

# Chapter 2

## Ultradense Array CGH and Discovery of Micro-Copy Number Alterations and Gene Fusions in the Cancer Genome

Ewa Przybytkowski, Adriana Aguilar-Mahecha, Sheida Nabavi,  
Peter J. Tonellato, and Mark Basik

### Abstract

The characterization of molecular alterations specific to cancer facilitates the discovery of predictive and prognostic biomarkers important to targeted therapeutics. Alterations critical to cancer therapeutics include copy number alterations (CNAs) such as gene amplifications and deletions as well as genomic rearrangements resulting in gene fusions. There are two genome-wide technologies used to detect CNAs: next generation sequencing (NGS) and dense microarray based comparative genomic hybridization, termed array CGH (aCGH). aCGH is a mature robust technology of lower cost and more accessible than NGS. This chapter describes the protocol steps and analysis required to obtain reliable aCGH results from clinical samples. Technical options and various necessary compromises related to the nature of clinical material are considered and the consequences of these choices for data analysis and interpretation are discussed. The chapter includes brief description of the data analysis, even though analysis is often performed by bioinformaticians. Today's cancer research requires collaboration of clinicians, molecular biologists, and mathematicians. Acquaintance with the basic principles related to the extraction of the data from arrays, its normalization and the algorithms available for analysis provides a baseline for mutual understanding and communication.

**Key words:** Breast cancer, Needle biopsies, Aneuploidy, Whole genome amplification, Array CGH, Copy number alterations, Gene fusions, Micro-aberration, Feature extraction, ADM-2

---

## 1. Introduction

### 1.1. Application of Array CGH in Cancer Research

Technological advances have dramatically increased our capacity to analyze and reveal a multitude of changes in cancer cells. The characterization of molecular alterations specific to cancer has enabled the discovery of novel predictive and prognostic biomarkers now an integral part of the development of novel targeted therapeutics. Molecular alterations critical to cancer therapeutics

include copy number alterations (CNAs) such as gene amplifications and deletions as well as genomic rearrangements resulting in gene fusions. DNA amplifications have been shown to contain important druggable oncogenes, such as the genes encoding for the HER2 and EGF receptors (1, 2). The discovery of chromosomal translocations in solid tumors, such as that involving the *ALK* gene resulting in a novel oncogenic fusion protein in lung adenocarcinoma, have also led to promising novel therapies directed against these changes (3, 4).

There are two genome-wide technologies used to detect CNAs associated with cancer: massively parallel next generation sequencing (NGS) (5) and microarray based comparative genomic hybridization (aCGH) (6). NGS requires access to expensive NGS platforms or services, extensive novel software and bioinformatics support, large computer storage and computational resources, uses considerable quantities of genomic DNA (>5 µg), and is therefore not easily accessible or of low cost. Even when the sequencing cost decreases, the computational resource and bioinformatics needs will remain. Array CGH (aCGH) is a mature technology with low cost competitive products, stable commercially produced software and standardized protocols and therefore provides a low cost, robust, and more easily accessible technology readily available to most wet lab scientists. In addition, aCGH allows the accurate characterization of gene copy number using as little as 0.5 µg of genomic DNA (7). Such sensitivity becomes important when one considers that genomic technologies are increasingly applied to minute tumor samples such as those obtained from clinical biopsies.

aCGH is a microarray-based technique detecting the relative efficacy with which two differentially fluorescently labeled DNAs (test and reference DNA) hybridize to the array of oligonucleotide probes immobilized on a slide. The probe sequences are generally designed to cover the entire genome with precision of coverage defined by the probe density across the genome. aCGH technology and probe density continue to rapidly improve allowing accurate detection of very small copy number variations (CNVs). Currently, ultradense arrays (also called high-resolution arrays) contain over 200,000 probes and interrogate the genome with an average resolution of 10–50 kb (8). Ultrahigh resolution allows better delineation of DNA breakpoints at regions of CNAs as well as the identification of very small, focal CNAs in addition to the large chromosomal regions involving several Mb thus bridging the gap between conventional cytogenetic technologies, which address gross, microscopic chromosomal aberrations (>1 Mb) and sequencing technologies, which detect variations at the level of a single nucleotide.

There are many commercially available arrays with varying genome coverage and resolution, but as yet, there are no clear standards for reporting CNAs in the cancer genome and only a few studies that discuss the comparison between platforms and the reproducibility of the data (7, 9).

Obtaining reproducible, high-resolution copy number data with high sensitivity and few false positives is the primary objective for any high throughput experimental study. According to the most comprehensive analysis published so far, reproducibility of replicate experiments of aCGH for finding CNV in DNA from normal healthy individuals is <70% (9) and may decrease even further when cancer biopsy samples are analyzed or when test DNA is subjected to whole genome amplification (WGA) prior to labeling and hybridization (8). Thus, the processing and analysis of the data must be carefully performed (see Subheading 3.8) and interpretation of results must take into consideration the quality of the biological material, the nature of the experimental protocols as well as analytical methods (see Subheading 3.9).

### **1.2. Methods for Array CGH Testing of Clinical Cancer Samples**

High throughput techniques such as ultradense aCGH are not only susceptible to the “garbage in-garbage out” principle but the vast amount of data produced in a semi-automated fashion amplifies the potential complications of producing poor quality data resulting in flawed interpretation. This is especially true when dealing with clinical samples which are often very small, sometimes partially degraded, and generally of heterogeneous composition. Several steps can be taken to avoid the “garbage-in” problem and thus vastly improve the likelihood of producing high quality data and corresponding interpretation: (1) Careful assessment of sample quality, (2) Rigorous quality control for each step in sample processing, (3) meticulous experimental procedure, and (4) well informed and statistically accurate data processing and analysis. These steps are reviewed herein and will lead to confidence in the data, results, and interpretation.

Clinical cancer samples are either formalin-fixed paraffin embedded (FFPE) specimens or snap-frozen tissues of tumor pieces or tumor biopsies. FFPE specimens provide relatively large numbers of archived samples with known clinical outcomes, but the quality of DNA obtained from these specimens is usually very poor (10). In this chapter we describe the use of snap-frozen tissues from core needle biopsies (see Subheading 3.1). Snap-frozen biopsies are an excellent source of high quality DNA, usually in sufficient amounts for aCGH.

An important cancer sample issue is tumor heterogeneity. Tumorigenesis is a micro-evolution in which variants of cells clonally expand within a tissue. This dynamic process results in genetic diversity within tumors and marked changes in tissue architecture. Thus tumor samples typically contain various proportions of tumor cells and normal cells as well as components of stroma and infiltrating lymphocytes (11). Consequently, DNA from tumor samples consists of a pool of heterogeneous DNA. To overcome the pooling problem and improve DNA homogeneity tumor tissue slices can be micro-dissected and only tumor cells selected for further analysis (12). This

strategy is often applied to FFPE samples. When fresh frozen biopsies are used, the pathologist may first screen the biopsies, and retain only those, which contain a high percentage of tumor cells (see Subheading 3.1). In the protocol shown in this chapter we describe an additional strategy to overcome tumor heterogeneity, using flow sorting of fluorescently labeled nuclei according to ploidy (see Subheadings 3.2 and 3.3). This procedure allows the selection of relatively pure tumor cell subpopulations especially for aneuploid tumors (13, 14). Diploid fractions, however, are less reliable since they may represent diploid tumor cells as well as normal contaminants.

While flow sorting improves tumor sample homogeneity and therefore the quality of DNA samples for further analysis, the post-cell sorted quantities of DNA are often insufficient for aCGH protocols (see Subheading 3.4 and Note 7) so WGA may be required (see Subheading 3.5). There are excellent WGA protocols (15), but all can introduce bias since various genome sequences have different amplification efficiencies (16, 17). The consequences of WGA seem to be more relevant when denser arrays are used and when analysis focuses on micro-aberrations (8).

aCGH requires a reference genomic DNA. The best reference for studies of cancer genome would be matched normal DNA from the same individual. However, this may be difficult to obtain for every clinical sample. Thus, in a protocol described here we used commercially available normal human genomic DNA (see Subheadings 3.5, 3.6, and 3.7). This compromise is not without consequences for data interpretation. We briefly discuss this problem in a section on data interpretation (see Subheading 3.9).

Finally, considering the potentially confounding factors of aCGH analysis of clinical cancer samples, validation of copy number changes with other molecular biology techniques such as polymerase chain reaction (PCR) is very important. Methods and strategies applied for validation of the aCGH results are not covered in this chapter.

---

## 2. Materials

### 2.1. General Lab Equipment

Mini-Sub DNA Cell Bio-Rad and power supply (Bio-Rad Power Pac 3000).

Gel-Doc EZ imaging system with UV tray (Bio-Rad).

DNA 120 SpeedVac Concentrator (Thermo Electron Corp.)

NanoDrop Spectrophotometer ND-1000. FLUOstar OPTIMA microplate reader (BMG LABTECH) equipped with 485 nm excitation filter and 520 nm emission filter (for PicoGreen Assay).

Microcentrifuge.

Vortex mixer.

Microwave oven.

Magnetic stir plate (2×) and stir bars (2×).

Single-channel pipettes (P10, P20, P200, P1000).

Sterile, low binding, barrier tips 100, 200, 20, and 10 µl.

Ice bucket.

Timer.

Circulating water bath.

Heat block.

Nuclease-free Eppendorf tubes (autoclaved). MICROTEST Flat Bottom Polystyrene, non-tissue culture treated 96 well plates (Falcon #351172) for PicoGreen Assay.

Powder-free gloves (latex) (see Note 1).

Sterile tweezers.

## **2.2. General Lab Reagents**

Molecular biology grade water from Milli-Q Synthesis system (Millipore) (autoclaved).

1× PBS, prepared from 10× stock (Fisher BioReagents #BP399-20).

0.5 M EDTA (Fisher BioReagents BP2483-500).

1× TE (pH 8.0), Molecular grade (Promega #V6231).

EtOH (100/96%).

UltraPure Agarose (Invitrogen #15515-027).

Nucleic Acid Stain SYBR safe (Invitrogen #S33102).

Gene Ruler 1 kb Ladder (Fermentas #SM0311).

## **2.3. For Biospecimen Collection and Processing**

Portable liquid nitrogen canister filled with liquid nitrogen.

Powder-free gloves (latex).

Bar-coded cryovials.

Sterile tweezers.

Cryogenic labels and permanent marker.

Log sheets.

Dry ice.

Standard pathology equipment and materials: cryomolds, optimal cutting temperature (O.C.T) compound, isopentane (2-methyl butane), cryostat, blades, glass slides, hematoxylin and eosin (H&E), etc.

## **2.4. If Sorting of Nuclei is Performed**

Sterile 35 × 10 mm petri dishes and sterile tweezers.

Extraction buffer: 10 mM Tris HCl pH 7.4, 146 mM NaCl, 2 mM CaCl<sub>2</sub>, 22 mM MgCl<sub>2</sub>, BSA 0.005%, Igepal CA-630 0.1%.

Sterile single use stainless surgical blades (Lance Paragon LTD, Sheffield, England).

20G×1 needles and 1 ml syringes.

DAPI stock solution, 10 mg/ml in DMSO (1,000×; final concentration of DAPI is 10 µg/ml).

DAPI working solution (100×), DAPI stock diluted in extraction buffer.

FACS tubes: BD Falcon 352235 12×75 mm, 5 ml polystyrene round bottom test tubes with a 35 µm nylon mesh cell-strainer cap.

BD FACSAria flow cytometer sorter with blue (488 nm) and violet (408 nm) lasers and 450/40 emission filter.

**2.5. For Isolation  
of DNA from Nuclei**

QIAmp DNA MicroKit (Qiagen #56304).

RNase A (100 mg/ml) (Qiagen #19101). Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen/Molecular Probes #P11496)

**2.6. For DNA  
Amplification**

Human Genomic DNA (*Female*): Promega #G1521.

Illustra GenomiPhi V2 DNA Amplification Kit GE Healthcare #256600-31.

**2.7. For DNA Labeling  
and Purification**

Bioprime Array CGH genomic labeling kit (Invitrogen #45-0048).

Fluorescent nucleotides Amersham; CY3-dUTP #CA95040-146L. CY5-dUTP #CA95040-158L.

Microcon Centrifugal Filters Ultracel YM-30 (Millipore #42410).

**2.8. For Hybridization,  
Washing, and  
Scanning of the Image**

SurePrint HD CGH Microarray Kits: G4411B Human Genome CGH Microarray Kit 1×244 K.

SurePrint G3 CGH Microarray Kits: G4447A SurePrint G3 Human CGH Microarray Kit 1×1 M (see Note 2).

Agilent High-Resolution Microarray Scanner Bundle G2565CA (include: Scanner, PC and LCD Monitor, Barcode reader and Feature extraction software). We used upgraded B scanner.

Hybridization chamber gasket slides, 5-pack (Agilent #G2534-60003).

Hybridization oven (Agilent #G2545A) and Hybridization oven rotator for Agilent Microarray hybridization chambers (Agilent #G2530-60029).

Hybridization Chambers (Agilent #G2534A).

Ozone-barrier slide covers (Agilent #G2505-60550).

Cot-1 DNA 1 mg/ml (Invitrogen #15279-101).

Agilent Oligo aCGH Hybridization Kit (Agilent #5188-5380).

Agilent Oligo aCGH Wash Buffer 1 and 2 set (Agilent #5188-522).

250 ml rectangular slide staining dishes (Wheaton).

1.5 l flat Pyrex baking dish for warming up the dish with washing buffer #2.

### 3. Methods

#### **3.1. Biopsy Collection and Processing**

Ideally, the clinical specimen should be preserved as soon as it is removed from the patient. The setting where it is relatively easy to do this is in the context of tissue biopsies. Biopsies for research purposes can be collected at the time of diagnosis or during the course of the disease.

1. The day the procedure is scheduled prepare all the material necessary for biospecimen collection (see Subheading 2.3), mark cryogenic labels with patient identifier and bring all to the room where the biopsy will take place.
2. When the specimen has been removed from the patient, take the core of tissue from the needle using sterile tweezers and put it into a pre-labeled bar-coded cryovial. Tightly close the cryovial and immediately submerge it into liquid nitrogen (see Note 3).
3. The sample can be stored for future analysis ( $-80^{\circ}\text{C}$  freezer or vapor phase of liquid nitrogen tank) or processed immediately.
4. Use a log sheet to write down the time of collection, preservation, long-term storage, or any deviation to the standard operation procedure (SOP).
5. The percent tumor cells present in the biospecimen has to be evaluated by a pathologist, thus the following steps need to be performed in a standard pathology lab.
6. Place on dry ice a beaker containing precooled isopentane (2-methyl butane) and the cryomold, let both chill.
7. Carefully remove the frozen biopsy from the cryovial using sterile tweezers, place on the chilled cryomold and cover with O.C.T compound avoiding bubbles.
8. Take the cryomold with forceps and submerge into the ice-cold isopentane for approximately 30 s (until the O.C.T has solidified).
9. Mount the O.C.T block containing the biopsy on to the cryostat (set between  $-25$  and  $-30^{\circ}\text{C}$ ) and proceed with cryosectioning of  $5\text{ }\mu\text{m}$  sections (standard pathology procedure).
10. Mount  $5\text{ }\mu\text{m}$  sections on glass slides and proceed with H&E staining (standard pathology protocol).
11. Have the H&E slides evaluated by a pathologist. Specimens containing at least 70% tumor cells are accepted for sorting experiments (see Note 4 and 5).



### **3.2. Preparation of Biopsies for Sorting Nuclei (See Note 6)**

1. Place frozen O.C.T. block containing biopsy on a small sterile petri dish.
2. Thaw O.C.T. block at room temperature until O.C.T. becomes soft and transparent.
3. Using sterile tweezers pull out biopsy from O.C.T. and place it on another sterile petri dish.
4. Add 100  $\mu$ l of extraction buffer to the biopsy.
5. Mince the tissue using a surgical scalpel blade.
6. Add 100  $\mu$ l of extraction buffer and continue mincing the tissue (see Note 7).
7. Transfer the buffer and debris to a sterile Eppendorf tube and put it on ice.
8. Wash petri dish with another 100  $\mu$ l of extraction buffer and add it to the Eppendorf tube.
9. Pass the suspension in the Eppendorf tube through a 20G needle five times using a syringe to disaggregate nuclei.
10. Add 3  $\mu$ l of a DAPI working solution (100 $\times$ ) per 300  $\mu$ l of extraction buffer and incubate on ice for 3 min.
11. Filter sample through the mesh (35  $\mu$ m) into the FACS tube.
12. Wash the filter with 200  $\mu$ l of extraction buffer containing 1 $\times$  DAPI.

### **3.3. Flow Sorting of Nuclei**

1. Nuclei are sorted according to DAPI intensity using BD FACS Aria flow cytometer/sorter. Excitation is achieved with violet laser (408 nm) and emission of DAPI is detected with a 450/40 filter (see Note 8).
2. For optimal DNA yield, extraction from at least 20,000 nuclei is recommended (see Note 9). Figure 1 shows two examples of sorting performed with different specimens.
3. After sorting keep the nuclear suspension on ice. Do not freeze and proceed with the isolation of DNA as soon as possible.

### **3.4. Isolation of DNA from Nuclei and Determination of DNA Quality**

It would be useful at this point to read carefully the manual published by Agilent, which describes and discusses in depth the experimental protocols and various options (18).

The following sections describe experimental choices used in our laboratory.

DNA from nuclei is isolated using QIAmp DNA MicroKit (Qiagen #56304) and a protocol designed by the manufacturer for genomic DNA isolation from tissues with some modifications.

1. Measure the volume of the nuclear suspension; if larger than 80  $\mu$ l, the sample needs to be concentrated. This can be achieved by centrifugation (microcentrifuge at maximum speed, 13,000 rpm (16000 $\times g$ ) for 25 min at 4  $^{\circ}$ C) (see Note 10).



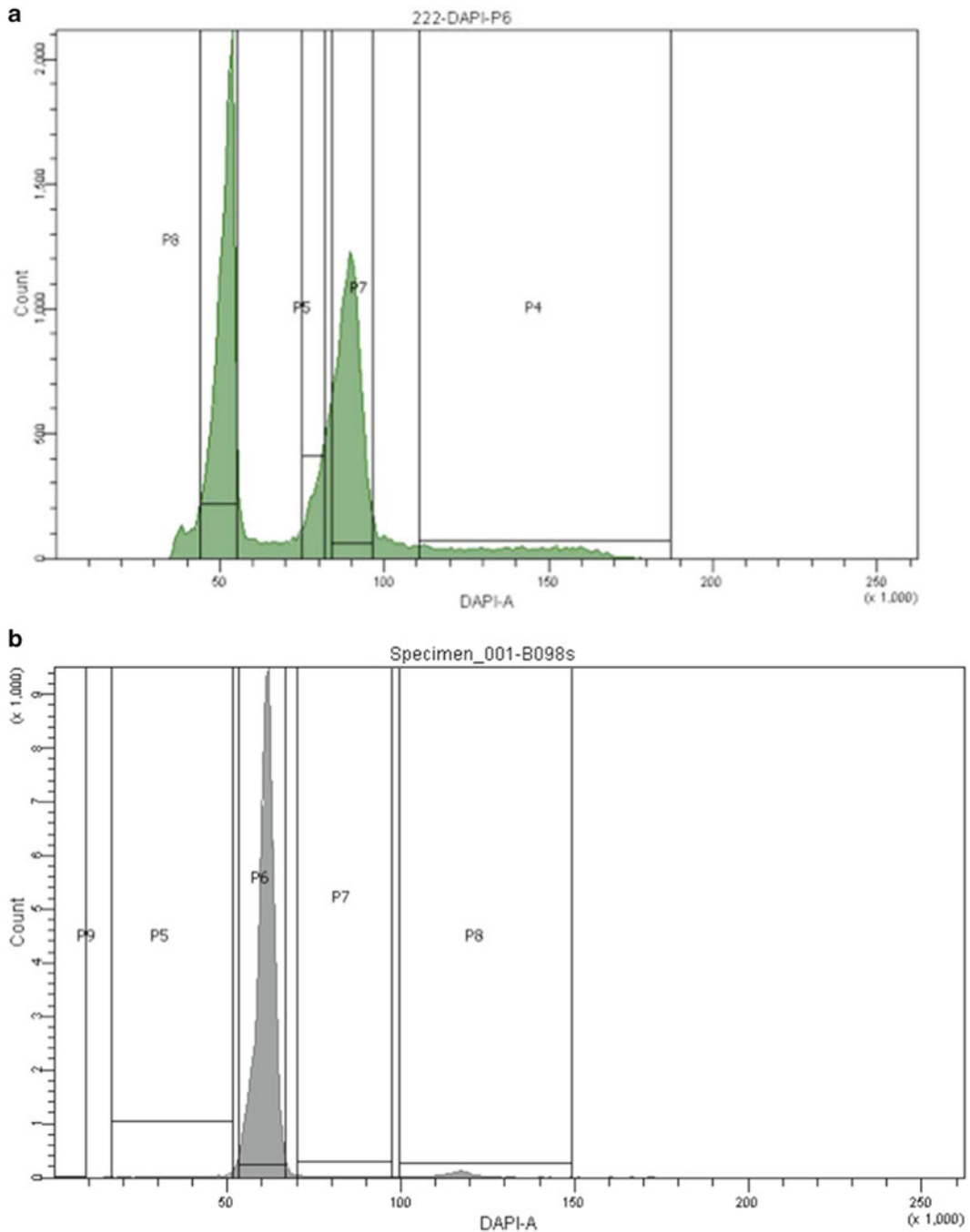


Fig. 1. FACS profiles of specimens. The numbers of nuclei are plotted against intensity of fluorescence from DAPI (a) Tumor specimen T-222 had two clearly defined subpopulations of nuclei, diploid P8 (105,246 nuclei were collected) and aneuploid P7 (96,166 nuclei were collected). We collected also two other, less defined subpopulations, P4 and P5. The number of nuclei in these two subpopulations was sufficient for DNA extraction (19,770 and 20,043 respectively). However, often this may not be the case and sufficient numbers of nuclei can be obtained only from one or two main subpopulations (b) Needle biopsy B-098 contained only one main diploid tumor subpopulation P6, which yielded 11,343,376 nuclei. The number of nuclei collected from P7 and P8 subpopulations was not sufficient to obtain enough DNA for array CGH analysis (9,600 and 3,891 respectively).

2. Add 100  $\mu\text{l}$  of buffer ATL to the 50  $\mu\text{l}$  suspension of nuclei to obtain total volume of 180  $\mu\text{l}$ .
3. Add 20  $\mu\text{l}$  of Proteinase K (provided in the kit), mix gently (see Note 11).
4. Incubate at 56 °C for 2–3 h (see Note 12).
5. Cool down the sample, centrifuge briefly to remove drops from inside the lid and add 4  $\mu\text{l}$  of RNase A, mix gently and incubate at room temperature for 2 min (digestion with RNase A is optional).
6. Add 200  $\mu\text{l}$  of AL buffer, mix gently.
7. Add 200  $\mu\text{l}$  of EtOH (100/96%), mix gently and let it sit at room temp for 5 min.
8. Transfer lysate to the QIAmp minElute Column and proceed with washing according to the protocol described in the kit.
9. Elute with 30  $\mu\text{l}$  of nanopure (Milli-Q) autoclaved water (allow 5 min for elution).
10. Centrifuge full speed for 1 min.
11. Elute again with another 30  $\mu\text{l}$  of water.
12. Centrifuge at full speed (13,000 rpm (16,000 $\times g$ )) for 1 min.
13. Use NanoDrop to determine DNA concentration and purity (see Note 13).
14. To obtain a more precise measurement of DNA concentration in the samples use PicoGreen ds DNA Assay Kit (Invitrogen/Molecular Probes #P11496).
15. If concentration of DNA is less than 10ng/ $\mu\text{l}$  (PicoGreen values), concentrate the samples using SpeedVac, (see Note 14).

### **3.5. DNA Amplification and Determination of Efficacy and Quality of Samples (Gel Electrophoresis)**

From this point on, continue the protocol by processing in parallel the test DNA and the reference DNA (matched normal DNA or commercially available normal human genomic DNA). It is essential that both samples be treated exactly the same.

DNA is amplified with illustra Genomiphi V2 DNA amplification Kit (GE Healthcare #25-6600-31) (see Note 15).

1. Set the heat block at 95 °C and water bath at 30 °C.
2. Mix 1  $\mu\text{l}$  (60 ng) of template DNA (test DNA or reference DNA) and 9  $\mu\text{l}$  of sample buffer (see Note 16).
3. Heat samples at 95 °C for exactly 3 min (not longer!) to denature DNA then cool on ice for 3 min.
4. Prepare amplification mix (9  $\mu\text{l}$  of reaction buffer+ 1  $\mu\text{l}$  of enzyme for each reaction), add 10  $\mu\text{l}$  of the mix to each tube and incubate samples for 2 h at 30 °C.
5. Set the temperature of heat block to 65 °C, which will be needed for the next step.

6. After 2 h at 30 °C, inactivate the enzyme by heating the reaction mix at 65 °C for 10 min.
7. Cool samples on ice.
8. Verify if amplification was successful by running 1 µl of each sample on 0.8% agarose gel (see Note 17).

### **3.6. DNA Labeling Purification and Determination of Efficiency of Labeling and DNA Quality**

Samples are fluorescently labeled using Invitrogen Array CGH genomic labeling kit (#45-0048), fluorescent nucleotides from Amersham: CY3-dUTP (#CA95040146L) and CY5-dUTP (#CA95040-158L) and manufacturer protocol:

1. Use total amount of each amplified sample, adjust their volume to 21 µl and add 20 µl of 2.5× Random Primers Solution.
2. Incubate at 95 °C for 5 min and then immediately cool on ice for 5 min.
3. On ice, add 5 µl of 10× dUTP Nucleotide Mix, 3 µl of CY5 dUTP (test DNA) or 3 µl CY3 dUTP (reference DNA) and 1 µl of Exo-Klenow Fragment.
4. Mix gently and incubate at 37 °C for 2 h.
5. Add 5 µl of Stop Buffer to each tube and place on ice. The reaction can be stored at -20 °C overnight.
6. Purify sample with Microcon Centrifugal Filters Ultracel YM-30 (Millipore #42410);  
Add 430 µl of TE buffer to each tube, load on a column, and centrifuge in microcentrifuge for 10 min at 10,000 rpm (6000×g). Proceed with washing and recovery as described by manufacturer.
7. Adjust volume of each sample to 80.5 µl with TE buffer.
8. Check the efficiency of labeling and the yield of DNA using NanoDrop. Typical yield is 7–10 µg of DNA and samples contain 50–80 pmol of labeled nucleotides per µg of DNA. The ratio of absorbance 260 nm/280 nm should be >1.8.

### **3.7. Hybridization, Washing and Scanning of the Image**

Hybridization is set according to the protocol described in detail in the manual from Agilent (Agilent Oligonucleotide Array-Based CGH for Genomic DNA Analysis, available on web).

1. Prepare hybridization mix: 50 µl of Cot-1 DNA (1 mg/ml, Invitrogen #15279-101), 52 µl of 10× blocking agent and 260 µl of 2× Hi-RPM Buffer (Agilent Oligo aCGH Hybridization Kit #5188-5380), and labeled DNAs (test and reference).
2. Incubate at 95 °C for 3 min following immediately by incubation at 37 °C for 30 min.
3. Use entire hybridization mixture for hybridization assembly and hybridize at 65 °C for 40 h.

4. Washing is performed with Washing Buffers from Agilent and according to their procedure, except the second wash at 37 °C is done for 3 min. We do not use stabilization or drying solutions (see Note 18).
5. Immediately after washing, slides are covered with Agilent Ozone-Barrier Slide Covers (#G2505-60550) and scanned using the Agilent scanner with 5  $\mu\text{m}$  (244 K array) or 3  $\mu\text{m}$  (1  $\times$  1 M arrays) resolution.

### 3.8. Image Processing and Data Analysis

Ultradense aCGH images generated by the Agilent scanner require processing performed by two highly specialized software packages. First, Agilent's *Feature Extraction* software is used to review assay quality and then to process the array spot intensities to extract the raw data. To conduct subsequent array data analysis we use Agilent's *Genomic Workbench* (GW). Although platform-independent analysis algorithms exist (both commercial and open source) GW tends to provide higher quality analysis with Agilent data (9). Each step in image processing and subsequent data analysis is highly dependent on the previous step and clear understanding of the implications and ultimately the functional interpretation and hypothesis development conditioned on the results requires significant insight into the potential Type I and Type II errors at each step. The potential origin of those errors is complicated. Generally, experimentalists without experience in the analysis would be wise to develop collaborations with bioinformaticians or others with the background, experience, and insight to navigate the iterative analysis described below and produce the highest possible quality results with accurate and informed interpretation.

#### 3.8.1. Image Processing and Feature Extraction

The Agilent *Feature Extraction* (FE) software for aCGH image processing is highly complex and requires 97 parameters (Agilent Feature Extraction Software Reference Guide, pp. 119–137) (19). Default settings should be used unless an expert in microarray image processing is available. Feature extraction software is used to:

1. View images of arrays obtained with the data generated by the Agilent scanner for preliminary inspection (.tiff files).
2. Extract the data and perform preliminary processing (.text files).
3. Create a quality control (QC) report on arrays (.pdf files).

To extract and process the data, the FE software first creates and aligns a grid to the image, identifies image spots, recognizes and rejects outlier pixels and then flags outlier features. Finally, signal and background biases are computed, calculation of signal log ratios is performed, and errors are estimated. The log ratio of fluorescence is the final readout of each spot on the array.

Explicit steps to run FE are found in the user guide (20) and briefly described in Agilent's Feature Extraction software tutorial (21).

### 3.8.2. aCGH Quality

Quality Control (QC) metrics integrated into the Agilent FE software facilitate the visual inspection of the overall quality of Agilent arrays. The QC Report generated by the software is an important review of the likelihood of success of the assay and the quality of the image analysis (see FE User Guide) (20).

Basic QC metrics provided by FE include the following:

1. **DLRSpread** (derivative log ratio spread): A measure of the log ratio noise calculated as the standard deviation (spread) of the log ratio differences between consecutive probes.
2. **BGNoiseRed/BGNoiseGreen** (background noise): Calculated as the standard deviations of negative control probes after rejecting nonuniform outliers (outliers significantly different from background), saturated features and feature population outliers (outliers the same as background), it is a measure of background fluorescence for each channel (Red and Green). High background fluorescence can result from inefficient washing procedure, contamination of washing dishes with detergent, or use of colored gloves when manipulating arrays (see Subheading 3.7 and Note 1).
3. **SignalIntensityRed/SignalIntensityGreen** (signal intensity): For each Red and Green channel, this metric is calculated as the median background-subtracted signal after rejecting non-uniform outliers and saturated features. For high quality arrays this metric should be more than 150. Low values (<50) are indicative of a failed array and marginal values (between 50 and 150), are indicative of noisy assays. Low values can result from poor quality DNA or loss during labeling and washing (see Notes 13 and 17 and Subheadings 3.6 and 3.7).
4. **SignalToNoiseGreen/SignalToNoiseRed** (signal to noise): For each channel this metric is calculated as **SignalIntensity** divided by **BGNoise**. The higher this ratio the better the quality of the array. If this ratio is low, there is significant noise in the data. A ratio over 100 indicates good DNA quantity and no significant problems with hybridization, washing, or scanning (see Subheading 3.7).
5. **ReproducibilityRed/ReproducibilityGreen** (reproducibility): For each channel, this metric calculates the median percent coefficient of variation (% CV) of background-subtracted signal for probes replicated on the chip, called noncontrol probes, after outlier rejection. It is a measure of the reproducibility of features. A large value of this metric (>0.2) indicates a problem with the hybridization procedure such as slide leakage.

### 3.8.3. Data Analysis

The primary goal of aCGH in cancer research is to identify regions of altered copy number such as amplifications and deletions in the cancer genome relative to the reference DNA. The aCGH data should also permit calculation of precise copy number for aberrant regions relative to the reference DNA. Analysis of aCGH assay results consists of a three-step process: (1) Loading the data, (2) Preprocessing, and (3) Identification of CNAs. Further analysis of multiple arrays or grouped assays as well as the analysis of correlations with phenotypes or medical outcomes may follow the three-step analysis of each individual array, but is outside the scope of this chapter.

#### 1. Loading the data

Step by step importation of FE files to GW is found in the GW User guide p. 50 (22). In addition, the correct microarray design files containing probe annotation and details can be downloaded (see Workbench user guide, pp. 51–54) (22). It is critical that the correct design files are loaded with the FE data files to ensure accurate data analysis.

#### 2. Preprocessing

Most aCGH-based aberration detection algorithms assume that the log ratio intensity values are approximately zero when no “true” aberration exists and adjacent probe noise signals are independent. However, this may not always be the case and GW has two preprocessing options to correct for biases before aberration detection algorithms are applied: centralization and GC correction. The centralization algorithm “zeros” the total data by adding or subtracting the same constant to all log ratio measurements thus setting the average log ratio for all array elements to zero. However, centralization should be applied with caution when highly aberrant cancer genomes are analyzed to avoid elimination of real aberration before a CNA detection algorithm is applied. This normalization can be used prior analysis with Aberration Detection Method 1 (ADM-1) and 2 (ADM-2) as well as circular binary segmentation (CBS) aberration detection algorithms.

GC correction compensates for local GC content of the genome, which tends to cause wavy artifacts in aCGH signal intensities. Such artifacts tend to interfere with detection algorithms and therefore lead to inaccurate aberration calls. GC correction algorithms are designed to remove this correlated signal. The effects of the algorithm are often negligible, but in the case when the baseline visually shows these artifacts we recommend using GC Correction when you perform CGH + SNP analysis.

Another GW normalization algorithm is Fuzzy Zero that applies a “global error model” to all aberrant intervals. In many samples, the log ratio errors of successive probes are correlated over wide genomic intervals and aberration detection methods which are based on estimating independent errors from probe to probe, (e.g., ADM algorithms) will underestimate the error for long intervals. Consequently, for such algorithms, long aberrations with low average log ratios are often incorrectly considered significant. The global error model is a more realistic approach to avoid such erroneous aberration. We recommend using the Fuzzy Zero algorithm with GW’s ADM-1 and ADM-2 aberration detection algorithms.

### 3. Identification of CNAs

Aberration detection algorithms typically interrogate genomic stretches or “windows” of log ratio intensities and corresponding probe quality measures to accurately identify genomic regions of altered copy number of the test genome relative to a reference genome. Agilent GW includes five such algorithms: Z-score, Aberration Detection Methods (ADM-1, and ADM-2), Hidden Markov Model (HMM), and CBS.

The Z-score method is a relatively simple method for detecting aberrant regions that are statistically different from the reference genome and tends to work well when genomic intervals have a high density of probes each with relatively high log ratios within a window defined by the user. The Z-Score method is sensitive to window size and Z-score “threshold.” Changing the window size tends to influence the length of detected aberration regions and changing the Z-score threshold affects the number of probes that are rejected and thus influence the aberration calls.

The ADM algorithms avoid the fixed window size limitation and thus may be more suitable to those users with limited aberration detection analysis experience. Both ADM-1 and ADM-2 sample adjacent probe’s log ratio of intensity to produce a robust estimation of the range of each aberrant segment. ADM-1 searches for intervals in which a statistical score exceeds a user specified threshold. The score is proportional to the absolute average log ratio of the genomic interval and to the square root of the number of probes in the interval. Intervals that have only a small number of probes with high log ratio values or those intervals that include relatively large number of probes with only slightly elevated log ratios will generate high score and therefore will be “called” aberrant. The statistical score represents the deviation of the average log ratios



from the expected value of zero, in units of standard deviation. The ADM-2 algorithm follows the same principle but includes a “quality” estimate based on the probe log ratio errors to produce a more robust aberration “call” when the assay is of generally lower quality. Consequently, we recommend the ADM-2 method for noisy arrays.

Both ADM-1 and ADM-2 depend on a judicious selection of the statistical score threshold. Although a threshold of six is recommended in the instruction manual, we recommend the use of a more conservative threshold (8–10). The threshold’s effect can be verified by visual inspection. Usually, when the array plots suggest the rejection of “true” aberrations, the threshold is too high. In addition, the aberrant regions defined by ADM-2 can be filtered (Aberration filters in a Discovery menu of GW). We have considered as aberrant only those regions, which contained minimum of five probes per region and/or those with minimum absolute  $\log_2$  ratio for region  $>0.3$ . The validity of aberration detection with ADM-1 and -2 is based on the assumption that errors from probe to probe are independent. However, this assumption is often incorrect and ADM-like algorithms tend to underestimate the error for long intervals. Consequently, long aberrations with low average log ratios are often incorrectly considered significant. To solve this problem we recommend application of the Fuzzy Zero algorithm (see section on data preprocessing, above).

The HMM aberration detection algorithm is based on a stochastic process which estimates the likelihood of observing a probe with aberrant signal in the presence of uncertainty in the signal measurement. In this algorithm, probes are sequentially binned according to whether they are estimated to likely measure an amplification, a deletion, or a region in which no copy number change occurred. The state with maximum probability (deletion, amplification, or no change) is assigned to each probe. HMM’s probabilistic approach assigns confidence measures or  $p$ -value to the estimate of each probe state. However, this algorithm tends to require significant computational time.

The CBS algorithm is similar to ADM-like algorithm in that it samples adjacent probes to arrive at a robust estimation of the true range of the aberrant segment. However, unlike ADM, CBS iterates within any given initial segment and measures the difference of log ratio means between two adjacent candidate segments to find a breakpoint. End points of the segment are then joined to test the combined mean log ratio. Consequently, this method identifies those break points that yield the greatest discrimination between segmental means. The output of the CBS algorithm does not include any score, and therefore, other

methods, especially ADM-1 and ADM-2, are better suited for aberration classification. In addition, because of searching iteratively for the best breakpoints the CBS algorithm is not as fast as ADM and Z-score algorithms.

### **3.9. Data Interpretation**

#### *3.9.1. Impact of Experimental Procedures on Data Generation, Quality, Analysis, and Interpretation*

The general design of high throughput experiments using clinical samples does not allow multiple measurements. To gain confidence about the data analysis and interpretation careful review and testing of the experimental protocols, data processing and analysis as well as setting standards of reproducibility with cell lines should be performed before clinical samples are tested. Protocols described in this chapter should aid your workbench preparation, but we suggest that in house testing and review be done before clinical specimens are used.

Data analysis and interpretation of aCGH experiments is a complex multi step process. Quality control at every step cannot be over-emphasized, but even when done carefully, experimental complications related to clinical samples (such as sample heterogeneity, genome amplification, and choice of reference) may have consequences for data interpretation and thus should be accounted for accordingly.

1. DNA Heterogeneity: We apply two independent steps to avoid problems with DNA heterogeneity; screening of specimens by the pathologist for percentage of tumor cells and flow sorting of tumor subpopulations according to DNA ploidy. However, some diploid samples may be more difficult to interpret than others and the conclusions about precise copy number for such samples have to be interpreted with caution.
2. WGA: The consequences of WGA for ultradense aCGH are discussed in detail in our paper (8), where we concluded that the artifacts related to WGA are more pertinent to the densest  $1 \times 1$  M Agilent arrays. We recommend using 244 K arrays when WGA is applied.
3. Reference Selection: The best reference for studies of cancer genome would be matched normal DNA from the same individual. However, this may be difficult to obtain for every clinical sample. The problems, which may arise due to the use of commercial normal human genomic DNA as a reference, are discussed in the following section.

#### *3.9.2. Impact of Scientific Questions on the Analytical Approach and the Interpretation*

aCGH is a hypothesis-free, genome-wide approach to investigate structural genomic rearrangements. The instability of a cancer genome is a well-accepted phenomenon that produces CNAs varying greatly in size and location within genomes (23). Analytical tools such as the algorithms described in Subheading 3.8 are meant to objectively evaluate the data to identify regions of CNAs within a cancer genome relative to the chosen reference. The next challenge is to interpret this overwhelming amount of information and

translate it into scientifically useful information and clinically important knowledge.

One way to do this is to focus the attention on patterns, which can be assigned to a particular phenotype determining its molecular signature for diagnostic or prognostic purposes. However, there is no standard way of describing complex cancer aberrations. In the literature, the regions of copy number changes are often described using cytoband nomenclature taken from cytogenetics. Few terms such as “focal amplifications/deletions” or “amplicons” are commonly used, but they are not precisely defined (Fig. 2a). Focal amplification/deletion are usually considered small (there is no indication about their size), while amplicons are usually considered large (again, no indication about their size) and focused around one area (e.g., ERBB2 amplicon on chromosome 17). However, the amplified areas within large amplicons may not be adjacent to each other, even though aCGH viewing tools display them like this. Parts of chromosomes may be translocated and rearranged, creating fusions between sequences, which in a normal genome are never close to each other. The complex structure of copy number changes within amplicons suggests that many events likely contributed to their existence (Fig. 2b).

DNA amplifications have been shown to contain important druggable oncogenes, such as the genes encoding for the HER2 and EGF receptors (1, 2). Thus, another common approach is search for such genes. The potential complication here is that some large amplifications or deletions contain many genes and finding those, which are pertinent to the phenotype (so-called drivers, as opposed to passengers, which are altered, but not important for tumor development) is not a trivial task.

In our approach we focused our attention on micro-CNAs, which we defined as ranging from 100 kb (the limit of detection for 244 K Agilent platform) or 20 kb (the limit for 1 M platform) up to 1 Mb in genomic length (submicroscopic limit). In searching for genes (biomarkers and/or drug targets), such small aberrations likely affect only a few genes, (in the MCF7 genome, most of such aberrations contain only one gene) (8). However, many smaller CNAs map to regions of known CNVs in the human genome. CNVs are structural genomic variants, which are responsible for the diversity between healthy individuals, but they also play a role in determining predisposition to some diseases such as autism or HIV infection (24). Some of these variants may be inherited and benign, but others could potentially contribute to tumor development. CNVs in the size range of micro-CNAs are very common in the human genome (25). Thus, when addressing micro-CNAs in the cancer genome we need to deal with the complex issues related to natural variation between individuals. The best way to avoid these issues is to use normal matched DNA as a reference for hybridization.

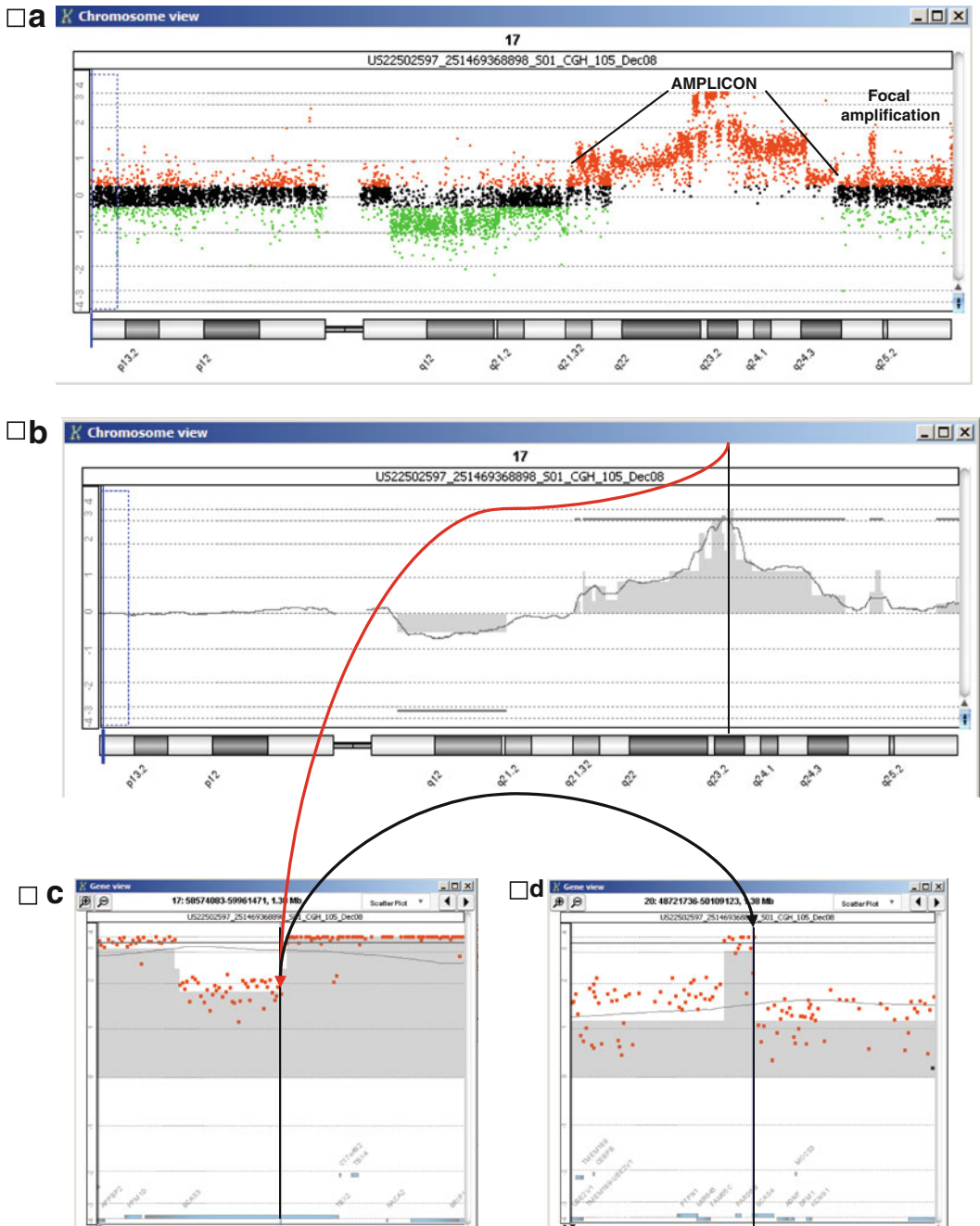


Fig. 2. Copy number alterations (CNAs) found on chromosome 17 in a genome of MCF-7 cells with ultradense aCGH platform (Agilent 244 K). (a) Chromosome view of a scatter plot showing data points for  $\log_2$  ratios (green: values below  $\log_2 = -0.3$ ; red: values above  $\log_2 = 0.3$ ; black: values above  $\log_2 = -0.3$  and below  $\log_2 = 0.3$ ) (b) Chromosome view of overlaid plots showing moving average (2 Mb window) for  $\log_2$  ratios of fluorescence between labeled MCF-7 DNA and the differentially labeled normal human reference, showing aberrations found with the ADM-2 algorithm (shaded areas) (c) Zoom-in on 1.38 Mb fragment of chromosome 17 showing close-up of the putative chromosomal breakage point which cut through the sequence of BCAS3 gene (genes are indicated by blue boxes) (d) Zoom-in 1.38 Mb fragment of chromosome 20 showing close-up of a putative chromosomal breakage point, which cut through the sequence of BCAS4 gene. The fusion of the BCAS3 gene on chromosome 17 with BCAS4 gene located on chromosome 20 was found by sequencing in the MCF 7 genome and validated with PCR (27).

Our second objective for ultradense aCGH application in studies of the cancer genome is to search for gene fusions. Recent evidence suggests that fusion events in epithelial cancers are more frequent than previously thought (26). Gene fusions are unique to cancers and make excellent candidates for biomarkers and/or drug targets (4). To identify gene fusions we shifted our focus in data analysis from genes mapped to the aberrant regions to those mapped to the breakpoints delineating CNAs and whose sequences are interrupted by amplifications/deletions. Example of how chromosomal breakage points may look like on aCGH data is shown in Fig. 2b–d.

When focused on finding gene fusions, we applied the following strategy:

1. All putative chromosomal breakage points, which are within the edges of segments of DNA copy number gains and losses as well as points of abrupt DNA copy number changes within called aberrations were found with the help of Agilent's ADM-2 algorithm and described according to the starting position of aberration (amplification or deletion) immediately adjacent to it.
2. A series of filtering strategies were applied to select the candidates for fusions: (1) putative breakage points, which map to known genes, and (2) putative breakage points associated with the difference in absolute average  $\log_2$  ratio  $>0.5$ . This can be followed by filtering strategies dictated by a biological approach: (1) putative breaks which recur within a studied cohort, or (2) putative break points within genes known to be involved in chromosomal translocations, or (3) putative break points within genes already known to be involved in gene fusions in other cancers, etc.

---

## 4. Notes

1. Do not use color nitrile gloves when handling arrays; the color can introduce background fluorescence.
2. The  $1 \times 1$  M array is used only when enough DNA (2.5  $\mu$ g) is available from biopsies and WGA is not necessary. Otherwise we use 244 K arrays (8).
3. The use of liquid nitrogen for snap freezing is the optimal method to preserve the molecular integrity of the sample. However, liquid nitrogen is often not accessible in the clinical setting making the adoption of this method challenging. To circumvent this limitation we have attempted to use RNeasy lysis buffer (Qiagen), a solution that allows the collection of tissue specimens at room temperature, preserving the integrity of the genetic material. Although, useful for many downstream

applications, RNAlater caused disintegration of nuclei and thus is not compatible with flow sorting prior aCGH.

4. It is possible to obtain enough nuclei from needle core biopsies with tumor content less than 70% as long as the core biopsy is at least 1mm × 5mm in size.
5. It is important to make sure that blood is collected from the patients before or after the biopsy procedure. Blood lymphocytes could be the source of matched normal DNA.
6. Flow sorting of nuclei is an excellent way to obtain pure tumor DNA from the dominant subpopulation in the specimen. However, this step can be skipped and DNA can be extracted directly from the frozen biopsy sample at the risk of having nonhomogenous DNA. The procedure for isolation of DNA in this case will be similar (Subheading 3.4) except that the digestion with Proteinase K should be carried overnight. However, flow sorting of nuclei from a biopsy may not only assure a more homogenous DNA sample, but it may also permit the isolation of different subpopulations from a single specimen, thus addressing the interesting issue of tumor evolution (11).
7. The minimum volume is 100 µl. More extraction buffer can be added if needed or if the biopsy is bigger, but keep in mind that increasing the volume of extraction buffer will dilute the sample. This will result in longer sorting time and a more dilute nuclear suspension.
8. To accurately determine the diploid reference position in a FACS profile, a diploid control sample (lymphoblastoid cell line of a normal person) is first run and then a small amount of nuclear suspension from the specimen is mixed into this control sample in order to establish FACS collecting gates.
9. The number of nuclei varies from biopsy to biopsy. We have always sorted the entire sample to get as many nuclei as possible. In some cases we obtain as many as a few millions of nuclei from one major fraction (Fig. 1). The amount of DNA obtained from this high number of nuclei would be sufficient to avoid WGA. However, the majority of specimens are not cellular and/or large enough to allow this. If WGA is not performed, DNA is digested with restriction enzymes exactly as described in the Agilent manual (Agilent Oligonucleotide Array-Based CGH for Genomic DNA Analysis). In this case the reference normal DNA will have to be digested as well.
10. The nuclear pellet is not visible in the Eppendorf tube. For better recovery of nuclei, samples with volume greater than 200 µl should be split and re-aliquoted in smaller volumes before centrifugation. After centrifugation, gently aspirate the supernatant using a pipette, do not use the vacuum. Leave a small volume of supernatant in the tube (about 80 µl).



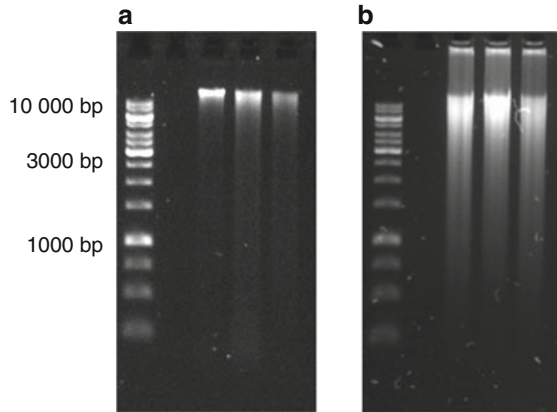


Fig. 3. Gel image of DNA samples. Samples were analyzed on an 0.8% agarose gel (a) DNA ladder and a clear single band of high molecular weight DNA, indicating a good quality sample (b) DNA Ladder and fragments of various molecular weights obtained after WGA.

11. Do not vortex, mix the solutions by gently tapping the tube with your finger.
12. We have never tested shorter times of digestion for this protocol. However, it is likely that digestion may be completed in less than an hour. This is true for samples containing nuclei. If the whole biopsy is used without sorting, samples should be digested overnight.
13. An excellent way to verify if the sample contains high molecular weight DNA is agarose gel electrophoresis. We could easily visualize quantities of about 30–50 ng of genomic DNA. A single band of high molecular weight and no smear indicates high quality (Fig. 3a). However, when the number of nuclei is small, the amount of DNA obtained might be too small to run on a gel. Our yield was about 100–200 ng of DNA from 20,000 nuclei and the following step requires a minimum of 60 ng. Thus, we most often relied on the NanoDrop and PicoGreen measurements for DNA quantitation and quality control at this step.
14. Bring down the volume to about 20  $\mu$ l. Frequently check the sample during drying to make sure that it will not dry completely. Complete drying and reconstitution may cause degradation of high molecular weight DNA.
15. If you encounter problems during this or the following steps with one of the samples, the test DNA or the reference DNA, it is preferable that you restart again from the amplification step with both samples in parallel. Do not adjust volumes or put aside any aliquots.
16. Although smaller amounts of DNA can be successfully amplified (we have amplified as little as 5 ng of DNA), we do not recommend decreasing this amount; the fewer the templates, the



greater the risk of amplifying nonspecific sequences. If your sample has a concentration below 60 ng/μl, you will need to concentrate it; pipette the volume corresponding to 60 ng into the tube, which will be used for the amplification reaction and place it in a SpeedVac for 1–2 min. Make sure that your sample will never dry. Add 9 μl of sample buffer directly to this tube and continue the protocol.

While pipetting very small volumes of DNA for amplification, make sure that you see the liquid in the tip after aspiration and then deposit it to the bottom of the tube, making sure that the tip touches the bottom. The tiny drop needs to be drawn to the tube by adhesion, since micro liter drops will not fall by themselves. Note that even without any DNA in your tube you may still end up with amplified DNA, which will be visible on an agarose gel. However, this will not be your sample, but a contaminant. Unfortunately, you may realize this only after you see the final result of aCGH. Note: Cleanliness is extremely important when doing WGA.

17. Gel electrophoresis. Prepare 0.8% agarose gel (Ultrapure agarose) in TBE buffer, add Syber safe Nucleic Acid Staining solution and pour into Bio-Rad minigel apparatus. To prepare the samples use 1 μl of amplified DNA, 9 μl of water and 2 μl of loading buffer. Run a gel at stable voltage (70 V) for 2 h. A typical image of amplified DNA is shown in Figs. 3b.
18. To obtain 37 °C temperature of washing buffer #2 in an open staining dish, we proceeded as follows: a 1.5 l flat Pyrex baking dish filled with water was placed on the top of heating stir plate. After the water in the baking dish reached 50 °C, the staining dish was placed in it. The temperature in the dish reached 37 °C 15–20 min later. At that moment the washing with buffer #1 (5 min room temp.) was started and immediately followed with wash #2 at 37 °C (3 min).

## References

1. Ross JS, Fletcher JA, Bloom KJ et al (2004) Targeted therapy in breast cancer: the HER-2/neu gene and protein. *Mol Cell Proteomics* 3:379–398
2. Shigematsu H, Gazdar AF (2006) Somatic mutations of epidermal growth factor receptor signaling pathway in lung cancers. *Int J Cancer* 118:257–262
3. Mano H (2008) Non-solid oncogenes in solid tumors: EML4-ALK fusion genes in lung cancer. *Cancer Sci* 99:2349–2355
4. Neal JW, Sequist LV (2010) Exciting new targets in lung cancer therapy: ALK, IGF-1R, HDAC, and Hh. *Curr Treat Options Oncol* 11:36–44
5. Meyerson M, Gabriel S, Getz G (2010) Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet* 11:685–696
6. Pinkel D, Albertson DG (2005) Array comparative genomic hybridization and its applications in cancer. *Nat Genet* 37(Suppl):S11–S17
7. Curtis C, Lynch AG, Dunning MJ et al (2009) The pitfalls of platform comparison: DNA copy number array technologies assessed. *BMC Genomics* 10:588

8. Przybytkowski E, Ferrario C, Basik M (2011) The use of ultra-dense array CGH analysis for the discovery of micro-copy number alterations and gene fusions in the cancer genome. *BMC Med Genomics* 4:16
9. Pinto D, Darvishi K, Shi X et al (2011) Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nat Biotechnol* 29:512–520
10. Klopffleisch R, Weiss AT, Gruber AD (2011) Excavation of a buried treasure—DNA, mRNA, miRNA and protein analysis in formalin fixed, paraffin embedded tissues. *Histol Histopathol* 26:797–810
11. Navin N, Kendall J, Troge J et al (2011) Tumour evolution inferred by single-cell sequencing. *Nature* 472:90–94
12. Callagy G, Jackson L, Caldas C (2005) Comparative genomic hybridization using DNA from laser capture microdissected tissue. *Methods Mol Biol* 293:39–55
13. Ruiz C, Lenkiewicz E, Evers L et al (2011) Advancing a clinically relevant perspective of the clonal nature of cancer. *Proc Natl Acad Sci USA* 108:12054–12059
14. Navin N, Krasnitz A, Rodgers L et al (2010) Inferring tumor progression from genomic heterogeneity. *Genome Res* 20:68–80
15. Hughes S, Lasken R (2005) Whole genome amplification. Scion Publishing Ltd, Bloxham
16. Hughes S, Lim G, Beheshti B et al (2004) Use of whole genome amplification and comparative genomic hybridisation to detect chromosomal copy number alterations in cell line material and tumour tissue. *Cytogenet Genome Res* 105:18–24
17. Pugh TJ, Delaney AD, Farnoud N et al (2008) Impact of whole genome amplification on analysis of copy number variants. *Nucleic Acids Res* 36:e80
18. Agilent Technology, Inc. Agilent Oligonucleotide Array-Based CGH for Genomic DNA Analysis Enzymatic Labeling for Blood, Cells, or Tissues (with a High Throughput option) Protocol (February 2010) #G4410-90010\_CGH\_Enzymatic\_Protocol\_v6.2.1.pdf
19. Agilent Technology, Inc. Agilent Feature Extraction Software (v10.7) Reference Guide (October 2009) #G4460-90026\_FE\_Reference.pdf
20. Agilent Technology, Inc. Agilent Feature Extraction Software (v10.5) User Guide (December 2008) #G4460-90019\_FE\_10.5\_User.pdf
21. Agilent Technology, Inc. [http://www.genomics.agilent.com/GenericA.aspx?pagetype=Custom&subpagetype=Custom&pageid=2151,resource on the web](http://www.genomics.agilent.com/GenericA.aspx?pagetype=Custom&subpagetype=Custom&pageid=2151,resource%20on%20the%20web).
22. Agilent Technology, Inc. Agilent Genomic Workbench 6.5 CGH Interactive Analysis User Guide (September 2010) #g3800-90028\_cgh\_interactive.pdf
23. Hicks J, Krasnitz A, Lakshmi B et al (2006) Novel patterns of genome rearrangement and their association with survival in breast cancer. *Genome Res* 16:1465–1479
24. Stankiewicz P, Lupski JR (2010) Structural variation in the human genome and its role in disease. *Annu Rev Med* 61:437–455
25. Speleman F, Kumps C, Buysse K et al (2008) Copy number alterations and copy number variation in cancer: close encounters of the bad kind. *Cytogenet Genome Res* 123:176–182
26. Prensner JR, Chinnaiyan AM (2009) Oncogenic gene fusions in epithelial carcinomas. *Curr Opin Genet Dev* 19:82–91
27. Hampton OA, Den Hollander P, Miller CA et al (2009) A sequence-level map of chromosomal breakpoints in the MCF-7 breast cancer cell line yields insights into the evolution of a cancer genome. *Genome Res* 19:167–177



<http://www.springer.com/978-1-62703-280-3>

Array Comparative Genomic Hybridization  
Protocols and Applications

Banerjee, D.; Shah, S.P. (Eds.)

2013, XII, 382 p., Hardcover

ISBN: 978-1-62703-280-3

A product of Humana Press