

# A scale-space method for detecting recurrent DNA copy number changes with analytical false discovery rate control

Ewald van Dyk<sup>1,2,3,4</sup>, Marcel J.T. Reinders<sup>3,4</sup> and Lodewyk F.A. Wessels<sup>1,2,3,4,\*</sup>

<sup>1</sup>Bioinformatics and Statistics group, Division of Molecular Carcinogenesis, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands, <sup>2</sup>Cancer Systems Biology Center, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands, <sup>3</sup>Delft Bioinformatics Laboratory, Faculty of EEMCS, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands and <sup>4</sup>Netherlands Bioinformatics Centre, Geert Grooteplein 28, 6525 GA Nijmegen, The Netherlands

Received September 3, 2012; Revised January 30, 2013; Accepted February 17, 2013

## ABSTRACT

Tumor formation is partially driven by DNA copy number changes, which are typically measured using array comparative genomic hybridization, SNP arrays and DNA sequencing platforms. Many techniques are available for detecting recurring aberrations across multiple tumor samples, including CMAR, STAC, GISTIC and KC-SMART. GISTIC is widely used and detects both broad and focal (potentially overlapping) recurring events. However, GISTIC performs false discovery rate control on probes instead of events. Here we propose Analytical Multi-scale Identification of Recurrent Events, a multi-scale Gaussian smoothing approach, for the detection of both broad and focal (potentially overlapping) recurring copy number alterations. Importantly, false discovery rate control is performed analytically (no need for permutations) on events rather than probes. The method does not require segmentation or calling on the input dataset and therefore reduces the potential loss of information due to discretization. An important characteristic of the approach is that the error rate is controlled across all scales and that the algorithm outputs a single profile of significant events selected from the appropriate scales. We perform extensive simulations and showcase its utility on a glioblastoma SNP array dataset. Importantly, ADMIRE detects focal events that are missed by GISTIC, including two events involving known glioma tumor-suppressor genes: CDKN2C and NF1.

## INTRODUCTION

DNA copy number alterations in cancer, typically recorded by array comparative genomic hybridization (aCGH), single nucleotide polymorphism (SNP) arrays and (more recently) sequencing, can reveal interesting genes that are important for diagnosis, prognosis and targeted therapeutics. However, genomic instability typically introduces random or passenger alterations that make it hard to distinguish recurring alterations (possibly harboring driver genes) from the rest in single sample (tumor) measurements.

A number of statistical methods have been developed to detect aberrations that recur at high frequencies across multiple samples. These methods include CMAR (1), Significance Testing for Aberrant Copy numbers (STAC) (2), Hierarchical Hidden Markov model (H-HMM) (3), Genomic Identification of Significant Targets in Cancer (GISTIC) (4), GISTIC2.0 (5), JISTIC (6) and Kernel Convolution: a Statistical Method for Aberrant Region deTectioN (KC-SMART) (7).

CMAR and STAC require discretized copy number alteration profiles where genomic regions take on one of three discrete states: a loss, no-aberration or a gain. Although this is partially justified because copy number changes in DNA are discrete in nature, measurements are typically performed on DNA extracted from a heterogeneous pool of cell populations, which could cause deviations from the expected discrete values. Therefore, CMAR and STAC disregard valuable information by ignoring the amplitude of gains or losses in single samples. H-HMM does not require discretized profiles but uses three hidden states to model losses, absence of aberrations and gains.

GISTIC2.0 requires non-discretized, but segmented, profiles. Segmentation (typically performed on single sample profiles) reduces measurement noise, but approximates a signal that varies across the genome with a

\*To whom correspondence should be addressed. Tel: +31 205127987; Fax: +31 206691383; Email: l.wessels@nki.nl

piecewise constant signal, requiring selection of segment boundaries (breakpoints). Breakpoints can be missed (in noisy profiles), and therefore, segmentation also introduces a form of discretization.

All methods used to detect recurring aberrations, in one way or another, aggregate (sum) all the sample profiles either in raw, segmented or discretized form. This results in a significant reduction in biological noise (passenger events) with respect to signal (recurring events). In addition, aggregation also reduces measurement noise, justifying an approach followed by, e.g. KC-SMART, that avoids segmentation all together and performs smoothing on the aggregated profile.

In particular, GISTIC2.0 and KC-SMART use a statistical framework that weighs both the amplitude and frequency of recurrence in its detection procedure. JISTIC is an adaptation of GISTIC, and all arguments used for GISTIC2.0 in this article also apply to GISTIC and JISTIC.

Possibly the single most desirable property of GISTIC2.0 is its ability to detect focal recurring events embedded in broader events (such as whole chromosome arms being deleted) through a peel-off algorithm requiring knowledge of segment boundaries provided by a segmentation algorithm. However, to the best of our knowledge, there are no approaches that analytically (without resorting to permutation tests) characterize the significance of recurring events and, at the same time, use a principled approach for automatic scale selection (required level of smoothing) while guaranteeing a specified error rate (average number of falsely detected recurrent events).

For an extensive review on (many more) methods, see (8).

Here we present ADMIRE (Analytical Multi-scale Identification of Recurring Events), a smoothing methodology, with the following features:

- Segmentation and/or calling are not required for the genomic profiles. Instead, reduction of measurement noise is achieved by performing smoothing on the aggregated profile;
- Automatic scale selection, or selection of the level of smoothing, is performed on the aggregated profile to increase the power for detecting recurrent events. For example, broad recurrent events are detected with higher significance if we allow for a higher level of smoothing. An important characteristic of the approach is that the error rate is controlled across all scales and that the algorithm outputs a single profile of significant events selected from the appropriate scales;
- A recursive procedure to detect statistically significant focal recurrent events that are embedded in broader events;
- An analytical method that controls the expected number of detected false-positive recurrent events (and therefore helps avoid time-consuming permutation tests)

## METHODS

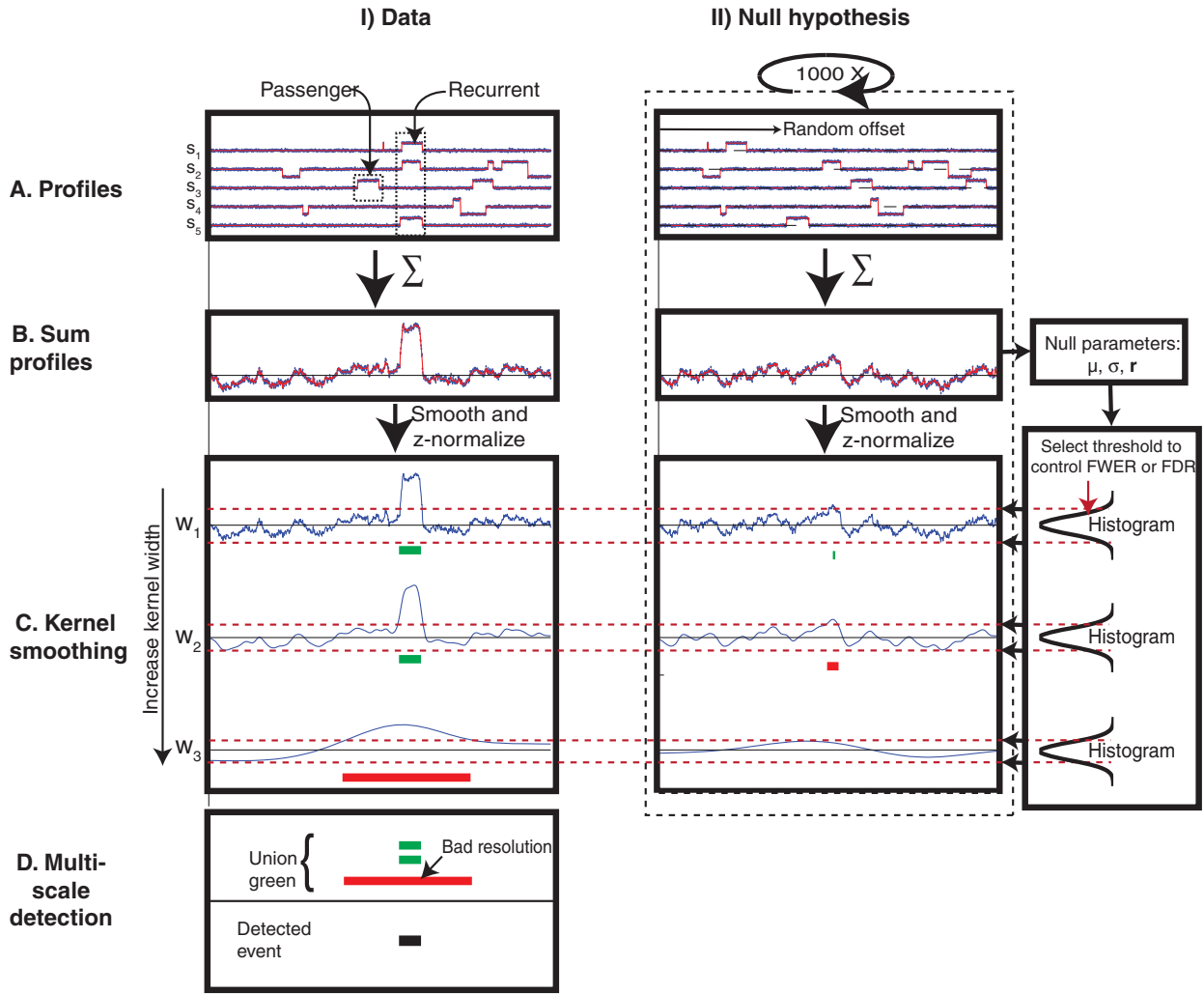
The ADMIRE methodology is summarized in Figure 1 and described in subsequent subsections. In this

example, and subsequent simulations, we simulate aCGH profiles, but any technique, such as SNP arrays (see RESULTS) or sequencing, might be used in principle. In Figure 1, the left column (Column I) illustrates the methodology on measured profiles, whereas Column II illustrates the construction of the null distribution (the expected behavior of the aggregated profile if none of the copy number alterations are recurrent). Multiple aCGH samples are summed [Figure 1B.I (Figure 1, Row B, Column I)] to obtain a single aggregated profile in which recurrent aberrations reveal high peaks compared with passenger events. This indicates that in our model, we consider both the frequency and amplitude of events, similar to the approach followed by GISTIC2.0 and KC-SMART. Next we perform kernel smoothing at different scales (Figure 1C.I) to reduce measurement noise. Figure 1A.II illustrates how we can simulate profiles that share no recurrent events by performing cyclic permutations on each profile individually, Figure 1B.II shows the summation of the resulting profiles to obtain a representative null hypothesis that closely resembles a stationary Gaussian random process and Figure 1C.II shows the kernel convolution per scale. In Figure 1 (Column II), these steps (permutation, summation and smoothing) are repeated 1000 times to obtain an empirical approximation of the null distribution per scale. These distributions are used to derive a threshold per scale corresponding to the desired false discovery rate (FDR) or family-wise error rate (FWER) of passenger events. The permutation test is shown for illustration purposes. ADMIRE avoids permutations altogether by exploiting an analytical relationship between the desired threshold and FDR or FWER. We apply the constant thresholds derived at each scale (kernel width) to obtain recurrent segments for each scale separately (Figure 1D.I). In Figure 1D.I and II, we regard only detected recurrent segments that are of sufficient resolution (the detected event is large compared with the kernel width) and take the union of all significant segments across all scales. The final step (not shown in Figure 1) involves a recursive procedure to detect focal recurrent events embedded in broad events. In the following sections, we will run through all these steps in more detail.

### Aggregation

Consider an ordered set of small genomic sequences  $\langle \text{seq}_i \rangle$  ( $\langle \rangle$  means the set is ordered) that are centered at genomic positions  $\langle p_i \rangle$  on a normal reference genome. Each such sequence has an average copy number  $\langle t_{s,i} \rangle$  across all cells in a specified tumor sample  $s$ . Furthermore, for a normal cell, we have a reference copy number for each sequence  $\langle n_i \rangle$  (typically  $n_i = 2$  for a diploid sequence). From now on we assume that we have an unbiased probe measurement of the log ratio (the base of the log is irrelevant for the subsequent analysis)  $a_{s,i} = \log(t_{s,i}/n_i)$ , where a positive (negative) value indicates a gain (loss) in the tumor sample.

To find recurring losses or gains, we simply add all sample profiles into one aggregated profile (Figure 1B.I).



**Figure 1.** Illustrating the steps involved for detecting recurring aberration in multiple copy number alteration profiles with the multi-scale ADMIRE approach. All plots in the left column, Column I, represent data with recurrent events, and Column II shows the exact same procedure when permuting the data to construct a cyclic shift null hypothesis. Column I: (A) Illustration of five (of 100) simulated aCGH profiles with recurring events and a number of passenger (random) aberrations. (B) The first step in detecting recurring events is to sum all profiles (100 samples) to a single aggregated profile. (C) A Gaussian kernel is convolved with the aggregated profile and z-normalized, as described in the text. This is done with many different kernel widths so that focal events can be detected with small kernels and broad events with larger kernels. Ultimately, constant thresholds (derived from the empirical null as outlined in Column II) will be applied on the smoothed signal (both upper and lower tail), as illustrated by the red dashed lines. (D) Illustration of how we combine all the events found on multiple scales. Basically, we take the union of all events found on all scales; however, for all kernels (except the smallest), we perform a filtering procedure to ensure the proper resolution. The procedure is simple in that we only keep those events that are substantially (20 times) larger than the kernel width (more on this in the text). Column II: Illustration of the permutation of profiles where each profile’s probes are cyclically shifted with a random offset (Panel A) and the summation of the resulting profiles (Panel B) to obtain a representative null hypothesis that closely resembles a stationary Gaussian random process with parameters  $\mu$ ,  $\sigma$  and the auto-correlation  $r$ . Panel C shows the kernel convolution per scale. In this illustration, we propose to repeat the steps in Panels A, B and C one thousand times to obtain an empirical approximation of the null distribution and use these distributions to derive a threshold per scale corresponding to the desired control of FDR and FWER. However, in this article, we derive an analytical relationship between the thresholds and FWER or FDR.

The aggregated probe values are given by:

$$a_i = \sum_{s=0}^{S-1} a_{s,i}, \tag{1}$$

where  $s \in \{0, 1, \dots, S - 1\}$  and  $i \in \{0, 1, \dots, P - 1\}$  are the sample and probe indices, respectively.

This process is the same as that proposed by KC-SMART and GISTIC2.0, with the fundamental exception that we do not split gains and losses. Little

power is lost by doing this, except for clear cases where a region (of the same size) is recurrently lost and gained. The major advantage of not splitting gains and losses is that relevant statistics (such as FDR control) become analytically tractable.

**The null hypotheses**

We propose to model the null distribution by performing random cyclic permutations. This implies that for genomic profile  $s$ , we push all probes by a random

number  $U_s$  positions to the right. The  $U_s$  probes that are pushed out of the genomic profile on the right are cycled around and fill the  $U_s$  empty positions that are created on the left of the profile. This process is performed for each sample independently (Figure 1A.II). We prefer this over random permutation of the probes in a sample profile because it destroys the recurrence structure but retains the auto-correlation between probes. After every sample has undergone a random cyclic shift, all the profiles are aggregated (Figure 1B.II). More specifically,

$$\begin{aligned} A_i^0 &= \sum_{s=0}^{S-1} A_{s,i}^0 \\ A_{s,i}^0 &= a_{s,R_{s,i}} \\ R_{s,i} &= i + U_s(\text{mod } P), \end{aligned} \quad (2)$$

where  $U_s$  is a uniform random variable covering  $\{0, 1, \dots, P-1\}$ . Note that each individual probe is identically distributed and identical to the distribution obtained from a permuting null hypothesis, as we randomly select one of the log ratios in each sample. It is also clear that the cyclic auto-correlation remains unchanged for each sample. Furthermore,  $A^0 = \{A_0^0, A_1^0, \dots, A_{P-1}^0\}$  is a homogeneous random process since the correlation between probes is independent of the probe labels and depends only on their relative ordering on the genome.

We can easily obtain analytical expressions for the mean, variance and auto-correlation of  $A^0$ :

$$\begin{aligned} \mu &= \sum_{s=0}^{S-1} \mu_s, & \mu_s &= \frac{1}{P} \sum_{i=0}^{P-1} a_{s,i} \\ \sigma^2 &= \sum_{s=0}^{S-1} \sigma_s^2, & \sigma_s^2 &= \frac{1}{P} \sum_{i=0}^{P-1} (a_{s,i} - \mu_s)^2 \\ r(\Delta i) &= \sum_{s=0}^{S-1} r_s(\Delta i), & r_s(\Delta i) &= \frac{1}{\sigma^2 P} \sum_{i=0}^{P-1} [(a_{s,i} - \mu_s) \times \\ & & & (a_{s,i+\Delta i(\text{mod } P)} - \mu_s)] \end{aligned} \quad (3)$$

Alternatively, we can represent the auto-correlation function with a  $P \times P$  diagonal-constant correlation matrix  $\mathbf{r}$  with  $r_{i,j} = r(i-j)$ .

Because we are summing multiple profiles, the random process will become multivariate Gaussian (a consequence of the central limit theorem), and the parameters in Equation 3 fully describe the random process.

Technically, it is more desirable to calculate a homogeneous auto-correlation measure based on genomic distance instead of probe index, as probes are not equally spaced. Nonetheless, the proposed scheme provides a good approximation.

### Smoothing with a fixed kernel width

As we do not assume that the input samples are segmented, and therefore contain substantial measurement noise, it is desirable to smooth the aggregated profile

(Figure 1C). We describe an optimal kernel smoothing methodology based on the assumption that the null hypothesis is a random Gaussian process. The idea is that if we fix the kernel type (e.g. Gaussian) and the kernel width (i.e. the number of nearby probes to average, in our case controlled by the standard deviation of the Gaussian kernel), we can normalize the smoothed (continuous) profile so that each point on the genome has exactly the same normal distribution (mean zero and variance one) in the null process. This way we can apply a constant threshold across the whole genome when detecting recurring aberrations.

The first step is to smooth the signal by convolving the aggregated profile with a kernel.

$$F_w^0(g) = k_w(g) * \sum_{i=0}^{P-1} A_i^0 \delta(g - p_i), \quad (4)$$

where  $g$  is the position on the genome,  $F_w^0(g)$  is the smoothed random process,  $k_w(g)$  is the kernel of width  $w$  ( $\exp(-0.5g^2/w^2)$  for a Gaussian kernel) and  $p_i$  is the genomic location of probe  $i$ .  $*$  represents the convolution operator and  $\delta$  is the Dirac delta function.

The smoothed function  $F_w^0(g)$  is a linear combination of  $\{A_0^0, A_1^0, \dots, A_{P-1}^0\}$  with coefficients  $\{k_w(g - p_0), k_w(g - p_1), \dots, k_w(g - p_{P-1})\}$  for any given point in space  $g$ .

We can calculate the exact mean and variance of  $F_w^0(g)$  as follows:

$$\begin{aligned} \mu_w(g) &= \mu m_w(g) \\ \sigma_w^2(g) &= \sigma^2 s_w^2(g), \end{aligned} \quad (5)$$

where

$$\begin{aligned} m_w(g) &= k_w(g) * \sum_{i=0}^{P-1} \delta(g - p_i) \\ s_w^2(g) &= \bar{k}_w^T(g) \mathbf{r} \bar{k}_w(g) \end{aligned} \quad (6)$$

$\bar{k}_w^T(g)$  is a  $1 \times P$  column vector equal to the kernel coefficients  $[k_w(g - p_0), k_w(g - p_1), \dots, k_w(g - p_{P-1})]^T$  and  $\mathbf{r}$  is the auto-correlation matrix.

We choose a threshold function such that  $F_w^0(g)$  has the same (single tale)  $P$ -value  $\rho$  for any given  $g$ . Therefore:

$$P(F_w^0(g) \geq t(g)) = \rho \quad (7)$$

As  $F_w^0(g)$  is Gaussian we get:

$$t(g) = \mu_w(g) + \sigma_w(g) t_\rho, \quad (8)$$

where  $t_\rho$  is a constant threshold that controls  $F_w^0(g)$  at a  $P$ -value  $\rho$ .

Equivalently, we can z-normalize  $F_w^0(g)$  to apply a constant threshold  $t_\rho$  represented by the z-normalized smoothed random process  $H_w^0(g)$ :

$$H_w^0(g) = \frac{F_w^0(g) - \mu m_w(g)}{\sigma s_w(g)} \quad (9)$$

It is worth mentioning that  $H_w^0$  is a differentiable (smooth) normal random process (with mean zero and variance one for all  $g$ ), but is non-homogeneous (unlike the discrete random process  $\langle A_i^0 \rangle$ ) due to unequal probe spacings.

**Counting significant events**

We ultimately seek to provide a list of genomic regions (broad or focal events) that are significantly recurring and therefore likely to be relevant in cancer development. In providing such a list, we are interested in controlling the expected proportion of regions that are in error (passenger events). We call this the event-based FDR. Before we can do this, it is important to first define what we mean by an event.

For a fixed threshold and kernel width, we define positive and negative excursion sets as follows:

$$\begin{aligned} a^+(h_w, t) &= \{g \in a | h_w(g) \geq t\} \\ a^-(h_w, t) &= \{g \in a | h_w(g) \leq -t\}, \end{aligned} \tag{10}$$

where  $h_w$  is the smoothed (and z-normalized) aggregate profile (see Equation 9) and  $a$  is the set of all  $g$  considered (the genome).  $a^+$  and  $a^-$  represent all genomic regions that are deemed recurrently gained and lost, respectively (relative to the threshold  $t$ ). Due to the Gaussian null hypothesis, we will focus all attention on  $a^+$  and realize that symmetric arguments exist for  $a^-$ .

We define positive recurring events to be the maximally connected subsets (ordered by inclusion) of  $a^+$ . For a smoothed aggregated profile  $h_w(g)$  and fixed threshold  $t$ , we represent the total number of detected events with  $\chi(h_w, t)$ . Note that counting the number of events is equivalent to counting the number of up-crossing on the threshold and adding one if the left boundary point is above the specified threshold.

**Analytical relationship between the threshold and the expected number of events found in the null hypothesis**

For  $H_w^0$  (smoothed non-homogenous Gaussian process), we can find an exact analytical expression that relates any given  $t$  to the expected number of events found ( $E[\chi(H_w^0, t)]$ ), with the only restriction being that the kernel selected must be differentiable up to the second order. A large amount of work has been done on finding  $E[\chi(H_w^0, t)]$  for homogeneous fields (9–12) and little on non-homogeneous fields (13). Therefore, we extend the theory for non-homogeneous (one-dimensional) processes in the supplementary Data (see the section entitled ‘Analytical expression for the Euler characteristic’) and show the final result here. More specifically, for a non-homogeneous process, the expected number of events is given by:

$$E[\chi(H_w^0, t)] = \frac{1}{2} \operatorname{erfc}\left(\frac{t}{\sqrt{2}}\right) + \frac{e^{-t^2/2}}{2\pi} \int_{g \in a} \sqrt{v_w(g)} dg, \tag{11}$$

where

$$\begin{aligned} v_w(g) &= \operatorname{Var}\left[\frac{d}{dg} H_w^0(g)\right] \\ &= \frac{\frac{d}{dg} \bar{k}_w^T(g) \mathbf{r} \frac{d}{dg} \bar{k}_w(g)}{\bar{k}_w^T(g) \mathbf{r} \bar{k}_w(g)} - \left(\frac{\bar{k}_w^T(g) \mathbf{r} \frac{d}{dg} \bar{k}_w(g)}{\bar{k}_w^T(g) \mathbf{r} \bar{k}_w(g)}\right)^2 \end{aligned} \tag{12}$$

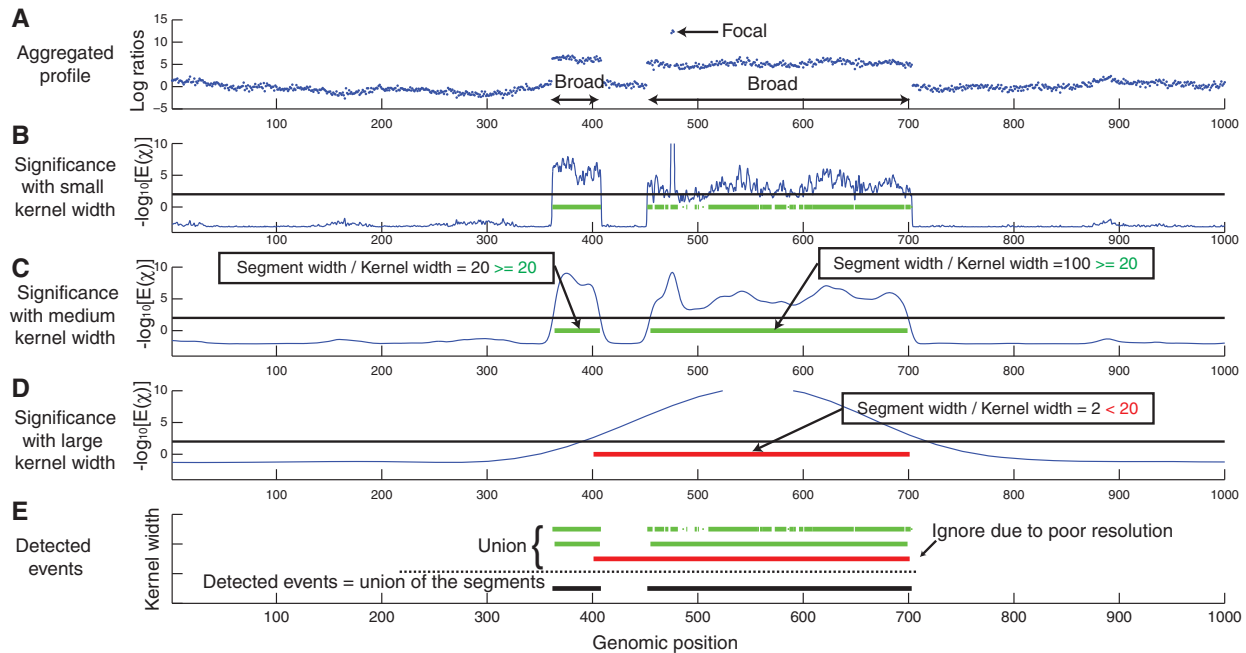
$v_w(g)$  is a function that represents the roughness of the random process (naturally the variance in the derivative) and depends entirely on the probe locations, smoothing and auto-correlation  $\mathbf{r}$  (and is independent of parameters  $\mu$  and  $\sigma$  because we z-normalized). For a rough random process (when we perform little smoothing), the integral in Equation 11 will be large and reflects the severity of multiple testing.

Note that we do not concern ourselves with estimating the full distribution of  $\chi(H_w^0, t)$ , but only the mean.  $E[\chi(H_w^0, t)]$  is a sufficient statistic for calculating the FDR (explained later).  $E[\chi(H_w^0, t)]$  is also an upper-bound for the FWER and becomes tight for practical FWERs ( $< 0.1$ ) (14).

We specifically used Gaussian kernels in this work, but Equation 11 hold for all kernels that are twice differentiable. For the application of detecting recurrent events, it is desirable to use a symmetric kernel that drops to zero, such as Gaussian, Student  $t$ , Cauchy or wavelet kernels. As the kernel is implemented in a discrete setting, it is also important to ensure that the kernel has a limited frequency bandwidth so that the smoothed aggregated profile can be sampled at a reasonable (Nyquist) rate.

**Multi-scale detection**

Previous sections indicate how we can control  $E[\chi(H_w^0, t)]$  for a fixed kernel width. GISTIC2.0 performs no smoothing on the aggregated profile (or effectively smooth with a small kernel width) and relies on noise reduction through segmentation on single profiles. Figure 2 shows that for unsegmented profiles, we can gain power by considering many kernel widths in parallel. For example if we try to detect broad recurring events, we gain power when increasing the kernel width. On the other hand, large kernel widths will reduce the resolution of profiles and estimated recurrent region boundaries will be inaccurate and focal events lost. This is illustrated in Figure 2. In Panel B, the resolution is high, resulting in accurate boundaries but low power causing the broad event to be shattered in many small events. In Panel D, the power is high, but the boundaries are inaccurate. Panel C shows a good compromise between boundary precision and power. Therefore, it is desirable to restrict the size of allowed kernels based on the size of detected events. To be more specific, at any given scale (except the smallest kernel width considered, as the resolution is assumed to be high), all detected events that have a detected width smaller than  $\alpha = 20$  times the kernel width will be ignored because they result in a poor resolution. Rather, these events are detected at a smaller scale to ensure a proper accuracy of the event boundaries. In fact, for  $\alpha = 20$ , at least 70% of any detected event will overlap with a real recurrent event (see the supplementary section entitled



**Figure 2.** Illustration showing how power can be gained by considering multiple scales (levels of smoothing). (A) A simulated aggregated profile with two broad recurring gains and one focal gain embedded in a broad event. (B) Significance level of the aggregated profile for little smoothing (small kernel width). Owing to the small kernel width, the resolution is high and the boundaries on the detected regions are fairly accurate. This is at the expense of power and results in hundreds of significant segments instead of two broad events. (C) Significant power is gained for intermediate kernel widths and the two broad events are found as desired. Furthermore, the resolution is high enough (the segment size is much greater than the kernel width) and therefore the boundaries of the significant events are sufficiently accurate (compared with the aberration size). (D) High power is observed for large kernel widths (significance level exceeds the threshold by far) but the resolution is so low that two events are merged into one and boundary estimates are poor. (E) We obtain the final estimate of recurring segments by taking the union of all detected events on all scales that reveal sufficient resolution. Note that the focal events embedded in broad events are completely missed. Furthermore, significance in these figures is represented by the expected number of events  $E[\chi]$  found across the whole genome (as predicted by the null hypothesis). The threshold is selected at  $E[\chi] = 0.01$ , a close upper-bound for the FWER of 0.01.

‘Details on multi-scale detection’).  $\alpha$  can be set by the user, and in the supplementary section entitled ‘Resolution parameter  $\alpha$  on simulated data’, we illustrate how different settings of  $\alpha$  influence results (see Supplementary Figure S1).

Finally, the union of all the remaining significant regions across all scales represents the recurring events in the data. This multi-scale procedure will more likely merge events that appear on the smallest scale than create new ones on a larger scale. This enables us to keep control over the number of detected events (see the supplementary section entitled ‘Details on multi-scale detection’ and Figure 5).

#### Updating the null parameters based on known recurrent events

Parameters  $\mu$ ,  $\sigma$  and  $\mathbf{r}$  will, in general, be conservative estimates for the non-recurrent null hypothesis if estimated on all probes, especially if a large proportion of these probes are recurrent. Therefore, it is desirable to ignore all probes that are known to be recurrent when estimating the null parameters. This is done iteratively by first calculating conservative parameter estimates (with all probes considered) and then removing all the probes that are deemed recurrent through the multi-scale detection procedure. If we re-calculate the null parameters (which will be less conservative) with the remaining probes only, more recurring events will be found. This process is

repeated until no more new recurring events can be found (see supplementary section ‘Details on updating the null parameters’ commenting on the convergence behavior). Although this method will drastically increase power, the null parameters will either be slightly optimistic or remain conservative if some recurrent events remain undetected.

#### Recursive multi-level detection of recurring aberrations

The events detected by the procedure as described thus far include focal and broad events, but we are not yet able to detect focal events that are embedded in broad events. To find those, we propose a recursive scheme that finds new events that are embedded in earlier detected events. For example, let us say that we find (among other) one broad recurrent gain that starts (ends) at genomic location  $g_s$  ( $g_e$ ). We re-estimate the null parameters  $\mu$ ,  $\sigma$  and  $\mathbf{r}$  from all probes between  $g_s$  and  $g_e$  and perform the multi-scale analysis to find smaller events embedded within this broad event. This procedure for finding a focal event within a broad event is illustrated in Figure 3. Again we iteratively update the null parameters until the null region converges (a new null region inside the broad event). Note that the boundaries of the detected broad event ( $g_s$  and  $g_e$ ) might be inaccurate and therefore embedded focal events might be detected at the border of the initial broad event. As these are a result of the boundary inaccuracy, we simply ignore them (unless, e.g. it is a focal gain within a gain). We repeat this recursive procedure until no more

events can be found and represent the results in recursive levels.

On a final note, not only does the recursive multi-level detection procedure allow us to detect recurring events embedded in broad recurring events, but also helps to improve our estimate on  $E[\chi]$ , as explained in the supplementary Data (see the section entitled ‘Details on recursive multi-level detection’).

**FDR control**

As we are able to predict the expected number of events found in the null hypothesis, we can also control the event-based FDR, the expected proportion of detected events that are false discoveries.

To see this, consider the Benjamini–Hochberg procedure (15) that controls the FDR at level  $q$  for  $m$  independent or positive dependent tests: Let  $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(m)}$  be the ordered observed  $P$ -values and  $m_0$  the number of true null hypotheses. If we reject the null hypotheses for tests  $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(k)}$ , where

$$k = \max\{i : m_0 p_{(i)} \leq iq\}, \tag{13}$$

then  $FDR \leq q$ . If we reject all tests with a  $P$ -value lower than  $\rho$ , then the expected number of false-positive tests  $E_\rho(\#FP) = m_0\rho$  (irrespective of the correlation that might exist between tests). Therefore, Equation 13 can be rewritten:

$$k = \max\{i : E_{p_{(i)}}(\#FP) \leq iq\} \tag{14}$$

For our application, Equation 14 is intuitive. For the  $i$ th detected event, if the ratio between the expected number of false-positive events ( $E(\chi)$ ) and the number of events  $i$  detected is smaller than the FDR ( $q$ ), then the FDR will be in control. We can lower the detection threshold until the inequality in Equation 14 is violated.

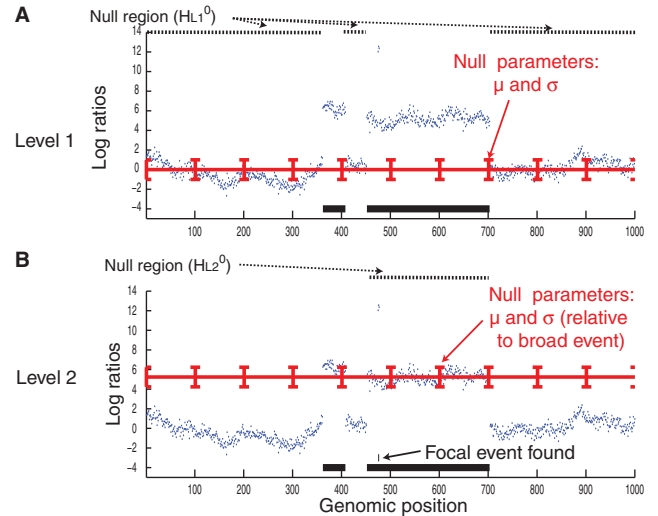
We propose the following procedure to find an appropriate value for  $E[\chi]$  to control the FDR at level  $q$ :

- Set  $n = 1$ ,  $E_1[\chi] = q$  and  $i_1 = 0$ ;
- REPEAT:
  - Detect recurrent events using ADMIRE with thresholds corresponding to  $E_n[\chi]$ . Count the number of detected events  $i_{n+1}$ ;
  - IF  $i_{n+1} \leq i_n$ : BREAK;
  - Set  $E_{n+1}[\chi] = (i_{n+1} + 1)q$ ;
  - Set  $n = n + 1$

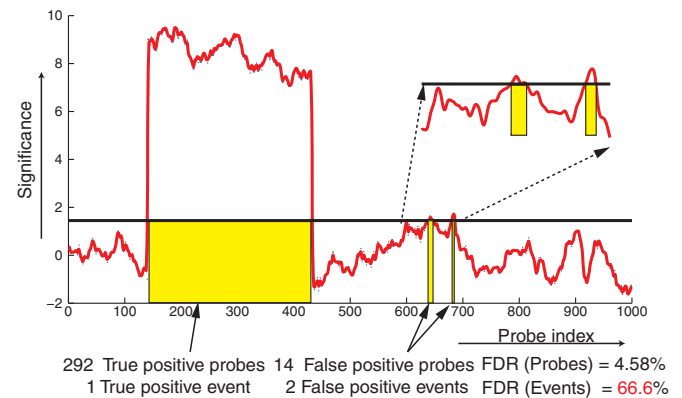
This methodology is different from that performed in GISTIC2.0. GISTIC2.0 regards each probe as an independent test (owing to the random permutation scheme) and uses the methodology proposed by Benjamini and Hochberg (16) to control the probe-based FDR (i.e. the proportion of false-positive probes). In contrast, ADMIRE performs event-based FDR, and this subtle, yet profound, difference is illustrated in Figure 4.

**RESULTS**

This section starts with an artificial, simulated dataset to illustrate several properties of ADMIRE. We start off by



**Figure 3.** Illustrating the recursive multi-level detection methodology. (A) On recursive level 1, we detect recurrent aberrations with the proposed multi-scale methodology. Note that the region in which we finally estimate the null parameters ( $\mu$ ,  $\sigma$  and  $\mathbf{r}$ ) is restricted to  $H_{L1}^0$ , as illustrated by the dotted line at the top of the figure. (B) On recursive level 2, we follow the exact same procedure, except this time, estimate the null parameters in the broad event  $H_{L2}^0$ . This allows us to detect embedded focal events inside broader events.



**Figure 4.** Probe-based versus event-based FDR control. Illustration on how controlling the probe-based FDR (expected proportion of detected probes that are false-positives) can introduce an unexpected proportion of focal events simply due to the presence of broad chromosomal aberrations.

demonstrating that the theoretical estimate of the expected number of events,  $E[\chi]$ , is indeed a good approximation of the empirically observed number of events under a wide range of experimental conditions. Then we move on to show that  $E[\chi]$  is a close upper-bound of the FWER and that the ADMIRE algorithm does control event-based FDR at the desired level. Finally, we demonstrate the properties of ADMIRE on a real-world glioma dataset.

**Datasets**

*Simulated datasets*

We simulate aCGH profiles on a genome consisting of  $2.4 \times 10^8$  bps and randomly select 12 000 probe positions for measurements. For each profile, we select 159 random breakpoints (160 segments) on the genome, of which a

random selection of 50% of the segments take on log ratios of 0 (all probes in these segments). The remaining segments randomly take on log ratios of  $-1$  and  $+1$ , representing passenger gains and deletions, respectively. We also add random Gaussian (measurement) noise to each profile (with variance  $\sigma_n^2$ ) for a specified signal to noise ratio (SNR) defined as  $\text{SNR} = 1/\sigma_n^2$ . For example, an SNR of  $10^{99}$  implies negligible measurement noise.

When recurrent events are added, we typically specify a width, location and frequency of recurrence across samples. For simplicity, probes covered by recurrent events take on values of  $-1$  or  $+1$  to represent recurrent gains or losses, respectively. For example, we might decide to add a recurrent event centered at  $1.2 \times 10^8$  bps,  $1 \times 10^6$  bps wide with a 30% frequency of occurrence across all samples.

In total, there are three global parameters that will be varied across experiments: (i) the number of samples to aggregate ( $s$ ), (ii) the SNR and (iii) the number of recurrent aberrations. For every recurrent aberration, we also specify the width, genomic location and frequency of occurrence.

For a detailed description on how we typically generate such a dataset, see the supplementary section entitled ‘KC-SMART vs. ADMIRE smoothing methodologies’ (under ‘Simulated data’).

### The glioma dataset

To demonstrate the properties of ADMIRE on real data, we used the dataset described by Beroukhim *et al.* (4) consisting of 141 high-quality glioma samples (107 primary Glioblastoma multiforme (GBM), 15 secondary GBMs and 19 lower-grade gliomas) to aggregate. DNA was hybridized on a Affymetrix  $\approx 100,000$  SNP array platform. Batch effects and systematic errors were removed using the exact methodology described by Beroukhim *et al.* (4) (see their Supporting information). All samples were segmented using Gain and Loss Analysis of DNA (GLAD) (17) to reduce measurement noise (this was done for both GISTIC2.0 and ADMIRE), and all known copy number variation probes were removed from the analysis.

### $E[\chi]$ simulations

We simulate aCGH profiles using the methodology proposed earlier; however, we do not add any recurrent aberrations.

To investigate whether our theoretical model of the expected number of detected events ( $E[\chi]$ ) is accurate for different thresholds, noise levels and kernel widths, we performed the following experiments. We varied the number of samples to aggregate,  $S$ , such that  $S \in \{2, 5, 10, 50, 100\}$ ; the SNR assumed two values,  $\text{SNR} \in \{1, 1 \times 10^{99}\}$  and the Gaussian kernel width was set to  $w \in \{4 \times 10^4, 8 \times 10^4, 1.6 \times 10^5, 3.2 \times 10^5, 6.4 \times 10^5, 1.3 \times 10^6\}$ . For combinations of these variables, we simulated 1000 artificial datasets.

In Figure 5A, we show the relationship between the analytical and empirical  $E[\chi]$  as the detection threshold is varied for a fixed kernel width of  $4 \times 10^4$  bps (two probes per kernel width, on average) and an SNR

of 1 ( $\sigma_n^2 = 1$  per sample). We show this result for all values of  $s$ .

Figure 5B is similar to Figure 5A, except that we do not add measurement noise. This serves to illustrate that our approach can also be applied to segmented data.

The main conclusion drawn from Figure 5A and B is that the analytically predicted  $E[\chi]$  becomes more accurate as we increase the number of aggregated samples due to the central limit theorem. For smaller sample sizes, the theoretical estimate is conservative.

In Figure 5C, we fix  $s$  to 100 and the SNR to  $1 \times 10^{99}$  and vary the kernel widths to show that the analytical estimate of  $E[\chi]$  remains accurate for all kernel widths. We also show that the empirical  $E[\chi]$  is smaller than the analytical  $E[\chi]$  if we perform the multi-scale detection.

Next we investigated the relationship between the empirical and theoretical estimate of  $E[\chi]$  on the glioma dataset. To obtain an empirical estimate of  $E[\chi]$ , we constructed a null hypothesis by repeating the cyclic permutation procedure, aggregation and kernel smoothing as outlined in Figure 1II., one thousand times on the glioma dataset. The results for  $S = 141$  and all kernel widths including the multi-scale analysis are depicted in Figure 5D. Overall, the theoretical prediction serves as a relatively tight upper-bound for the empirical estimate, but depends on the kernel width. More specifically, the estimate of  $E[\chi]$  becomes more accurate for larger kernels owing to adjacent probes being averaged (and again the central limit theorem suggests better convergence).

Overall, this experiment shows that the analytical  $E[\chi]$  is sufficiently accurate and that the multi-scale procedure produces conservative results.

### FWER simulations

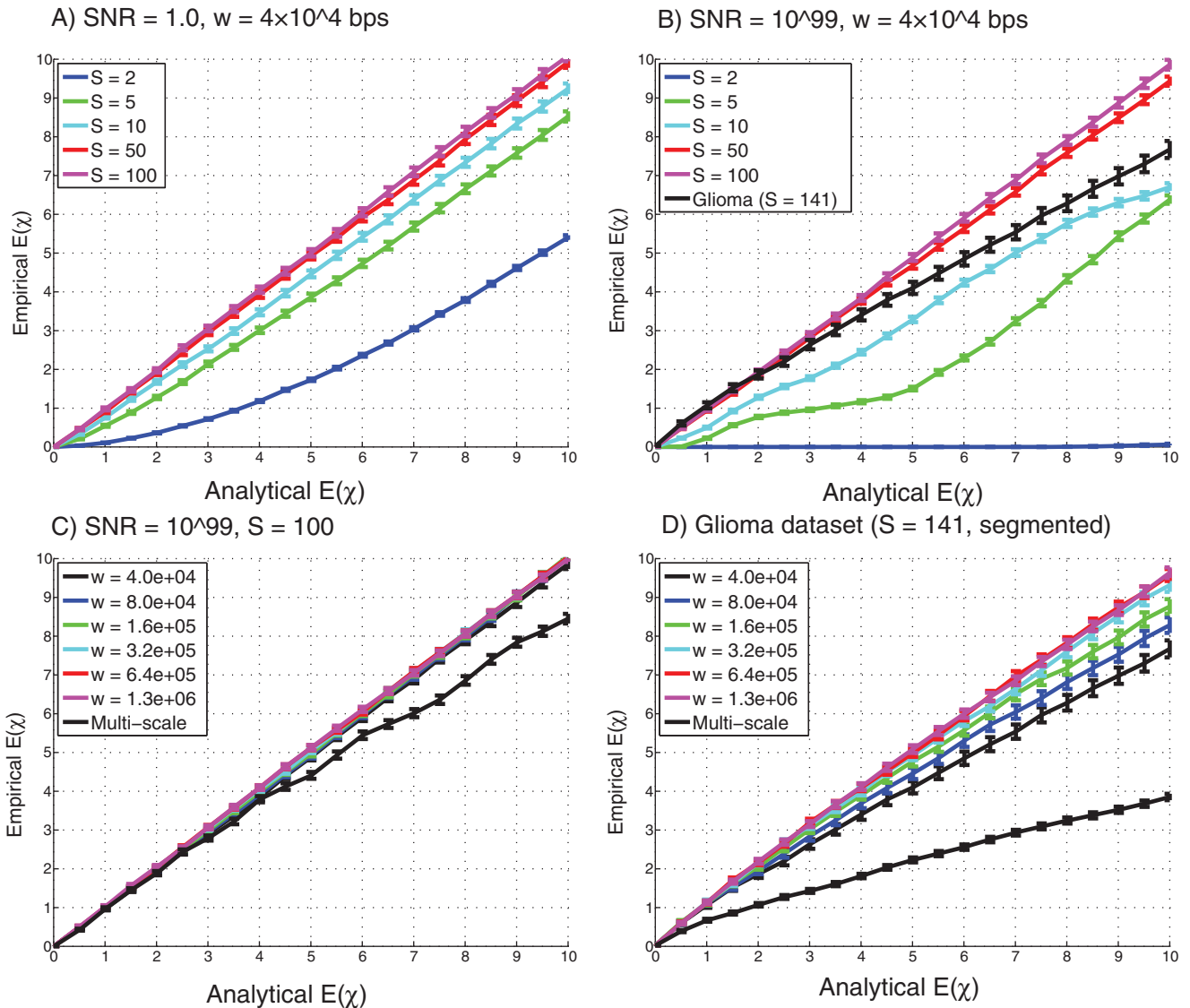
We observed earlier that  $E[\chi]$  is a close upper-bound for the FWER (14), and in this section, we perform simulations to verify this fact. We simulated aCGH profiles using the same methodology proposed earlier.

We fix the number of samples to aggregate to 100 and only add one recurrent event centered at 120 Mbps with a given width,  $w_a$ , and a 30% chance of occurrence per sample.

In every simulation, we also fix the kernel width and therefore do not perform a multi-scale analysis. Neither do we search for embedded events through recursion. However, we do update the null parameters iteratively based on known recurrences. See the supplementary section entitled ‘FWER control for simulated data’ for a detailed description of the experiment.

Figure 6.A depicts a typical power plot as a function of aberration size and kernel width—for an elaborate collection of these plots for different SNRs, see Supplementary Figure S3. This plot shows how the power changes (for the analytical FWER fixed at 5%) for detecting recurring aberrations of different sizes (one event per simulation) while varying the kernel width. We can observe that for a fixed kernel width, the power decreases as the aberration size decreases. In fact, there is an abrupt drop in power when the aberration size equals the kernel width, as indicated by





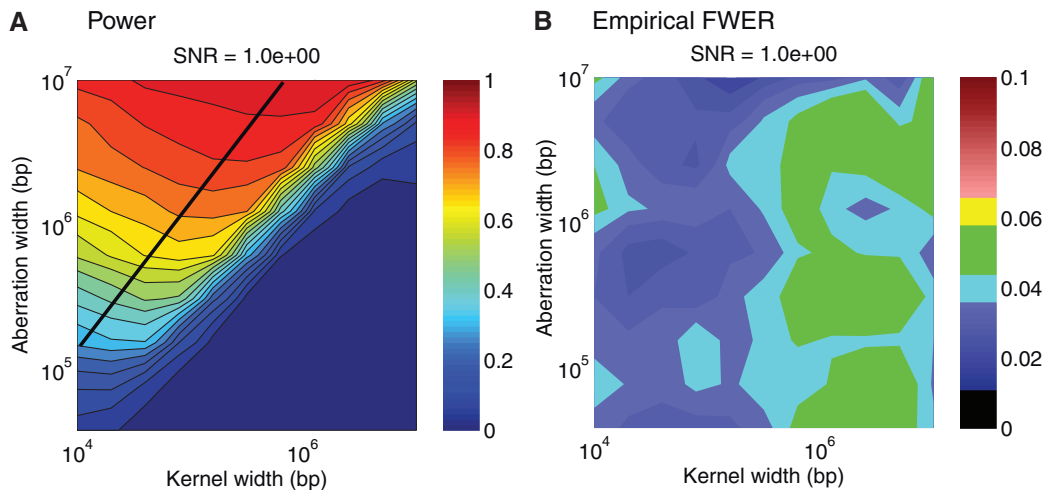
**Figure 5.** Illustration of the relationship between the analytical estimates of  $E[\chi]$  (x-axis) and that measured across 1000 simulations (y-axis) of aCGH profiles containing only passenger events. (A) We fix the kernel width to be small (40 kb) and the SNR at 1 to represent measurement noise. We vary the number of samples to aggregate for each simulation experiment. (B) A similar experiment on simulated aCGH profiles where we added no measurement noise ( $\text{SNR} = 1 \times 10^{99}$ ) and therefore effectively work with segmented samples. The black line depicts the result obtained when using cyclic permutation to create a null hypothesis on the glioma dataset. (C) The number of simulated samples to aggregate is fixed at 100 and the kernel width is varied, showing good theoretical predictions for all kernels. The black line indicates the mean number of events detected when we apply multi-scale selection. (D) Similar results are depicted when using cyclic permutations to create a null hypothesis on the glioma dataset. The genome size for the simulated data is only  $12 \times 10^8$  bps, whereas the glioma dataset consists of all probes stretching from chromosome 1 to 22. Error bars indicate the standard error of the empirical  $E[\chi]$ .

the diagonal ridge in the panel. In general, we can conclude that as long as the aberration is larger than the kernel width (region above the diagonal line), we have more power to detect the aberration. Figure 6B shows that the measured FWER (the chance of detecting one or more false-positives) is close to that predicted by  $E[\chi]$ , as expected. From these simulations it is clear that for any recurrent aberration of a fixed width, a fixed kernel width can be selected to gain optimal power. If the kernel width becomes too large, we observe a drastic loss in power, as indicated by the lower right corner in Figure 6A. Note that in contrast, Figure 2 suggests that larger kernels increase the power, but if we extend

Figure 2 to show even larger kernels, the significance levels will drop drastically.

**FDR simulations**

For the FDR experiments, we expanded the simulated dataset described previously to include recurrent events of different sizes and to have overlapping recurrent events. This will allow the possibility to estimate the capacity of the *complete* ADMIRE algorithm to control the FDR. More specifically, we expanded the simulated dataset by adding  $N_b = 2$  broad ( $20 \times 10^6$  bps, 1000 probes, on average) and  $N_m = 5$  medium-size ( $2 \times 10^6$



**Figure 6.** (A) A representative plot of the power for detecting a recurring aberration as a function of the aberration size and kernel width for the SNR fixed at 1. In this experiment, we added only a single recurring aberration per experiment and fixed  $E[\chi]$  at 5%. The black line indicates the maximum allowed kernel width at which an aberration can be detected if we apply filtering with  $\alpha = 20$  in the multi-scale methodology. See Supplementary Figure S3 for similar plots at different SNRs. (B) The empirical FWER. The green regions indicate that the measured FWER is within 1 standard deviation of the expected 5% FWER.

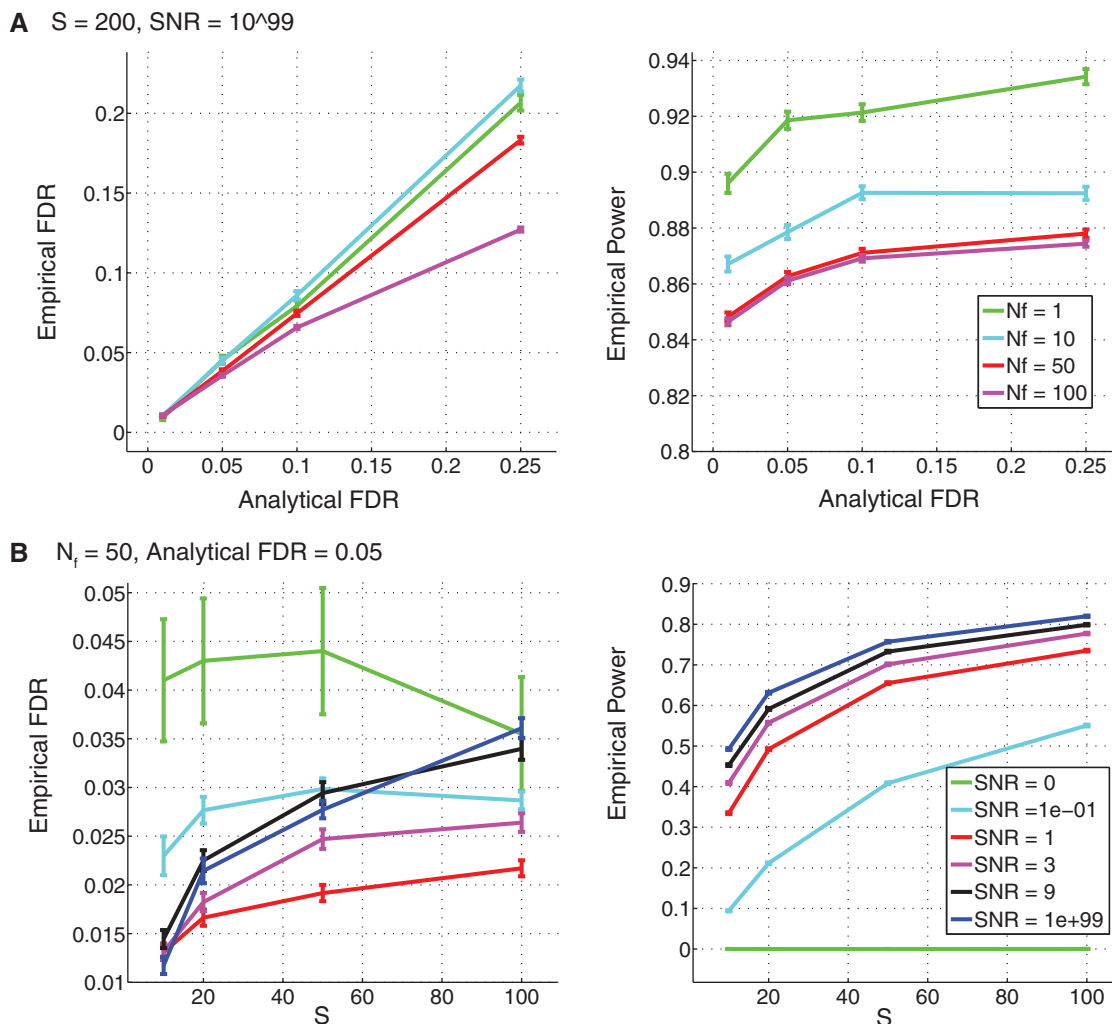
bps, 100 probes, on average) non-overlapping recurrent events at random locations (albeit consistent between samples) on the genome. Furthermore, we added a varying number ( $N_f$ ) of recurrent focal events (100 kb, five probes, on average) across the genome (potentially overlapping with the broad- and medium-size events). For each recurrent event, we select a random frequency (between 0 and 1) of occurrence across samples.

The complete ADMIRE algorithm has been applied with a specified analytical FDR. The number of samples to aggregate ( $s$ ) is varied, as well as the SNR and the number of focal recurring events ( $N_f$ ). An event is considered a true-positive if at least 70% of the detected region overlaps with a true recurrent region (the multi-scale detection procedure with  $\alpha = 20$  filtering guarantees an overlap of at least 70%). The number of true-positive events is then the sum of the number of true recurrent broad (maximum two), medium (maximum five) and focal events found. The empirical FDR is calculated by averaging the proportion of falsely detected events across 1000 simulation experiments. Likewise, the empirical power is the average proportion of true recurring events that are detected. For example, when we add only one recurring focal event, we hope to detect eight true events (two broad, five medium and one focal). If, for example, we detect four of the eight recurrent events and one extra false event in one simulation, the measured FDR would be 20% and the power 50%.

In Figure 7A, we fix the number of samples to aggregate to 200 and the SNR is set high (zero measurement noise and profiles are segmented). We vary the number of focal events and the analytical FDR and represent the measured FDR and power. In Figure 7B, we fix the number of focal recurring events to 50 and the analytical FDR to 5%, while varying the number of samples that are aggregated and the SNR.

From Figures 7A and B, it is clear that the empirical FDR is smaller than that predicted analytically. The three main reasons for this are the following:

- Inaccurate estimation of the null random process parameters  $\mu$ ,  $\sigma$  and  $\mathbf{r}$ . The higher the number of true-positives missed, the more conservative the null parameter estimates are and the true FDR will be smaller than predicted. Ultimately, this estimate will be most conservative if we estimate null parameters across the whole genome. In Figure 7A, we can clearly see that the FDR decreases when we increase the number of recurrent events. This is because for a fixed threshold, the expected number of undetected events is proportional to the total number of events (this is a simple consequence of how we generated the data). Therefore, for a larger number of recurrent events, the expected number of events that go undetected will be large and therefore the null parameters will be more conservative. This is especially noticeable in Figure 7B, where the FDR is fixed at 5% and we vary the SNR. For an SNR of zero, the null parameters will be accurate (as recurrent events do not exist) and we expect the FDR estimate to be close to the predicted value, whereas for an SNR of one, the null parameters will include a significant proportion of the recurrent signal. This situation improves again for higher SNRs owing to an increase in power;
- The multi-scale procedure in Figure 2 also ensures a conservative estimate on  $E[\chi]$ , as illustrated in Figure 5C (and D);
- If the number of samples to aggregate ( $s$ ) is small, the Gaussian model becomes inaccurate for the null hypothesis. This explains the reduced FDR for small values of  $s$  values in Figure 7B. Note that for SNR = 0, the Gaussian model is accurate no matter how many profiles we aggregate.



**Figure 7.** The relationship between the theoretically predicted analytical FDR and empirical FDR and power for a simulated dataset. (A) The empirical FDR (left panel) and power (right panel) as a function of the analytical FDR (varied between 1 and 25%) for the number of true focal recurrent events assuming the following values,  $N_f \in \{1, 10, 50, 100\}$ , while keeping the number of samples to aggregate per simulation fixed at 200, i.e.  $S = 200$ . Furthermore, we do not add any noise, as the  $\text{SNR} = 10^{99}$ , implying that all samples are segmented. (B) The empirical FDR (left panel) and power (right panel) as a function of the number of samples to aggregate  $S$  for the SNR assuming the following values,  $\text{SNR} \in \{0, 0.1, 1, 3, 9, 10^{99}\}$ , while keeping the number of focal recurrent events and FDR fixed at 50 ( $N_f = 50$ ) and 5%, respectively.

The power curves in Figure 7A (right panel) counterintuitively suggest that we lose power when increasing the number of focal events. However, if we consider that medium-size ( $N_m = 5$ ) and broad-size ( $N_b = 2$ ) events are detected with much higher power, it becomes obvious. For example, if we add only one focal event, then 7/8 of all recurrent events are of medium or broad size, whereas for 100 focal events, this ratio is only 7/107.

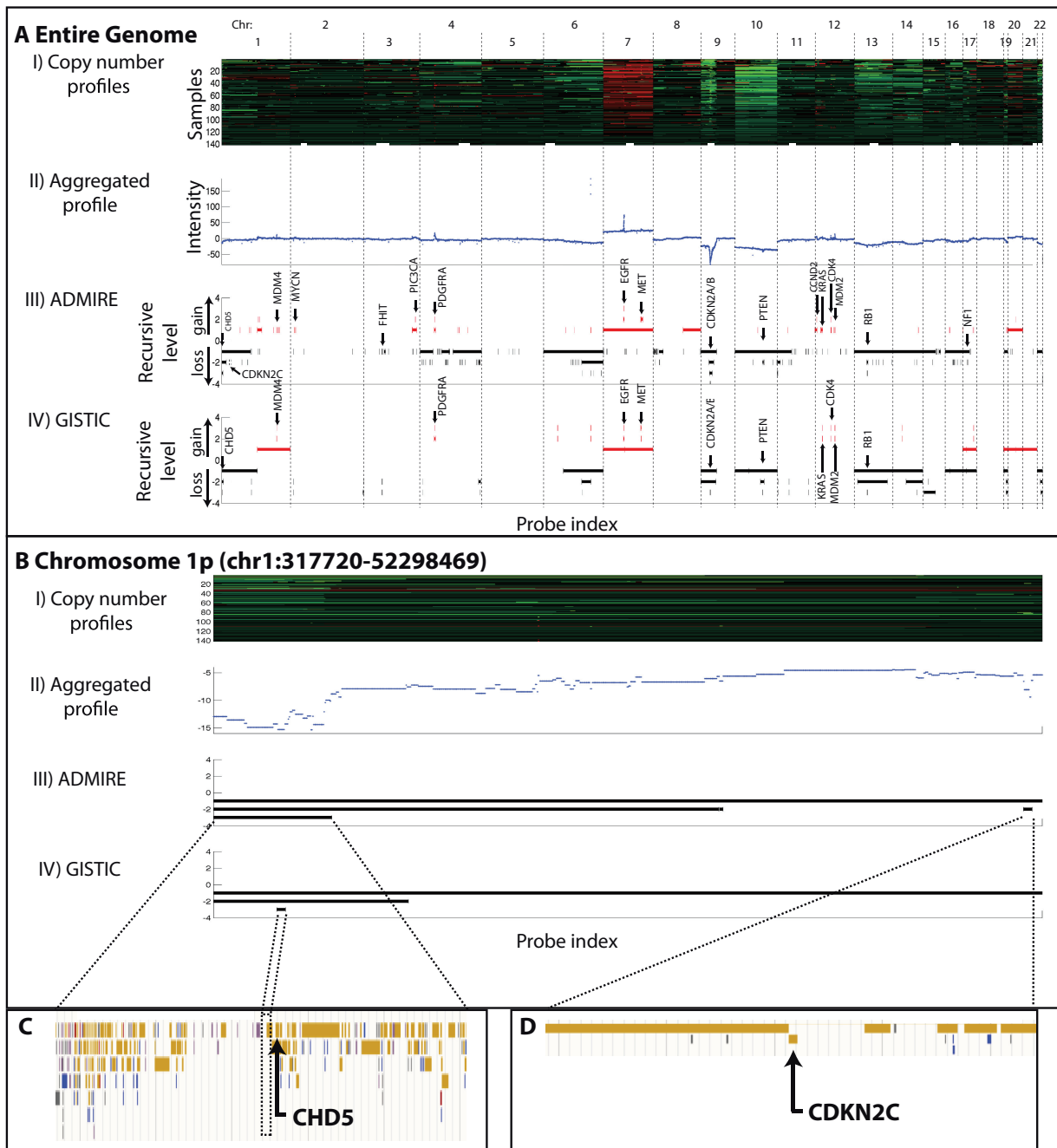
#### Application on glioma data

We compare the recurring events found by both ADMIRE and the latest version of GISTIC2.0 at 25% FDR on the glioma dataset described earlier. The results in Figure 8 reveal that ADMIRE finds many more events (in total 223 focal and broad events) than GISTIC2.0 (50 focal and broad events). All the known glioma tumor suppressors and oncogenes found by GISTIC2.0 are also recovered by ADMIRE. Although GISTIC2.0 performs

probe-based FDR, and is therefore expected to be optimistic (see Figure 4), there are many sources of power loss that are overcome by ADMIRE as follows:

- Substantial power is gained, as regions that are known to be significantly recurrent are ignored when estimating the null parameters;
- We account for the auto-correlation in the genomic profiles (in the null hypothesis), and as nearby probes reveal high positive correlations, the severity of multiple testing is reduced;
- By considering multiple scales (levels of smoothing), we gain substantial power for detecting broader events.

We give a multi-level representation of the events found by both ADMIRE and GISTIC2.0 in Figure 8. GISTIC2.0 dichotomizes events into focal and broad



**Figure 8.** Comparison of detected recurring events detected by ADMIRE and GISTIC2.0 on the glioma dataset. (A) Summary of the recurrent aberrations found by both ADMIRE and GISTIC2.0 on the entire genome. (A.I) The SNP array profiles for 141 glioma samples. Red (green) represents amplifications (deletions). (A.II) The sum of all the SNP array profiles. (A.III) A multi-level representation of the recurring events found by ADMIRE at 25% event-based FDR. The first recursive level shows all the broad and focal events that are not embedded in broad events. The second level shows more focal (or less broad) events embedded in broad first-level events, etc. (A.IV) Results found by GISTIC2.0 at 25% probe-based FDR. The first level (+1/-1 for gains or losses, respectively) represents all the broad recurrent events found at the chromosome arm level. After removing segments that stretch across whole chromosome arms, all segments with q-values below 0.25 are represented on the second level. Finally, focal regions are detected using the RegBouncer algorithm and represented on the third level. Therefore, red events (positive levels) represent recurring gains (levels move upwards) and black (negative levels) represents deletions (with levels moving downwards). (B) A zoom of the result in Panel A, showing the first part of chromosome 1p. (C) The top recursive level (most focal) event found by ADMIRE containing the CHD5 gene. It is interesting to note that GISTIC2.0 finds a much more focal area close to CHD5; however, with careful observation of the aggregated profile in (B.II) it is obvious that no focal event can be called with high significance by ADMIRE at this point. (D) Shows the recurring region found by ADMIRE containing the known glioma tumor suppressor gene CDKN2C that was missed by GISTIC2.0.

(chromosome arm-length) recurrences. All events found by GISTIC2.0 on a chromosome-arm level are indicated on the first level (+1 for gains or -1 for losses). After removing aberrations that stretch across whole chromosome arms, GISTIC2.0 also finds probes that are significantly recurrent with *q*-values below the 25% probe-based FDR level. All regions defined by these probes are represented on a second level (+2 for gains and -2 for losses). GISTIC2.0 uses an arbitrated peel-off algorithm to identify multiple potential target regions inside each significant region below the *q*-value threshold. The boundaries of these regions are then fine-tuned using an algorithm called RegBouncer (5). These regions are then represented on the third level. In contrast, ADMIRE makes no such distinction and simply adds levels until convergence. ADMIRE only adds more focal regions on a higher level if it can be proved significantly recurrent (below 25% event-based FDR) with respect to its immediate background (the level below).

Visually it is clear that ADMIRE shows an increase in power for detecting broad events (due to the multi-scale approach), as can be seen, for example, when looking at the third level of recurrent deletions in chromosome 1p in Figure 8B.III (containing CHD5). In contrast, GISTIC2.0 only finds a focal recurrent aberration (close to CHD5). The aggregated profile in Figure 8B.II reveals that indeed the broad event (third recursive level) detected by ADMIRE is likely a real event (of the same width), but it is difficult to prove significance of the focal event found by GISTIC2.0 relative to this background. It is possible to look for maximal peaks inside the broad event to help guide us towards genes that are likely relevant, but cannot be significantly distinguished from neighboring genes. In this sense, ADMIRE is more conservative at detecting focal events than GISTIC2.0.

One can argue that it is important to detect broad events with high power (justifying the multi-scale methodology). To see why, consider a single scale analysis (with little or no smoothing). One might not have the necessary power to detect some broad events; however, random (passenger/measurement noise) focal events that surpass the threshold (in combination with the broad event) will lead to shattered positives. In contrast, the multi-scale procedure will likely detect the broad event, and if not, we regard the overlapping focal events (that surpass the threshold) to be non-random (with respect to its immediate background).

ADMIRE detects a number of focal events that are missed by GISTIC2.0, including two events involving known glioma tumor suppressor genes: CDKN2C and NF1. The focal recurrent event overlapping with CDKN2C is showcased in Figure 8D. NF1 is showcased in Supplementary Figure S4.

## DISCUSSION

ADMIRE is an algorithm designed to assist in the discovery of broad and focal (potentially overlapping) recurring events. It does not require segmentation of single sample genomic profiles and therefore admits heterogeneous

samples that do not display clear breakpoints in copy number. ADMIRE performs a kernel smoothing methodology on the aggregated profile that optimizes the power for detecting recurring events if the null hypothesis closely resembles a Gaussian random process. Our previous algorithm, KC-SMART, is an example of another kernel smoothing methodology. Compared with KC-SMART, ADMIRE shows a drastic increase in power, especially for focal aberrations, when we fix the FWER at 5% (see Supplementary Figure S2).

Furthermore, ADMIRE performs analytical event-based FDR control instead of probe-based FDR. The user thus receives a list of recurrent regions for which the expected proportion of false regions is lower than that specified by the FDR.

From a technical perspective, ADMIRE gains power in detecting recurring events by accounting for the auto-correlation between probes (reduces the severity of multiple testing), performing a multi-scale smoothing methodology (especially helps for detecting broad events) and perhaps most importantly by estimating the behavior of passenger events (the null hypothesis) in regions that do not contain known recurrent events. Although it might be regarded as unimportant to detect broad events with high power (as focal events are expected to be of greater importance when searching for relevant genes), we argue that this is of central importance, as one might expect that for every broad event missed, a number of potentially false focal events might be detected in this region simply due to passenger events revealing peaks in an elevated region (shattered events) in the aggregated profile.

We introduced an analytical expression for the expected Euler characteristic, which simply counts up-crossings and not explicitly how long the signal remains above the amplitude threshold (the so-called sojourn time). Intuitively this could present a problem, but ADMIRE solves the problem by using the scale space to automatically tune the power to match the aberration width.

We also introduced a method that allows us to control the FDR (based on the expected Euler characteristic) without resorting to time-consuming permutation tests. We are therefore able to perform complex procedures, such as updating the null-process parameters in the recursive multi-level detection scheme, within a realistic time frame.

The methodology is justified from a theoretical perspective and justified with empirical simulations. Also when we test the method on a glioblastoma dataset, we find many more potentially interesting recurrent events (including two known glioma tumor suppressors CDKN2C and NF1) that approximately form a superset of those found by GISTIC2.0. Note that ADMIRE does not make a binary distinction between broad and focal events since multiple levels of increasingly focal events are derived from the data.

On a final note, the amount of primary memory used by ADMIRE depends on the probe locations and the minimum kernel width specified. If the whole human genome is covered with probes (say 3 million or more probes) and the minimum kernel width specified is 1 kb, the maximum memory usage will be 2 GB, which might be

smaller than the dataset itself. Computation time is largely influenced by the number of recurrent aberrations detected, which might take up to 8 h on an Intel Core i7-950 processor for a dataset consisting of 3 million probes, 200 samples and 200 recurrent events.

## AVAILABILITY

ADMIRE can be downloaded at <http://bioinformatics.nki.nl/admire/>. This includes a zipped file with the required Matlab code and glioma dataset.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Methods, Supplementary Figures 1–4 and Supplementary References [18,19].

## ACKNOWLEDGEMENTS

The authors thank Jeroen de Ridder and Theo Knijnenburg for valuable discussions regarding different aspects of the work. Special thanks go to Guillem Rigauil for checking and providing invaluable insight in many theoretical parts.

## FUNDING

This work was part of the BioRange programme of the Netherlands Bioinformatics Centre (NBIC), which is supported by the Netherlands Genomics Initiative (NGI). Funding for open access charge: The Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands.

*Conflict of interest statement.* None declared.

## REFERENCES

- Rouveirol, C., Stransky, N., Hupé, P., La Rosa, P., Viara, E., Barillot, E. and Radvanyi, F. (2006) Computation of recurrent minimal genomic alterations from array-CGH data. *Bioinformatics*, **22**, 849–856.
- Diskin, S.J., Eck, T., Greshock, J., Mosse, Y.P., Naylor, T., Stoeckert, C.J.Jr., Weber, B.L., Maris, J.M. and Grant, G.R. (2006) STAC: A method for testing the significance of DNA copy number aberrations across multiple array-CGH experiments. *Genome Res.*, **16**, 1149–1158.
- Shah, S.P., Lam, W.L., Ng, R.T. and Murphy, K.P. (2007) Modeling recurrent DNA copy number alterations in array CGH data. *Bioinformatics*, **23**, i450–i458.
- Beroukhir, R., Getz, G., Nghiemphu, L., Barretina, J., Hsueh, T., Linhart, D., Vivanco, I., Lee, J.C., Huang, J.H., Alexander, S. *et al.* (2007) Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc. Natl Acad. Sci. USA*, **104**, 20007–20012.
- Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhir, R. and Getz, G. (2011) GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.*, **12**, R41.
- Sanchez-Garcia, F., Akavia, U.D., Mozes, E. and Pe'er, D. (2010) JISTIC: identification of significant targets in cancer. *BMC Bioinformatics*, **11**, 189.
- Klijn, C., Holstege, H., de Ridder, J., Liu, X., Reinders, M., Jonkers, J. and Wessels, L. (2008) Identification of cancer genes using a statistical framework for multiexperiment analysis of nondiscretized array CGH data. *Nucleic Acids Res.*, **36**, e13.
- Rueda, O.M. and Diaz-Uriarte, R. (2010) Finding recurrent copy number alteration regions: a review of methods. *Curr. Bioinformatics*, **5**, 1–17.
- Adler, R.J. and Hasofer, A.M. (1976) Level crossings for random fields. *Ann. Probab.*, **4**, 1–12.
- Adler, R.J. (1976) On generalising the notion of upcrossings to random fields. *Adv. Appl. Probab.*, **8**, 789–805.
- Worsley, K.J. (1995) Estimating the number of peaks in a random field using the Hadwiger characteristic of excursion sets, with applications to medical images. *Ann. Stat.*, **23**, 640–669.
- Worsley, K.J., Marrett, S., Neelin, P., Vandal, A.C., Friston, K.J. and Evans, A.C. (1996) A unified statistical approach for determining significant signals in images of cerebral activation. *Hum. Brain Mapp.*, **4**, 58–73.
- Worsley, K.J., Andermann, M., Koulis, T., MacDonald, D. and Evans, A.C. (1999) Detecting changes in nonisotropic images. *Hum. Brain Mapp.*, **8**, 98–101.
- Nichols, T. and Hayasaka, S. (2003) Controlling the familywise error rate in functional neuroimaging: a comparative review. *Stat. Methods Med. Res.*, **12**, 419–446.
- Benjamini, Y. and Yekutieli, D. (2001) The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.*, **29**, 1165–1188.
- Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B*, **57**, 289–300.
- Hupé, P., Stransky, N., Thiery, J.P., Radvanyi, F. and Barillot, E. (2004) Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. *Bioinformatics*, **20**, 3413–3422.
- Berdiev, B.K., Xia, J., McLean, L.A., Markert, J.M., Gillespie, G.Y., Mapstone, T.B., Naren, A.P., Jovov, B., Bubien, J.K., Ji, H.L. *et al.* (2003) Acid-sensing ion channels in malignant gliomas. *J. Biol. Chem.*, **278**, 15023.
- Vila-Carriles, W.H., Zhou, Z.H., Bubien, J.K., Fuller, C.M. and Benos, D.J. (2007) Participation of the chaperone Hsc70 in the trafficking and functional expression of ASIC2 in glioma cells. *J. Biol. Chem.*, **282**, 34381–34391.