Dennin
Dr rer nat habil (Christian-Albrechts-University, Kiel;
University of Lübeck, UKSH, Germany)
Academic Director, Head of Virology, Rtrd,
Department of Infectious Diseases and Microbiology,
University of Luebeck, UKSH, Campus Luebeck,
160 Ratzeburger Allee, D-23538 Luebeck, Germany.
Tel: +49 451 500 45317
Fax: +49 451 500 45318
E-mail: Reinhard.dennin@uksh.de

References

1. Yamagishi H. Role of mammalian circular DNA in cellular differentiation. *BioEssays*. 1986;**4(5)**:218–221. <https://doi.org/10.1002/bies.950040508>
2. Kunisada T, Yamagishi H, Sekiguchi T. Intracellular location of small circular DNA complexes in mammalian cell lines. *Plasmid*. 1983;**10(3)**:242–250. [https://doi.org/10.1016/0147-619X\(83\)90038-0](https://doi.org/10.1016/0147-619X(83)90038-0)
3. Krolewski JJ, Schindler CW, Rush MG. Structure of extrachromosomal circular DNAs containing both the Alu family of dispersed repetitive sequences and other regions of chromosomal DNA. *J Mol Biol*. 1984;**174(1)**:41–54. [https://doi.org/10.1016/0022-2836\(84\)90364-4](https://doi.org/10.1016/0022-2836(84)90364-4)
4. Rush MG, Misra R. Extrachromosomal DNA in eucaryotes. *Plasmid*. 1985;**14(3)**:177–191. [https://doi.org/10.1016/0147-619X\(85\)90001-0](https://doi.org/10.1016/0147-619X(85)90001-0)
5. Cohen S, Houben A, Segal D. Extrachromosomal circular DNA derived from tandemly repeated genomic sequences in plants. *Plant J*. 2008;**53(6)**:1027–1034. <https://doi.org/10.1111/j.1365-313X.2007.03394.x>
6. Cohen S, Segal D. Extrachromosomal circular DNA in eukaryotes: possible involvement in the plasticity of tandem repeats. *Cytogenet Genome Res*. 2009;**124(3–4)**:327–338. <https://doi.org/10.1159/000218136>
7. Barreto SC, Uppalapati M, Ray A. Small circular DNAs in human pathology. *Malays J Med Sci*. 2014;**21(3)**:4–18.

8. Shibata Y, Kumar P, Layer R, Willcox S, Gagan JR, Griffith JD, et al. Extrachromosomal micro DNAs and chromosomal microdeletions in normal tissues. *Science*. 2012;**336**(6077):82–86. <https://doi.org/10.1126/science.1213307>
9. Sal'nokov KV. Extrachromosomal DNA in mammalian cells. (English Abstract) *Tsitologiya*. 1990;**32**(11):1061–1071.
10. Dennin RH, Wo J. Episomal DNA in human cells—a link with mobile genetic elements? Sections of the hepatitis C virus 5'-NCR sequence in PBMC's DNA of HCV negative human individuals. EMBO-EMBL Symposium-The Mobile Genome, Heidelberg, Germany, September 16–19, 2015. Abstract 84.
11. Cohen S, Agmon N, Sobol O, Segal D. Extrachromosomal circles of satellite repeats and 5S ribosomal DNA in human cells. *Mobile DNA*. 2010;**1**:11. <https://doi.org/10.1186/1759-8753-1-11>
12. Seeger Ch, Mason WS. Molecular biology of hepatitis B virus infection. *Virology*. 2015;**479–480**:672–686. <https://doi.org/10.1016/j.virol.2015.02.031>
13. Cara A, Klotman ME. Retroviral E-DNA: persistence and gene expression in non-dividing immune cells. *J Leukoc Biol*. 2006;**80**:1013–1017. <https://doi.org/10.1189/jlb.0306151>
14. Schmidt H, Taubert H, Lange H, Kriese K, Schmidt WD, Hoffmann St, et al. Small polydispersed circular DNA contains strains of mobile genetic elements and occurs more frequently in permanent cell lines of malignant tumors than in normal lymphocytes. *Oncol Rep*. 2009;**22**:393–400. https://doi.org/10.3892/or_00000450
15. Dillon LW, Kumar P, Shibata Y, Wang YH, Willcox S, Griffith JD, et al. Production of extrachromosomal micro DNAs is linked to mismatch repair pathways and transcriptional activity. *Cell Rep*. 2015;**11**(11):1749–1759. <https://doi.org/10.1016/j.celrep.2015.05.020>
16. Gaubatz JW. Extrachromosomal circular DNAs and genomic sequence plasticity in eukaryotic cells. *Mutat Res*. 1990;**237**(5–6):271–92. [https://doi.org/10.1016/0921-8734\(90\)90009-G](https://doi.org/10.1016/0921-8734(90)90009-G)
17. Burns KH, Boeke JD. Human transposon tectonics. *Cell*. 2012;**149**:740–752. <https://doi.org/10.1016/j.cell.2012.04.019>
18. Mourier T. Transposable elements and circular DNAs. *Mob Genet Elements*. 2016;**6**(6):e1240748. <https://doi.org/10.1080/2159256X.2016.1240748>
19. Kemp JR, Longworth MS. Crossing the LINE toward genomic instability: LINE-1 retrotransposition in cancer. *Front Chem*. 2015;**3**:68. <https://doi.org/10.3389/fchem.2015.00068>
20. Batzer MA, Deininger PL. Alu repeats and human genomic diversity. *Nat Rev Genet*. 2002;**3**:370–379. <https://doi.org/10.1038/nrg798>
21. Regev A, Cohen S, Cohen E, Bar-Am I, Lavi S. Telomeric repeats on small polydisperse circular DNA (spcDNA) and genomic instability. *Oncogene*. 1998;**17**(26):3455–3461. <https://doi.org/10.1038/sj.onc.1202250>
22. Shoura MJ, Gabdank I, Hansen L, Merker J, Gotlib J, Levene SD, et al. Intricate and cell-type-specific populations of endogenous circular DNA (eccDNA) in *C. elegans* and *H. sapiens*. *G3*. 2017;**7**(10):3295–3303. <https://doi.org/10.1534/g3.117.300141>
23. Kuttler F, Mai S. Formation of non-random extrachromosomal elements during development, differentiation and oncogenesis. *Semin Cancer Biol*. 2007;**17**(1):56–64. <https://doi.org/10.1016/j.semcancer.2006.10.007>
24. Alves JM, Lopes AM, Chikhi L, Amorim A. On the structural plasticity of the human genome: chromosomal inversions revisited. *Curr Genomics*. 2012;**13**(8):623–632. <https://doi.org/10.2174/138920212803759703>
25. Jørgensen TS, Kiil AS, Hansen MA, Sørensen SJ, Hansen LH. Current strategies for mobilome research. *Front Microbiol*. 2015;**5**:Article 750. <https://doi.org/10.3389/fmicb.2014.00750>
26. Møller HD, Bojsen RK, Tachibana C, Parsons L, Botstein D, Regenberg B. Genome-wide purification of extrachromosomal circular DNA from eukaryotic cells. *J Vis Exp*. 2016;**(110)**:e54239. <https://doi.org/10.3791/54239>

27. Huang H, Qian J, Proffit J, Wilber K, Jenkins R, Smith DI. FRA7G extends over a broad region: coincidence of human endogenous retroviral sequences (HERV-H) and small polydispersed circular DNAs (spcDNA) and fragile sites. *Oncogene*. 1998;**16(18)**:2311–2319. <https://doi.org/10.1038/sj.onc.1200202>
28. Wildschutte JH, Williams ZH, Montesion M, Subramanian RP, Kidd JM, Coffin JM. Discovery of unfixed endogenous retrovirus insertions in diverse human populations. *Proc Natl Acad Sci USA*. 2016;**113(16)**:E2326–E2334. <https://doi.org/10.1073/pnas.1602336113>
29. Feschotte C, Gilbert C. Endogenous viruses: insights into viral evolution and impact on host biology. *Nat Rev Genet*. 2012;**13(4)**:283–296. <https://doi.org/10.1038/nrg3199>
30. Goodier JL. Restricting retrotransposons: a review. *Mobile DNA*. 2016;**7**:16. <https://doi.org/10.1186/s13100-016-0070-z>
31. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. *Nat Rev Genet*. 2009;**10(10)**:691–703. <https://doi.org/10.1038/nrg2640>
32. Lanciano S, Carpentier M-C, Llauro C, Jobet E, Robakowska-Hyzorek D, Lasserre E, et al. Sequencing the extrachromosomal circular mobilome reveals retrotransposon activity in plants. *PLoS Genet*. 2017;**13(2)**:e1006630. <https://doi.org/10.1371/journal.pgen.1006630>
33. Tyler-Smith C, Yang H, Landweber LF, Dunham I, Knoppers BM, Donnelly P, et al. Where next for genetics and genomics? *PLoS Biol*. 2015;**13(7)**:e1002216. <https://doi.org/10.1371/journal.pbio.1002216>
34. Genovese G, Handsaker RE, Li H, Kenny EE, McCarroll StA. Mapping the human reference genome's missing sequence by three-way admixture in Latino genomes. *Am J Hum Genet*. 2013;**93**:411–421. <https://doi.org/10.1016/j.ajhg.2013.07.002>
35. International Human Genome Sequencing Consortium (IHGSC). Initial sequencing and analysis of the human genome. *Nature*. 2001;**409**:860–921. <https://doi.org/10.1038/35057062>
36. Knierim E, Lucke B, Schwarz JM, Schuelke M, Seelow D. Systematic comparison of three methods for fragmentation of long-range PCR products for next generation sequencing. *PLoS One*. 2011;**6(11)**:e28240. <https://doi.org/10.1371/journal.pone.0028240>
37. Poptsova MS, Il'icheva IA, Nechipurenko DY, Panchenko LA, Khodikov MV, Oparina NY, et al. Non-random DNA fragmentation in next-generation sequencing. *Sci Rep*. 2014;**4**:4532. <https://doi.org/10.1038/srep04532>
38. Lister R, Pelizzola M, Dowen RH, Hawkins D, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. 2009;**462**:315–322. <https://doi.org/10.1038/nature08514>
39. Yavartanoo M, Choi JK. ENCODE: a sourcebook of epigenomes and chromatin language. *Genomics Inform*. 2013;**11(1)**:2–6. <https://doi.org/10.5808/GI.2013.11.1.2>
40. Turner KM, Deshpande V, Beyter D, Koga T, Rusert J, Lee C, et al. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature*. 2017;**543(7643)**:122–125. <https://doi.org/10.1038/nature21356>
41. Mishra S, Whetstone JR. Different facets of copy number changes: permanent, transient, and adaptive. *Mol Cell Biol*. 2016;**36**:1050–1063. <https://doi.org/10.1128/MCB.00652-15>
42. Storlazzi CT, Lonoce A, Guastadisegni MC, Trombetta D, D'Addabbo P, Daniele G. Gene amplification as doubleminutes or homogeneously staining regions in solid tumors: origin and structure. *Genome Res*. 2010;**20**:1198–1206. <https://doi.org/10.1101/gr.106252.110>
43. Deininger PL, Batzer MA. Alu repeats and human disease. *Genome Res*. 1999;**67**:183–193. <https://doi.org/10.1006/mgme.1999.2864>
44. Kelley DR, Schatz MC, Salzberg StL. Quake: quality-aware detection and correction of sequencing errors. *Genome Biol*. 2010;**11**:R116. <https://doi.org/10.1186/gb-2010-11-11-r116>
45. Saha S, Rajasekaran S. EC: an efficient error correction algorithm for short reads. *BMC Bioinformatics*. 2015;**16(Suppl 17)**:S2. <https://doi.org/10.1186/1471-2105-16-S17-S2>

46. Heydari M, Miclotte G, Demeester P, Van de Peer Y, Fostier J. Evaluation of the impact of Illumina error correction tools on de novo genome assembly. *BMC Bioinformatics*. 2017;**18**:374. <https://doi.org/10.1186/s12859-017-1784-8>
47. Tom JA, Reeder J, Forrest WF, Graham RR, Hunkapiller J, Behrens TW, et al. Identifying and mitigating batch effects in whole genome sequencing data. *BMC Bioinformatics*. 2017;**18**:351. <https://doi.org/10.1186/s12859-017-1756-z>
48. Pal S, Aluru S. In search of perfect reads. *BMC Bioinformatics*. 2015;**16**(Suppl 17):S7. <https://doi.org/10.1186/1471-2105-16-S17-S7>
49. Knoppers BM, Nguyen MT, Sénécal K, Tassé AM, Zawati MH. Next-generation sequencing and the return of results. *Cold Spring Harb Perspect Med*. 2016;**6**(10):a026724. <https://doi.org/10.1101/cshperspect.a026724>
50. Park ST, Kim Y. Trends in next-generation sequencing and a new era for whole genome sequencing. *Int Neurol J*. 2016;**20**(Suppl 2):S76–83. <https://doi.org/10.5213/inj.1632742.371>
51. Ke R, Mignardi M, Hauling T, Nilsson M. Fourth generation of next-generation sequencing technologies: promise and consequences. *Hum Mutat*. 2016;**37**(12):1363–1367. doi:10.1002/humu.23051
52. Garsed DW, Marshall OJ, Corbin VDA, Hsu A, Di Stefano L, Schroder J. The architecture and evolution of cancer neochromosomes. *Cancer Cell*. 2014;**26**:653–667. <https://doi.org/10.1016/j.ccell.2014.09.010>