Aneuploidy in the human brain and cancer
van den Bos, Hilda

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2017

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 05-08-2019
Single-cell sequencing to quantify genomic integrity in cancer

Hilda van den Bos, Bjorn Bakker, Diana CJ Spierings, Peter M Lansdorf, Floris Foijer

The International Journal of Biochemistry & Cell Biology, in press
Abstract
The use of single-cell DNA sequencing (sc-seq) techniques for the diagnosis, prognosis and treatment of cancer is a rapidly developing field. Sc-seq research is gaining momentum by decreased sequencing costs and continuous improvements in techniques. In this review, we provide an overview of recent advancements in the field of sc-seq in cancer and we discuss how sc-seq can contribute to improved care for cancer patients. Sc-seq has made it possible to study the genomes of individual cancer cells from primary tumors, metastases and circulating tumor cells, revealing inter- and intra-tumor heterogeneity, which cannot be detected using other methods. We review studies on individual human cancer cells in relation to prognosis and treatment response. Finally, future perspectives of sc-seq in cancer diagnosis and treatment are discussed with a focus on the use of circulating tumor cells to monitor therapy response and the development of personalized treatments based on knowledge about the genomic heterogeneity.

Keywords: single-cell sequencing, intra-tumor heterogeneity, tumor evolution, circulating tumor cells

Highlights
- Single-cell sequencing allows in depth analysis of the genomes of single tumor cells, which is impossible using other techniques
- Single-cell sequencing allows analysis of rare cell types such as circulating tumor cells
- Single-cell sequencing may provide future applications in the diagnosis, monitoring and treatment of cancer
Introduction
Classical sequencing of DNA extracted from millions of tumor cells provides useful information about point mutations and copy number alterations (CNAs) that are present in most tumor cells. Although tumor heterogeneity can be identified to a certain extent by sequencing multiple regions of a tumor, mutations and CNAs present at low frequencies cannot be reliably identified by bulk sequencing. In contrast, sequencing of single cells (sc-seq) allows in-depth analysis of the genomes of normal and tumor cells. As the cost of sequencing the genomes of individual cells continues to decrease and the protocols become more easily accessible, the range of applications increases and the potential of sc-seq is starting to be uncovered. Such applications include quantification of aneuploidy, smaller CNA, and the mutational landscapes and thereby the genomic heterogeneity of a tumor. With this information, phylogenetic trees can be built to unravel tumor evolution. Also, rare cells isolated from biopsies or circulating tumor cells can be analyzed.

In this review, we discuss how single-cell DNA sequencing can contribute to the diagnosis, treatment and prognosis of cancer patients. Over the last several years many single-cell DNA sequencing protocols have been developed. The majority of these protocols depend on whole genome amplification (WGA) before library construction. Several WGA methods are available, each having its own advantages and disadvantages (reviewed in 3). Although WGA is necessary to generate sufficient genomic coverage for analysis of point mutations, it comes at the cost of introducing PCR amplification biases, which can obscure the detection of CNAs. To circumvent this, protocols have been developed that do not require WGA. These protocols eliminate the amplification bias, but only allow for shallow sequencing and are therefore suitable for detection of CNAs, but not for mutation analysis 3,4. A major advantage of these pre-amplification-free methods is the much lower price per cell and higher throughput, allowing for many more cells to be analyzed at once. Finally, a recently published protocol allows single cell CNA analysis on formalin fixed, paraffin embedded (FFPE) tissue 5. Since it is common practice to store FFPE samples from tumors, this protocol opens up the opportunity to analyze an extensive repository of samples. Besides single-cell DNA sequencing, other sequencing methods that can advance cancer research have been developed, such as single-cell RNA sequencing, droplet PCR and sequencing of circulating cell free DNA. Although these methods can greatly contribute to the improvement of cancer diagnosis and monitoring, it is beyond the scope of this review to discuss these in detail.

Single cell aneuploidy, CNA, and mutation analysis
Since the first report on whole genome sequencing of single breast cancer cells 6, a series of single cell sequencing studies have been published analyzing many different cancer types. A summary of these studies is given in Table 1. Sc-seq can reveal aneuploidy, CNAs and/or mutations in individual cancer cells from primary tumors, metastases and other (rare) tumor cell sources, such as circulating tumor cells (CTCs) and malignant pleural effusion. The heterogeneity of a tumor and/or its metastasis and CTCs can be determined, and the presence of one or more clones and evolutionary history can be mapped. Sc-seq does not only allow the CNA pattern in a tumor or metastasis to be revealed, but also allows studies on the timing or order in which CNAs arose and subclones emerged. By sampling a tumor, its metastasis or CTC’s multiple times, the response to treatment can be monitored.
over time (Figure 1). Several studies have investigated the clonal evolution of tumor cells by building phylogenetic trees from this sc-seq information. For example, the group of Navin has shown that breast cancer tumors experience short burst of aneuploid rearrangements followed by stable clonal expansion. Using single-cell whole-exome sequencing on two colorectal cancer (CRC) patients, Wu et al. demonstrated that although these CRC tumors were monoclonal in origin, several subclones were found based on accumulation of novel driver mutations.

With the use of single-cell DNA sequencing questions about the mutation frequency in tumors can be answered. Even though tumor cells typically have a higher mutation frequency than somatic cells, it has been difficult to distinguish whether this is caused by an increased mutation rate or a higher proliferation rate of tumor cells as the proliferation rate is difficult to measure. While bulk sequencing has suggested a 210-fold increase in mutation rate in tumor cells compared to somatic cells, sc-seq suggests this rate to be much lower: at 1-13.3 times the normal mutation rate. However, since only a few patients and cancer types have been investigated so far, more studies are needed to draw more general conclusions.

Circulating and disseminated tumor cells
One of the strengths of sc-seq is the possibility to analyze rare cells, such as cells isolated from a small tumor biopsy or circulating tumor cells (CTCs). The analysis of CTCs is a very promising development in the monitoring of cancer. It provides a minimal invasive way to monitor the progression of a tumor as well as the response to treatment. CTCs are isolated from the blood using an antibody staining for EpCAM and CD45, and/or by size. The number of CTCs per ml of blood already has prognostic value: high numbers of CTCs are associated with poor prognosis in several cancer types, e.g. breast, prostate, lung and colorectal cancer (reviewed in ). Subsequent sequencing of CTCs and comparing them to primary and metastatic tumor cells yields further valuable information. Questions such as, ‘Do the CTCs have a specific karyotype/mutation landscape that allows these cells to disseminate’ or ‘Are cells at random released from the tumor?’ can be answered. Heitzer et al. showed that CTCs in metastatic colon cancer have similar CNA and mutation profiles. Other studies confirm similarity between the CNA profile of CTCs and the primary tumor in lung and prostate cancer. In contrast, Gao et al. found the CNA pattern in CTCs to be more homogeneous than in the primary tumor of colon, breast, gastric and prostate cancer. In addition, they found the CNA pattern to resemble the pattern found in metastases, suggesting that only a certain CNA pattern allows tumor cells to be released from the primary tumor and initiate the growth of metastases. Demeulemeester et al. sequenced disseminated single tumor cells (DTCs), isolated from the bone marrow of 6 breast cancer patients. Based on CNA analysis and matching phylogenetic trees the authors concluded that these DTCs separated from the primary tumor relatively late.
The authors concluded that these DTCs separated from the primary tumor relatively late. Based on CNA analysis and matching phylogenetic trees, the authors demonstrated that although these CRC tumors resemble the pattern found in metastases, suggesting that only a certain CNA pattern allows tumor cells to be released from the primary tumor and initiate the growth of metastases. This is consistent with previous studies showing that CTCs can be found in the bloodstream of patients with various types of cancer, including colon, breast, gastric, and prostate cancer. In addition, they found the CNA pattern to be more homogeneous than in the primary tumor, which can be helpful in distinguishing between primary tumors and metastases. One of the strengths of single-cell sequencing is the possibility to analyze rare cells, such as CTCs, and/or by size-fractionating sorting. This allows for a minimal invasive way to monitor the development in the monitoring of cancer. It provides a valuable information of CTCs and comparing them to primary and metastatic tumor tissue.

Table 1. Summary studies using single cell sequencing in cancer

<table>
<thead>
<tr>
<th>Cancer type</th>
<th>Tissue</th>
<th># patients</th>
<th># single cells sequenced, before/ after QC</th>
<th>WGA method</th>
<th>Mutations /CNAs</th>
<th>Region sequenced</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute lymphoblastic leukemia</td>
<td>Bone marrow</td>
<td>6</td>
<td>1479 cells total / ~50% passing QC</td>
<td>MDA (Genome Phiv2)</td>
<td>Mutations</td>
<td>Targeted sequencing</td>
<td>17</td>
</tr>
<tr>
<td>Acute lymphoblastic leukemia</td>
<td>Bone marrow</td>
<td>3</td>
<td>288 cells total / 214 passing QC</td>
<td>none</td>
<td>CNAs</td>
<td>Whole genome</td>
<td>18</td>
</tr>
<tr>
<td>Acute myeloid leukemia</td>
<td>Bone marrow</td>
<td>3</td>
<td>36 cells total / 35 passing QC</td>
<td>LA-PCR (PicoPLEX)</td>
<td>Mutations</td>
<td>Targeted sequencing: &gt;1,900 loci</td>
<td>19</td>
</tr>
<tr>
<td>Bladder carcinoma</td>
<td>Primary tumor</td>
<td>1</td>
<td>66 cells total / 44 passing QC</td>
<td>MDA (REPLI-g)</td>
<td>Mutations</td>
<td>Exome</td>
<td>20</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Primary tumor and liver metastasis</td>
<td>2</td>
<td>200 cells total/ QC NA</td>
<td>DOP-PCR (Genome Plex)</td>
<td>CNAs</td>
<td>Whole genome</td>
<td>6</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Primary tumor</td>
<td>2</td>
<td>195 cells total/ QC NA</td>
<td>MDA (REPLI-g)</td>
<td>Mutations and CNAs</td>
<td>Exome and whole genome</td>
<td>10</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Primary tumor</td>
<td>2</td>
<td>384 cells total/ 332 passing QC</td>
<td>DOP-PCR (Seqplex)</td>
<td>CNAs</td>
<td>Whole genome</td>
<td>21</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>DTCs in bone marrow</td>
<td>6</td>
<td>63 cells total / 45 passing QC</td>
<td>DOP-PCR (Genome Plex)</td>
<td>Mutations and CNAs</td>
<td>Whole genome</td>
<td>9</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Primary tumor</td>
<td>12</td>
<td>1119 cells total/ QC NA</td>
<td>DOP-PCR (Genome Plex)</td>
<td>CNAs</td>
<td>Whole genome</td>
<td>12</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Primary tumor</td>
<td>4</td>
<td>676 cells total/ QC NA</td>
<td>DOP-PCR (Genome Plex)</td>
<td>CNAs</td>
<td>Whole genome</td>
<td>5</td>
</tr>
<tr>
<td>Colon cancer</td>
<td>Primary tumor</td>
<td>1</td>
<td>Total NA/ 67 passing QC</td>
<td>MDA (REPLI-g)</td>
<td>Mutations</td>
<td>Exome</td>
<td>22</td>
</tr>
</tbody>
</table>
Table 1. (continued)

<table>
<thead>
<tr>
<th>Cancer type</th>
<th>Tissue</th>
<th># patients</th>
<th># single cells sequenced, before/ after QC</th>
<th>WGA method</th>
<th>Mutations /CNAs</th>
<th>Region sequenced</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colon cancer</td>
<td>Primary tumor</td>
<td>2</td>
<td>165 cells total / 124 passing QC</td>
<td>MDA (REPLI-g)</td>
<td>Mutations and CNAs</td>
<td>Exome</td>
<td>11</td>
</tr>
<tr>
<td>Colon, breast, gastric and prostate cancer</td>
<td>CTCs and primary tumor</td>
<td>23</td>
<td>102 cells total / QC NA</td>
<td>MALBAC</td>
<td>CNAs</td>
<td>Exome and Whole genome</td>
<td>8</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>CTCs</td>
<td>11</td>
<td>68 cells total / QC NA</td>
<td>MALBAC</td>
<td>Mutations and CNAs</td>
<td>Exome</td>
<td>23</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>CTCs</td>
<td>2</td>
<td>8 cells total / QC NA</td>
<td>LA-PCR (Ampli1)</td>
<td>CNAs</td>
<td>Whole genome</td>
<td>24</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>CTCs</td>
<td>31</td>
<td>253 cells total / QC NA</td>
<td>LA-PCR (Ampli1)</td>
<td>CNAs</td>
<td>Whole genome</td>
<td>25</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>Primary tumor and metastases</td>
<td>1</td>
<td>586 cells total / 346 passing QC</td>
<td>None</td>
<td>CNAs</td>
<td>Whole genome</td>
<td>26</td>
</tr>
<tr>
<td>Myeloproliferative neoplasm</td>
<td>Bone marrow</td>
<td>1</td>
<td>90 cells total / 58 passing QC</td>
<td>MDA (REPLI-g)</td>
<td>Mutations</td>
<td>Exome</td>
<td>27</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>CTCs</td>
<td>5</td>
<td>138 cells total / 42 passing QC</td>
<td>MDA (RepliPHI)</td>
<td>Mutations</td>
<td>Exome</td>
<td>28</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>CTCs</td>
<td>1</td>
<td>41 cells total / QC NA</td>
<td>DOP-PCR (Genome Plex)</td>
<td>CNAs</td>
<td>Whole genome</td>
<td>29</td>
</tr>
<tr>
<td>Renal carcinoma</td>
<td>Primary tumor</td>
<td>1</td>
<td>25 cells total / QC NA</td>
<td>MDA (REPLI-g)</td>
<td>Mutations</td>
<td>Exome</td>
<td>30</td>
</tr>
</tbody>
</table>

1 Control cells from cell lines are not included.

**Figure 1. Single cell sequencing in cancer.** By sequencing (rare) tumor cells or circulating tumor cells, aneuploidy, copy number variations and/or mutations can be identified. This enables characterization of the heterogeneity of a tumor, identification of its evolutionary path, and allows for monitoring of treatments.

Bone marrow DTCs can form a dormant reservoir that evades therapy, and might cause metastasis over time. Interestingly, sc-seq revealed that only ~50% of the cells that were previously identified as tumor cells based on morphological features, were actual tumor cells. Moreover, a number of studies have found that CTCs from certain cancer types have specific CNA patterns, which might contribute to diagnostics. A study comparing CTCs from patients with breast, gastric, prostate or colon cancer revealed reproducible CNA patterns among patients with the same cancer type, with the exception of breast cancer. The notion that breast cancer is a multi-subtype disease might explain the difference with the other cancer types studied. These observations fit well with recent mouse models for CIN cancer that show that chromosomal instability leads to aneuploid cancers that exhibit recurrent karyotypes that are tissue-specific. Another study found similar CNA patterns in different patients with lung adeno carcinoma. Therefore, CNA pattern in CTCs might reveal the type of cancer as well as shed light on the different cancer subtypes. Moreover, CTCs might be suitable for...
non-invasive tumor monitoring over time. Dago et al. analyzed CTCs from a prostate cancer patient before and after treatment 29. Indeed, two therapy resistant clones were identified after treatment, one of which was already present before treatment. Importantly, another study revealed that CTCs can identify which patients will or will not respond to treatment as researchers were able to discriminate chemo sensitive from chemo refractory small-cell lung cancer based on the CNV patterns of CTCs 25.

In summary, sequencing CTCs holds great potential. However, it must be taken into account that capturing these cells has so far been difficult and needs further optimization. Furthermore, not all tumor cells might be captured, such as EpCAM negative tumor cells. Therefore, we need to further invest in methodologies to isolate and analyze CTCs, as they will become an essential tool in the future to monitor disease onset, progression and therapy response.

Conclusion and future perspectives

We have only just started to explore the great potential that sc-seq of cancer cells offers for expanding our knowledge on e.g. tumor evolution, metastatic potential, monitoring therapy response, and development of therapy resistance. Sc-seq can contribute to unravelling fundamental mechanisms of tumor formation and progression: the order in which tumor cells acquire CNAs and mutations can help to identify driver events. Sequencing of CTCs has the potential to provide a non-invasive method for early cancer diagnosis, monitoring therapy response and relapse, although CTCs are not found in all cancer patients.

The continuing decrease in sequencing costs, together with improved library preparation and robust whole genome amplification protocols, as well as the constantly improving toolboxes for bioinformatics analysis all help to fully exploit the potential of sc-seq. Together, these important developments will help to increase the number of cells analyzed per study as well as the number of patients included. Such larger studies involving more patients, studying larger numbers of cells, and sampling at multiple time points will allow us to study the evolution of cancers and the role of intra tumor heterogeneity at unprecedented resolution in the near future.

Funding

This work was supported by a Dutch Cancer Society grant (KWF grant RUG-2012-5549) to FF and PML, an Advanced ERC grant to PML and a Groningen Foundation for Paediatric Oncology (SKOG) grant to FF and PML.

References

5. Martelotto, L. G. et al. Whole-genome single-cell copy number profiling from formalin-fixed
29. Dago, A. E. et al. Rapid phenotypic and genomic change in response to therapeutic pressure in


Copy number alterations assessed at the single-cell level revealed mono- and polyclonal seeding patterns of distant metastases in a small cell lung cancer patient.

Hilda van den Bos*, Paranita Ferronika*, Aaron Taudt, Diana C.J. Spierings, Ali Saber, Thijo J.N. Hiltermann, Klaas Kok, David Porubsky, Anthonie J. van der Wekken, Wim Timens, Floris Foijer, Maria Colomé-Tatché, Harry J.M. Groen, Peter M. Lansdorp, and Anke van den Berg

*These authors contributed equally to the paper

Adapted from Annals of Oncology (2017) April 11