SUPPLEMENTARY METHODS

Analytical expression for the Euler characteristic

In this Section we derive an analytical expression for the expected number of events \( E[\chi] \) found above a fixed threshold and kernel width (see Equation 11 in the main text) for a suitably regular (as defined later) non-homogenous Gaussian (multi-variate) random process \( H(g) \) in a compact domain \( a = \{ g : b_1 \leq g \leq b_r \} \), where \( b_1 \) and \( b_r \) are constant boundaries (for example, the boundaries of a chromosome arm). Furthermore, we assume that \( H(g) \) has mean zero and variance one at any given location \( g \). Therefore the random process is uniquely defined by the non-homogenous correlation function \( r(g_1, g_2) \) or alternatively \( r_g(\Delta g) = r(g, g + \Delta g) \). From now on, any realization of the random process \( H(g) \) will be represented with a lowercase \( h(g) \).

The null process derived in Equation 9 meets the required criterion. However, as will be shown later we do not need an explicit solution for \( r_g(\Delta g) \). Instead, the same information is captured by the expected variation in the derivative of \( H(g) \) (see Equation 12).

Much work has been done on finding the expected number of events above a fixed threshold for homogeneous Gaussian random fields of any dimensionality (1, 2). Little work has been done on treating non-homogenous fields except for work done in neuro-imaging (3). Unfortunately, this work is not directly compatible with our application and therefore we derive the necessary Equation to relate \( E[\chi] \) with the desired threshold.

To start, we defined the excursion set \( a^+ \) of any realization \( h(g) \) as the region (a subset of \( a \)) where \( h(g) \) is higher than or equal to the fixed threshold \( t \):

\[
a^+(h,t) = \{ g \in a : h(g) \geq t \} \tag{S1}\]

We define the ‘number of events’ in \( a \) to be the number of maximally connected subsets (ordered by inclusion) of \( a^+ \) and denote it with the one dimensional Euler characteristic \( \chi(h,t,a) \). It is clear that \( \chi(h,t,a) \) is related to the number of up-crossings on \( t \) if \( h(g) \) is continuous. If, for any sample function \( h(g) \), \( h(b_1) \) is smaller than \( t \) (on the left boundary), the Euler characteristic is equal to the number of up-crossings. On the other hand if \( h(b_1) \) is higher than or equal to the threshold the Euler characteristic will be the number of up-crossings plus one. We denote the number of up-crossings with \( \chi_{DT}(h,t,a) \) since this is exactly the differential topology (DT) characteristic for a one dimensional function (1). \( \chi(h,t,a) \) and \( \chi_{DT}(h,t,a) \) are related as follows:

\[
\chi(h,t,a) = I(h(b_1) - t) + \chi_{DT}(h,t,a), \tag{S2}\]

where

\[
I(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{S3}\]

We are interested in evaluating the expected Euler characteristic:

\[
E[\chi(H,t,a)] = P[H(b_1) \geq t] + E[\chi_{DT}(H,t,a)]
= \frac{1}{2} \text{erfc}(\frac{t}{\sqrt{2}}) + E[\chi_{DT}(H,t,a)] \tag{S4}\]

Suitably regular processes We denote the first and second order derivatives (if they exist) of the process \( H(g) \) by \( H'(g) \) and \( H''(g) \) respectively. We define a random process to be suitably regular on \( a \) (for a fixed threshold \( t \)) if the following conditions hold as described by (2) and adapted for our one dimensional application:

1. A sample function \( h(g) \) has a.s. (almost surely) continuous derivatives up to second order with finite variance in an open neighbourhood of \( a \). This condition is satisfied if we use kernels for smoothing with the same requirements (such as a Gaussian kernel).

2. The set \( \{ g \in a : h(g) = t \land h'(g) = 0 \} \) is a.s. empty. In other words the probability density function \( p[H'(g)]H(g) = t \) should be finite at zero and the set \( \{ g \in a : h(g) = t \} \) a.s. finite.

3. The set \( \{ g \in \{ b_1, b_r \} : h(g) = t \} \) is a.s. empty.

A random process is therefore considered to be suitably regular if any realization \( h \) is a.s. sufficiently smooth (ensured by the kernel convolution) and consist of a finite number of points where \( t \) is crossed, the derivative being non zero at each point.

First we observe that a suitably regular Gaussian random process has a differentiable correlation function \( r_g(\Delta g) \) with respect to \( \Delta g \) up to second order. We also observe that:

\[
\text{Var}\left(\frac{d}{dg} H(g)\right) = \lim_{\Delta g \to 0} \text{Var}\left[\frac{H(g + \Delta g) - H(g)}{\Delta g}\right]
= \lim_{\Delta g \to 0} \frac{1}{(\Delta g)^2} E\left[\left(H(g + \Delta g) - H(g)\right)^2\right]
= \lim_{\Delta g \to 0} \frac{2(1 - r_g(\Delta g))}{(\Delta g)^2}
= -\lim_{\Delta g \to 0} \frac{r'_g(\Delta g)}{\Delta g}
= -\lim_{\Delta g \to 0} r''_g(\Delta g), \tag{S5}\]

where \( r'_g \) and \( r''_g \) represent the first and second order derivative \( r_g \) with respect to \( \Delta g \). In the last two steps we consecutively applied L’Hôpital’s rule.

Therefore, \( r'_g \) comply with the following properties:

\[
r_g(0) = 1
r'_g(0) = 0
r''_g(0) = -\text{Var}\left(\frac{d}{dg} H(g)\right) \tag{S6}\]

The expected number of up-crossings Before we derive an analytical expression for the expected number of up-crossing, let’s first derive a few useful lemmas.

Let us define a function \( q : \mathbb{R} \to \mathbb{R} \):

\[
q(r) = \lim_{\Delta g \to 0^+} \frac{(k \circ r)(\Delta g)}{\Delta g}, \tag{S7}\]
where
\[
k(r) = \frac{1}{2\pi |\Sigma(r)|^{1/2}} \times \int_{t}^{0} e^{-\frac{1}{2}r^\top [\Sigma^{-1}(r)] r} dxdy
\]
\[
\Sigma(r) = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}
\] (S8)
and \( F \) is the set of all functions \( r \) for which the limit is finite.

**LEMA 1.** Consider a suitably regular random Gaussian process \( H \) with \( \forall g \in \mathbb{R} \) \( H(g) \sim N(0,1) \) and correlation function \( r_e(\Delta g) = r(g,g+\Delta g) \). If \( \forall g \in a, r_g \in F \), then
\[
E[\chi_{DT}(h,t,a)] = \int_{b_t}^{b_h} q(r_g) dg
\] (S9)

**PROOF.** The DT characteristic is simply the number of points \( g \in a \) satisfying the following conditions:

1. \( h(g) = t \)
2. \( h'(g) > 0 \)

Note that the number of points satisfying \( h(g) = t \) for \( g \in a \) representing up and down crossings (say \( c_1 < c_2 < \ldots < c_k \)) are a.s. finite and distinct due to the regularity conditions. Let us define grid points via sequences \( a^g : 2^n \to a \) such that:
\[
a^g(0) = b_t \]
\[
a^g(i + 1) = a^g(i) + \Delta g^n,
\] (S10)
where \( \Delta g^n = \frac{b_h - b_t}{2^n} \) and denote the set of grid points by \( \{a^g\} = \{a^g(i) | i < 2^n\} \). It is easy to prove by induction that for any \( k,l \in \mathbb{N}, k < l \) implies that \( \{a^g\} \subseteq \{a^g\} \).

Next, let us define the infinite sequence \( \{\chi_n | n < \mathbb{N}\} \), with elements
\[
\chi_n = \sum_{i=0}^{n-1} f^n(h,t,a^n(i)) \Delta g^n,
\] (S11)
where
\[
f^n(h,t,g) = \begin{cases} 1 & \text{if } h(g) < t \leq h(g + \Delta g^n) \\ 0 & \text{otherwise} \end{cases}
\] (S12)

It is rather easy to show that the sequence \( \{\chi_n | n < \mathbb{N}\} \) is:

- increasing;
- bounded from above by \( \chi_{DT}(h,t,a) \);
- there a.s. exist an \( n \in \mathbb{N} \) for which \( \chi_n = \chi_{DT}(h,t,a) \), since there exist a.s. only a finite number of crossings;

Therefore we conclude that:
\[
\lim_{n \to \infty} \chi_n = \lim_{n \to \infty} \sum_{i=0}^{n-1} f^n(h,t,a^n(i)) \Delta g^n
\]
\[
= \chi_{DT}(h,t,a)
\] (S13)

Now let us consider the expected value of and given \( \chi_n \):
\[
E[\chi_n] = \sum_{i=0}^{n-1} E[f^n(h,t,a^n(i))] \Delta g^n,
\] (S14)
It is clear at this point that the sequence \( \{E[\chi_n] | n < \mathbb{N}\} \) will converge to \( E[\chi_{DT}(h,t,a)] \) since, for any realization of the process, \( \{\chi_n | n < \mathbb{N}\} \) converges to \( \chi_{DT}(h,t,a) \). Therefore we set out to compute
\[
E[\chi_{DT}(h,t,a)] = \lim_{n \to \infty} E[\chi_n]
\]
\[
= \lim_{n \to \infty} \sum_{i=0}^{n-1} E[f^n(h,t,a^n(i))] \Delta g^n
\]
\[
= \int_{b_t}^{b_h} \lim_{n \to \infty} E[f^n(h,t,g)] dg
\] (S15)
if the limit inside the integral is finite for all \( g \in a \). Therefore the problem is reduced to finding the expectation of \( f^n(h,t,g) \), which is fully determined by the joint probability density function of \( x = h(g) \) and \( y = h(g + \Delta g^n) \) with correlation \( r = r_g(\Delta g^n) \) (see Equation S12):
\[
p(x,y) = \frac{1}{2\pi |\Sigma(r)|^{1/2}} e^{-\frac{1}{2}[(x-x^*)\Sigma^{-1}(r)(x-x^*)]}
\] (S16)
where
\[
\Sigma(r) = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}
\] (S17)

From Equation S12 we get:
\[
E[f^n(H,t,g)] = \frac{1}{\Delta g^n} P[x < t, y \geq t]
\]
\[
= \frac{(k \circ r_g)(\Delta g^n)}{\Delta g^n}.
\] (S18)

We are interested in finding an expression for (if it exists):
\[
q(r_g) = \lim_{n \to \infty} E[f^n(H,t,g)] = \lim_{\Delta g \to 0^+} \frac{(k \circ r_g)(\Delta g)}{\Delta g}
\] (S19)
This concludes the proof of lemma 1.

**LEMA 2.** Consider two functions \( r_1 \) and \( r_2 \) that are differentiable up to the second order in open intervals \((0, \delta_1)\) and \((0, \delta_2)\) respectively, where \( \delta_1, \delta_2 > 0 \). Let us assume that \( r_1 \in F \) and that:
\[
\lim_{g \to 0^+} r_1(g) = \lim_{g \to 0^+} r_2(g) = 1
\]
\[
\lim_{g \to 0^+} r'_1(g) = \lim_{g \to 0^+} r'_2(g) = 0
\]
\[
\lim_{g \to 0^+} r''_1(g) = \lim_{g \to 0^+} r''_2(g) < 0
\] (S20)

If these conditions hold, then \( r_2 \in F \) and \( q(r_1) = q(r_2) \).
PROOF. Note that there exist $\delta_1$ and $\delta_4$ such that $0 < \delta_1, \delta_4 < \min(\delta_3, \delta_2)$ such that $r_1$ and $r_2$ have negative first and second order derivatives in the intervals $(0, \delta_3)$ and $(0, \delta_4)$ respectively and furthermore the images are equal $r_1([0, \delta_3]) = r_2([0, \delta_4]) = (e, 1)$ for some $e \in (0, 1)$. Since $r_2$ is decreasing in the interval $(0, \delta_3)$, its inverse exists and we define a function with domain $(0, \delta_3)$ and range $(0, \delta_4)$:

$$w(g) = (r_2^{-1} \circ r_1)(g) \quad (S21)$$

First note that $w(g)$ is an increasing function, since both $r_1$ and $r_2^{-1}$ are decreasing. Also, as expected:

$$\lim_{g \to 0^+} w(g) = \lim_{g \to 0^+} r_2^{-1}(r_1(g)) = \lim_{r_1 \to 1^-} r_2^{-1}(r_1) = 0 \quad (S22)$$

Furthermore, for $g \in (0, \delta_3)$:

$$\frac{dw}{dg} = \frac{dr_1}{dg} \cdot \frac{dr_2}{dw} \quad (S23)$$

Note that the denominator $\left(\frac{dr_2}{dw}\right)$ and its derivative (with respect to $g$) is strictly non-zero in the interval $g \in (0, \delta_3)$ since $r_2$ have negative first and second order derivatives. Also note that $\lim_{g \to 0^+} \frac{dr_1}{dg} = \lim_{g \to 0^+} \frac{dr_2}{dw} = \lim_{g \to 0^+} \frac{dw}{dg} = 0$ (See Equation S20). Therefore we can apply L'Hôpital:

$$\lim_{g \to 0^+} \frac{dw}{dg} = \lim_{g \to 0^+} \frac{\frac{dr_1}{dg}}{\frac{dr_2}{dw}} = \frac{1}{\lim_{g \to 0^+} \frac{dw}{dg}} = 1 \quad (S24)$$

Where we noticed that $\lim_{g \to 0^+} \frac{dr_1}{dg} = \lim_{g \to 0^+} \frac{dr_2}{dw} = \lim_{g \to 0^+} \frac{dw}{dg} = 0$ (See Equation S20).

To summarize,

- $w$ has domain $(0, \delta_3)$ and range $(0, \delta_4)$;
- $w$ is an increasing function with first and second order derivatives;
- $\lim_{g \to 0^+} w(g) = 0$;
- $\lim_{g \to 0^+} w'(g) = 1$;

On a different note, $k$ is clearly differentiable in the domain $(\epsilon, 1)$ since $\Sigma$ is positive definite.

Now we have everything we need to prove the lemma. For any function $r_2$ (including $r_1$) satisfying Equation S20:

$$\lim_{g \to 0^+} (k \circ r_2)(g) = 0 \quad (S25)$$

since $r_2 \to 1^-$ and $\Sigma$ becomes singular. Therefore by L'Hôpital:

$$q(r_2) = \lim_{g \to 0^+} \frac{dk}{dr_2} \frac{dr_2}{dg} = 0 \quad (S26)$$

if the limit exist.

We know that $q(r_1)$ is finite (since $r_1 \in \mathcal{F}$) and again by L'Hôpital:

$$q(r_1) = \lim_{g \to 0^+} \frac{(k \circ r_2 \circ w)(g)}{g} = \lim_{g \to 0^+} \frac{dr_2}{dw} \frac{dw}{dg} = 0 \quad (S27)$$

which concludes the the proof for lemma 2.

**LEMMA 3.** For a stationary correlation function $r$:

$$q(r) = e^{-r^2/2} / 2\pi \sqrt{-r''(0)} \quad (S28)$$

**PROOF.** Consider a suitably regular stationary random Gaussian process $H$ with correlation function $r$ and $\forall g \in \mathbb{R} H(g) \sim N(0, 1)$. Lemma 1 states that if $q(r)$ is finite then

$$E[\mathcal{L}_D|H, t, [0, 1]] = \int_0^1 q(r)dg = q(r) \quad (S29)$$

In the second step we used the fact that $r$ is constant for all $g$ in a stationary process.

Also, by Equation 3.2 of Theorem 3.1 in (2) when applied to a one dimensional field:

$$E[\mathcal{L}_D|H, t, [0, 1]] = e^{-r^2/2} / 2\pi \sqrt{-r''(0)} \quad (S30)$$

**THEOREM 1.** Consider a suitably regular (non-stationary) random Gaussian process $H$ with $\forall g \in \mathbb{R} H(g) \sim N(0, 1)$ and correlation function $r_1(\Delta g) = r(g, g + \Delta g)$. Then

$$E[\mathcal{L}_D|H, t, a] = e^{-r^2/2} / 2\pi \int_{\beta_r} \sqrt{\text{Var}[dH(g)]dg} \quad (S31)$$

**PROOF.** According to Lemma 1, we simply need to find an expression for $q(r_2)$ for all $g$. The properties specified in Equation S6 hold for all $r_5$. Certainly for each such $r_5$ there
exist a stationary \( r^*_d \) with the same properties. By Lemma 2 and 3:

\[
q(r^*_d) = q(r^*_d) = e^{-r^*_d/2} \sqrt{-r^*_d} f(0) = e^{-r^*_d/2} \sqrt{-r^*_d} f(0)
\]

\[
= e^{-r^*_d/2} \sqrt{-r^*_d} f(0) = e^{-r^*_d/2} \sqrt{-r^*_d} f(0)
\]

\[
= e^{-r^*_d/2} \sqrt{-r^*_d} f(0) = e^{-r^*_d/2} \sqrt{-r^*_d} f(0)
\]

\[
\text{Details on multi-scale detection}
\]

Resolution parameter \( \alpha \): The filter parameter \( \alpha \) is related to the minimum possible spatial overlap between a detected event \( D \) and a real recurrent event \( R \), which we indicate with the similarity coefficient \( J_1 \):

\[
J_1(D, R) = \frac{|D \cap R|}{|D|} \tag{S33}
\]

To see this, consider for example a Gaussian kernel truncated at \(+3\sigma\) (the kernel weights after \(+3\sigma\) are negligible and can safely be ignored for computational purposes) and a true recurrent event \( R \) with unknown boundaries \( g_s \) and \( g_e \). Clearly at a fixed scale \( w \) the smoothed signal is influenced by the recurrent event on the interval \([g_s - 3\sigma, g_e + 3\sigma]\). Therefore if we detect a recurrent event larger than \(3\sigma\) we are guaranteed that \( J_1(D, R) \geq \frac{\alpha}{\Delta} \cdot 100\% \) unless a false positive occurs. For a symmetric kernel that is decreasing with distance from the center, the minimum overlap is restricted to \( \frac{\alpha - 2\kappa}{\alpha} \cdot 100\% \), where we choose \( \kappa \) such that:

\[
\frac{\int_{-\infty}^{\infty} k_w(g)dg}{\int_{-\infty}^{\infty} k_w(g)dg} \ll 1 \tag{S34}
\]

The idea is simply to keep the kernel weights low at a distance \( \kappa \) \((\kappa = 3 \text{ for a Gaussian kernel})\) so that the contribution from a significant recurrence at that distance is negligible in the null region.

Small \( \alpha \) means that the aberration widths are allowed to be comparable in size to the kernel width. Figure 6 in the main text shows that the power increases when we allow for the kernel width to be similar in size of the aberration. However, the resolution \( (J_1) \) decreases.

If we let \( \alpha \rightarrow \infty \) we gain good resolution, but the power would be equivalent to that in a single scale analysis (at the smallest scale). As a compromise we choose \( \alpha = 20 \) that guarantees a minimum overlap \( J_1 \) of 70\% (a parameter that the user can vary). In the final step of the multi-scale selection we take the union of all recurrent events surviving the \( \alpha \) filtering and as a consequence all maximally connected regions resulting from the union also have a minimum overlap of 70\% (or whatever specified by the user) with true events. If the smallest kernel width considered is sufficiently small, we do not particularly care whether this overlap is high since we expect the detected event to be sufficiently zoomed into potentially interesting driver genes.

Usually, it is not a good idea to sacrifice power for a high resolution by letting \( \alpha \rightarrow \infty \) (single scale analysis), because large events might be shattered into smaller pieces as illustrated in Figure 2.B. Although shattered events are strictly not false positives, we might employ a more strict overlap condition such as the Jaccard similarity coefficient to call true positives:

\[
J_2(D, R) = \frac{|D \cap R|}{|D \cup R|} \tag{S35}
\]

For a small shattered event \( D, J_1 = 1.0 \) and \( J_2 \) would be small. We therefore define two different types of undesirable events for an overlap threshold \( o \) (e.g. 70\%):

- False positives: Events for which \( J_1 < o \);
- Shattered positives: Events for which \( J_1 \geq o \) and \( J_2 < o \);

For a fixed threshold \( o \) the shattered positive rate will decrease for a decreasing value of \( \alpha \), whereas the FDR will remain unchanged (for false positives) as long as \( \alpha \) remains large enough to ensure a high \( J_1 \).

Simulation results where we vary the \( \alpha \) parameter is illustrated in the supplementary section entitled ‘Resolution parameter \( \alpha \) on simulated data’. Here we also observe the effect that \( \alpha \) has on the shattered positive rate.

The expected number of events It is important to ask what effect the proposed multi-scale detection has on the expected number of significant regions (from now on referred to as \( E(\mathcal{X}) \)) found in the null-hypothesis if we keep \( E(\mathcal{X}) = E(\mathcal{X}(H_0^0, t)) \) constant and the same on each scale. First note that if we consider a very small kernel (say 1 kbp) then we are performing almost no smoothing and any region above a fixed threshold is considered significant with a high resolution. In this case our estimate on \( E(\mathcal{X}(H_1^0, t)) \) will be accurate (since we don’t remove any segments that don’t survive the 20\% filter rule). On the other hand, if for larger kernels, we only retain regions that are at least 20\% in size, our estimate \( E(\mathcal{X}(H_1^0, t)) \) will be much higher then the actual expectation on that particular scale. For any region \( a = [g_s, g_e] \) let us define \( \gamma_a \) to be the total number of events found in \([g_s, g_e]\) on the smallest kernel considered. Furthermore, let \( G(a) \) be the property ‘\( a \) is a detected event for some \( H_a(g) \) and survives the 20\% filtering rule’. We make the following assumption on the null-hypothesis:

\[
P(\mathcal{X} > 1|G(a)) > P(\mathcal{X} = 0|G(a)) \tag{S36}
\]

In effect we assume that if there exist a large region that is considered significantly elevated, then we expect the chance to be higher for this region to contain multiple significant ‘hits’ on a small scale then no ‘hits’ at all.

If this assumption holds, then clearly the multi-scale procedure will more likely merge events on the smallest possible scale than create new ones on a larger scale, and since we control the expected number of events on the smallest scale at \( E(\mathcal{X}) \), we expect:

\[
E(\mathcal{X}) < E(\mathcal{X}) \tag{S37}
\]

and therefore have \( E(\mathcal{X}) \) under control.
Details on updating the null-parameters

Consider a smoothed aggregated profile with no recurrent events which can be modeled with a Gaussian random process with parameters $\mu$, $\sigma$ and $r$. In the main text we proposed a methodology to iteratively update these null-parameters by detecting peaks that surpass a given threshold $t$ and re-estimating the parameters by ignoring these significant regions. Even with no recurrent events there is always a possibility of detecting false positives and consequently an updated estimate on $\sigma$ will be biased (and generate optimistic results) after one or more iterations. We set out to prove that this bias (assuming there are no recurrent events) will converge after infinite iterations and that the effect on the estimated number of false events ($E[\chi]$) will be negligible for reasonable thresholds $t$. If recurrent events are present, $\sigma$ will be a conservative estimate and therefore we consider only the worst case scenario with no recurrences.

Consider the variance of the random Gaussian process $\sigma$ and the respective threshold $t$ to control $E[\chi]$. Furthermore, assume that $t_i$ is the expected threshold on iteration $i$. If there are no recurrent events (unknown for real data), $t_1 = t$, the desired threshold. On the second iteration, all measurements larger than $t_1$ will be ignored when calculating a new variance estimate and will therefore be biased. In fact, we can predict how $t_i$ changes across each iteration:

$$t_{i+1} = t \sqrt{1 - \text{erfc}(\frac{t_i}{\sqrt{2\sigma}})} - \sqrt{\frac{2t}{\pi \sigma} e^{-\frac{1}{4} (\sigma^2)}} \tag{S38}$$

Using Equation 11 we can also predict how the estimate on $E[\chi]$ increase with each iteration. Therefore, for a fixed starting threshold $t$ we can predict the ratio between the predicted $E[\chi]$ after infinite iterations and the $E[\chi]$ for the desired threshold $t$:

$$R(t/\sigma) = e^{\frac{1}{4}(\frac{t}{\sigma})^2 - \frac{1}{8}(\frac{t}{\sigma})^4}} \tag{S39}$$

$R$ is a decreasing function of $t$ strictly larger than one that only converge for values $t/\sigma > 2.1617$. In practice we are typically interested in thresholds $t/\sigma > 3$ and therefore $R < 1.1632$. This illustrates that the iterative procedure only leads to a slight over-estimation of $E[\chi]$ by a factor of maximally 1.1632 (if no recurrent events are present). In real data, where many recurrent events are present, our estimate on $E[\chi]$ will likely remain conservative.

Details on recursive multi-level detection

Not only does the recursive multi-level detection procedure described in the text allow us to detect recurring events embedded in broad recurring events, but also helps to improve our estimate on $E[\chi]$ (the expected number of false recurring events found) when not all parts of the genome (the recurring events) can be described by the null-hypothesis. We illustrate this concept in Figure 3 in the main text. Note that the region in which we estimate the null-parameters $\mu$, $\sigma$ and $r$ is restricted to $H_0^{\text{FR}}$ in Figure 3A as illustrated by the dotted line at the top of the figure. However in Equation 11 (in the main text) we integrate across the whole genome and therefore the expected number of detected events in $H_0^{\text{FR}}$ will be lower than $E[\chi]$ (the expected number of events if the null-model held across the whole genome) by a factor $\int_{H_0^{\text{FR}}} \chi^2 / \int_G \chi^2$ (the erfc term is small).

In the second recursion step (Figure 3.B) we follow the exact same procedure except this time estimate the null-parameters in the broad event $H_0^{\text{BR}}$. This allows us to detect embedded focal events inside broader events. Again we make sure the test is genome wide (although we only estimate and detect events in $H_0^{\text{BR}}$, we still integrate across the whole genome in Equation 11 in the main text) and therefore the expected number of null-events found inside $H_0^{\text{BR}}$ will differ from $E[\chi]$ (the expected number of events if the null-model held across the whole genome for the new parameters) by a factor $\int_{H_0^{\text{BR}}} \chi^2 / \int_G \chi^2$. The total number of expected random events found in both recursion steps will be $\int_{H_0^{\text{FR}}} + \int_{H_0^{\text{BR}}} \chi^2 / \int_G \chi^2$. In fact, if we consider all recursion steps the expected number of falsely detected events (focal, broad and embedded) will approach $E[\chi]$ and we effectively avoid the need for step up/down multiple testing procedures. The only added assumption here is that the expected number of false focal events discovered inside false broad events is insignificant.
SUPPLEMENTARY RESULTS

Resolution parameter $\alpha$ on simulated data

**Summary** We investigated the effect that the resolution parameter $\alpha$ (which is related to the accuracy of event boundaries) has on the FDR and the power for detecting recurrent aberrations of different genomic lengths. We generated simulated data using the methodology proposed in the section entitled 'FDR simulations'. Specifically, we added $N_0 = 2$ broad (20 $\times$ 10$^6$ bps), $N_m = 5$ (2 $\times$ 10$^6$ bps) medium and $N_f = 50$ (100 $\times$ 10$^3$ bps, with an average of five probes) focal recurrent events. For each simulation, the recurrent events were placed at random locations (potentially overlapping) with a recurrence frequency (across the samples) randomly selected between zero and one. Passenger events and noise were simulated using the procedure described in the section entitled ‘Data sets’. We fixed the number of aCGH samples to aggregate to $S = 200$ and the analytical FDR level to 5%. We vary the SNR and $\alpha$ parameter. Finally, we distinguish between true and false positives based on an overlap threshold $\alpha = 70\%$.

**Results** Figure S1 illustrates the empirical FDR and power for detecting recurrent aberrations (on 1000 simulations) of different sizes, for different SNRs and $\alpha$ parameters. True and false positives are discriminated based on an overlap threshold $\alpha = 70\%$. Before we continue it is important to realize that for large $\alpha$ values we are effectively ignoring all scales except the smallest (i.e. we perform little smoothing), since none of the events detected at a larger scale will survive the $\alpha$ filtering. This is in contrast to $\alpha = 0$, where all detected events on all scales are combined.

In Figure S1.A.I we measure the expected proportion of detected events for which $J_1 < 70\%$ (false positives). As predicted in the supplementary section entitled ‘Details on multi-scale detection’, for $\alpha$ value above 20, this proportion is below 5% (the FDR). For $\alpha$ values smaller than 20 the accuracy on event boundaries become poor and the measured FDR grows high.

In Figure S1.A.II we measure the expected proportion of detected events that are true positives, but $J_2 < 70\%$ (shattered positives). It is important to observe that these errors are reduced when we allow for $\alpha$ to grow small (and effectively use the full scale space). This effect is observed even for segmented samples (SNR = 1e99). This illustrates that the scale space methodology is not only useful for reducing measurement noise, but also biological noise (i.e. passenger events).

In Figure S1.A.III we illustrate the measured FDR if we consider shattered positives to be false. That is, we regard all detected events with a Jaccard similarity index below 70% as false positives. It is interesting to note that for $\alpha = 20$, the FDR is close to minimal, but still higher than the desired FDR level of 5%. This is because the shattered positive rate cannot be controlled, but reduced due to the multi-scale analysis - an aspect ignored by single-scale methods. Furthermore, it is not surprising that $\alpha = 20$ is close to the minimum FDR, since we chose a corresponding overlap threshold of $\alpha = 70\%$.

In Figure S1.B.I we illustrate the average proportion of recurrent focal events (of which there are 50 per simulation) that are detected (Jaccard similarity coefficient above 70%) while varying $\alpha$. Note that for small $\alpha$ values we lose power. This is mainly due to low precision on event boundaries. For large $\alpha$ values we see a minimal reduction in power. This is because little value is added by the scale-space for such small events, since they are detected at small scales only.

In Figure S1.B.II we illustrate the power for detecting broad events. Again for low $\alpha$ values we lose considerable power due to low precision on event boundaries. However, for very large $\alpha$ values we also see a drastic decrease in power. This is mainly due to scattering and the multi-scale procedure becomes invalid.

Finally, in Figure S1.B.III we illustrate the power for detecting all events (focal and broad included). The power observed is similar to that in Figure S1.B.I since most events are focal per simulation (50 as apposed to two).

**Discussion** Not only do we reduce the shattered positive rate with the multi-scale procedure if we select $\alpha$ properly ($\alpha = 20$ for an overlap of 70%), but we also increase the power for detecting true positives (especially broad events). This is true even for segmented samples, which illustrates that smoothing not only reduces measurement noise but also biological noise (passenger events).

Furthermore, in Figure S5.C and D, we see that the multi-scale procedure reduces our estimate on $E[\chi]$ which will result in a conservative FDR estimate. However, this does not imply that the power for detecting recurrent events are reduced compared to a single-scale analysis (since we join events on all scales, you cannot lose events found at the smallest scale). This simulation study serves to illustrate that the multi-scale procedure (with proper selection of $\alpha$) can drastically increase the power for detecting events (especially broad events), which makes a conservative estimate on the FDR well worth it.

**KC-SMART vs. ADMIRE smoothing methodologies**

**Summary** Both KC-SMART (4) and the newly proposed method perform kernel smoothing on aggregated profiles and applies a constant threshold on the smoothed signal to decide which locations represent recurring aberrations or not. To make these methods directly comparable, we do not split positive and negative aberrations into two problems for KC-SMART.

The new method performs smoothing with Equation 9 (in the main text) and relies on data dependent parameters $\mu$, $\sigma$ and $r$ (defined in Equation 3 in the main text) and the platform dependent probe locations, whereas KC-SMART uses only probe locations to normalize the smoothed signal. KC-SMART uses the following normalization scheme:

$$h_{KC}(g) = \frac{f_w(g)}{k_w(g) \sum_{i=0}^{E-1} \delta(g - p_i)}$$ (S40)

The purpose of this experiment is to show that the new smoothing method increases the power of detecting recurring aberrations if we apply a constant threshold (theoretically justified by Equation 7 in the main text). We do this by simulating 1000 aggregated profiles (with only one recurring aberration) and calculate the proportion of these tests that reveal the recurring aberration with both smoothing methods.
Simulation results illustrating the effect that the resolution parameter $\alpha$ (on x-axis of each plot) has on FDR and power for a simulated case study. Simulations were performed for SNR $\in \{1, 3, 9, 1e+99\}$ and repeated 1000 times. In each experiment we add two broad, five medium and 50 focal recurrent events. In each plot, the vertical black line indicates $\alpha = 20$. A.I) All detected events that are covered by less than 70% ($J_i < 0.7$) of all true recurrent regions are considered a false positive. The y-axis represents the average ratio of detected events that are false positives (measured FDR). A.II) All detected events that are covered by more than 70% of a true recurrence ($J_i \geq 0.7$), but with a Jaccard similarity coefficient below 70% is considered a shattered positive. The y-axis represents the average ratio of detected events that are shattered positives. A.III) All detected events that have a Jaccard similarity coefficient below 70% for all recurrent regions are considered a strict false positive. The y-axis represents the average ratio of detected regions that are strict false positives. B.I) The y-axis represents the power for detecting focal events with a Jaccard similarity coefficient above 70%. B.II) This is the same as B.I) except that we show the power for detecting broad events. B.III) Here we show the power for detecting all events (focal, medium and broad).

(new and KC-SMART) when controlling the (two-tailed) FWER at 5%. Since we know the location of the recurring aberration, we do not need to rely on the cyclic-shift null hypotheses and estimate the required threshold from these 1000 simulations to control the FWER. Therefore the power of KC-SMART can be compared directly to the newly proposed method.

**Simulated data**

- Genome size: 240 Mbps, $a = \{g|0 \leq g \leq 120$ Mbps$\}$;
- Number of probes: 12000;
- Probe positions: Random positions on the genome (each probe has only 1 bp width);
- Number of samples to aggregate: 100;
- Segmentation: Each sample is considered to consist of 160 segments with 159 random breakpoints;
- Segment amplitudes: All probes within a segment takes on a value of either $-1$, 0 or +1 with probabilities 25%, 50% and 25% respectively. Therefore, for each sample we expect 80 aberrated segments (with amplitude $-1$ or $+1$) that are not recurring;
- Recurring segment center: $c_a = 120$ Mbps;
- Recurring segment widths: We perform parallel experiments with widths $(w_a)$ 40 kbps, 80 kbps, 160 kbps, 320 kbps, 640 kbps, 1.28 Mbps, 2.56 Mbps, 5.12 Mbps and 10.24 Mbps;
- Recurring segment amplitude: $+1$;
- Probability of recurring segment per sample: Each sample (of the 100 samples to aggregate) has a 30% chance of containing a recurring aberration;
- Signal to noise ratio (SNR): For a given SNR, Gaussian noise is added to each sample with mean zero and variance $1/$SNR. Parallel experiments are performed at SNRs of 0 (no random or recurring aberrations), 0.1, 1.0, 3.0, 9.0 and $10^{99}$ (no noise);
Experimental procedure For each parallel experiment (for data sets with a given recurring aberration width and SNR) we perform smoothing across multiple scales. We do this for kernel widths $w_k = 10$ kbps, 20 kbps, 40 kbps, 80 kbps, 160 kbps, 320 kbps, 640 kbps, 1.28 Mbps, 2.56 Mbps, 5.12 Mbps and 10.24 Mbps.

Define the recurring aggregated region to be $a_r = [c_a - w_a/2 - 3w_k, c_a + w_a/2 + 3w_k]$. Define the non-recurring aggregated region to be $a_n = a - a_r$. We perform the following steps.

• STEP 1: Generate an aggregated profile using the methodology proposed earlier;
• STEP 2: Smooth the aggregated profile using a specified kernel width (estimate null parameters from $a_n$);
• STEP 3: Calculate and store the maximum peak of the smoothed profile in the region $a_n$ and $a_r$;
• STEP 4: Repeat steps one to three 1000 times. Therefore we will have 1000 maximum peak values for both $a_n$ and $a_r$;
• STEP 5: Define $t$ to be the 97.5th highest maximum peak in $a_n$. This will be the thresholds that approximately controls the FWER (two tail) at 5%;
• STEP 6: Define $p$ (the power) to be the proportion of maximum peaks in $a_r$ above $t$;
• STEP 7: Repeat steps one to six for the two different smoothing methods (KC-SMART and the new method);

STEP 8: Set the null-region estimate equal to the whole genome $\hat{a}_n = a$;

STEP 9: Define $\hat{a}_r = \{ g \in a | h_{a}(g) \geq t \}$. Set $\hat{a}_n = a - \hat{a}_r$;

STEP 10: Repeat steps 3 to 5 until $\hat{a}_n$ converges. Now we have a final estimate for $t$;

STEP 11: Determine whether any peaks are above $t$ in the non-recurring region $a_n$ and the recurring region $a_r$;

STEP 12: Repeat steps one to seven 1000 times;

STEP 13: Define $p$ (the power) to be the proportion of simulations that reveal one or more peaks in $a_r$ above $t$;

STEP 14: Define FWER (empirical) to be the proportion of simulations that reveal one or more peaks in $a_n$ above $t$;

Results Figure S3 shows the power obtained for all combinations of SNR, aberration width and kernel width from the proposed smoothing method and the measured FWER (which was controlled analytically).

It is interesting to note that for segmented data (no Gaussian noise on any segments) the analytical FWER control becomes conservative. This is mainly due to the fact that the expected Euler characteristic is an upper bound on the FWER and nearby peaks are highly correlated (since the measurement noise is small). To see why, consider:

$$E[\chi] = P[\chi = 1] + 2P[\chi = 2] + 3P[\chi = 3] + ... \approx (\chi)$$

FWER = $P[\chi = 1] + P[\chi = 2] + P[\chi = 3] + ...$ (S41)

If the correlation between probes extend much further than the kernel width (for example, smoothing with a small kernel on segmented samples), we get situations like:

$$P[\chi = 1] \approx P[\chi = 2] \approx P[\chi = 3],$$

and therefore the expectation becomes conservative.

FWER control for simulated data

Summary This test is very similar to the one proposed in the KC-SMART comparison. The big difference is that we do not use our knowledge of the recurring aberration locations in order to control the FWER. In other words we can only find estimates on the recurrent ($a_r$) and null ($a_n$) regions. We treat the method as a black box which outputs a smoothed aggregated profile and a threshold that is analytically determined to control the FWER at 5%.

Experimental procedure

• STEP 1: Generate an aggregated profile using the methodology proposed in Section;
• STEP 2: Set the null-region estimate equal to the whole genome $\hat{a}_n = a$;
• STEP 3: Estimate null-parameters $\mu, \sigma^2$ and $r$ using only probes in $\hat{a}_n$;
• STEP 4: Smooth the aggregated profile using a specified kernel width ($w_k$) and calculate the analytical threshold $t$ that controls $E[\chi]$ at 2.5% (remember we are performing a two tailed test);
• STEP 5: Set $\hat{a}_r = \{ g \in a | h_{a}(g) \geq t \}$. Set $\hat{a}_n = a - \hat{a}_r$;
• STEP 6: Repeat steps 3 to 5 until $\hat{a}_n$ converges. Now we have a final estimate for $t$;
• STEP 7: Determine whether any peaks are above $t$ in the non-recurring region $a_n$ and the recurring region $a_r$;
• STEP 8: Repeat steps one to seven 1000 times;
• STEP 9: Define $p$ (the power) to be the proportion of simulations that reveal one or more peaks in $a_r$ above $t$;
• STEP 10: Define FWER (empirical) to be the proportion of simulations that reveal one or more peaks in $a_n$ above $t$;

References

1. 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.

2. Adler,R.J. (1976) On Generalising the Notion of Uppcrossings to Random Fields. *Adv. Appl. Prob.*, 8, 789–805.

3. 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.

4. Klijn,C., Holstege,H., de Ridder,J., Liu,X., Reinders,M., Jonkers,J. and Wessels,L.J. (2008) Identification of cancer genes using a statistical framework for multexperiment analysis of nondiscretized array CGH data. *Nucleic Acids Res.*, 36, e13.

5. Berdyshev,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.

6. 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.

7. Supplementary Data — 2013/1/30 — 18:02 — page 8 — #8
Figure S2. KC-SMART and the new method (ADMIRE) are similar in the sense that both perform kernel smoothing on an aggregated profile and then apply a constant threshold to decide whether an aberration is recurring or not. We simulated aCGH profiles with a known recurring aberration as described earlier and test the power of detecting this aberration with a) KC-SMART and b) the new method (ADMIRE) when fixing the FWER at 5%. Each plot represents the power (proportion of recurring aberration detected in 1000 simulations) while varying the aberration width ($w_a$) and kernel width ($k_w$). Inspection reveals an increase in power with ADMIRE (especially for small kernels).
Figure S3. In this experiment we again simulate 1000 aggregated profiles. Then for different recurring aberration widths, kernel widths and SNRs we calculate a) the power and b) the estimated FWER. The color bar in the FWER plots should be interpreted as follows: green represents one standard deviation from 5% FWER, cyan and yellow represent FWERs within two standard deviations, blue and red represent extremes below or above two standard deviations and black represents a FWER below 1%. 
Figure S4. Example of ADMIRE on a part of Chromosome 17 where no recurrent aberrations are found by GISTIC2.0. A.I) The SNP-array profiles for 141 Glioma samples. Red (green) represents amplifications (deletions). A.II) The aggregated (sum) of all the samples. A.III) A multi-level representation of the recurring events found by ADMIRE at 25% event-based FDR (across the whole genome). The first recursive level shows a broad deletion. If we re-estimate the null-parameters in this broad event we find two extra focal events. B) Shows that the left most focal event overlaps with the NF1 Glioma tumor suppressor gene. C) The second focal recurrent loss is located in the ASIC2 gene. High-grade glioma tumors are associated with the low functional expression of ASIC2 (5, 6).