Copy-number-variation and copy-number-alteration region detection by cumulative plots

Wentian Li*, Annette Lee and Peter K Gregersen

Address: The Robert S Boas Center for Genomics and Human Genetics, Feinstein Institute for Medical Research, North Shore LIJ Health System, Manhasset, NY 11030, USA

Email: Wentian Li* - wli@nslij-genetics.org; Annette Lee - anlee@nshs.edu; Peter K Gregersen - peterg@nshs.edu

* Corresponding author

Background

Regions with copy number variations (in germline cells) or copy number alteration (in somatic cells) are of great interest for human disease gene mapping and cancer studies. They represent a new type of mutation and are larger-scaled than the single nucleotide polymorphisms. Using genotyping microarray for copy number variation detection has become standard, and there is a need for improving analysis methods.

Results

We apply the cumulative plot to the detection of regions with copy number variation/alteration, on samples taken from a chronic lymphocytic leukemia patient. Two sets of whole-genome genotyping of 317 k single nucleotide polymorphisms, one from the normal cell and another from the cancer cell, are analyzed. We demonstrate the utility of cumulative plot in detecting a 9 Mb (9 ×10^6 bases) hemizygous deletion and 1 Mb homozygous deletion on chromosome 13. We also show the possibility to detect smaller copy number variation/alteration regions below the 100 kb range.

Conclusion

As a graphic tool, the cumulative plot is an intuitive and a scale-free (window-less) way for detecting copy number variation/alteration regions, especially when such regions are small.

Background

Most efforts in genetic mapping of human diseases focus on single-nucleotide-polymorphism (SNP): individual nucleotide base that may differ from one person to another. If the cause of a polymorphism is due to diverging paths in population genetic history, such as in multiple ethnic groups, it can be used as an ancestry or ethnic identity marker [1]. If the polymorphism is a functional mutation (non-synonymous or promoter-region polymorphism) [2] underlying a human disease, then it is the focus of attention in case-control genetic analyses [3].

A new type of genetic polymorphism emerged recently as another source of mutation that may lead to human diseases: the copy number variation (CNV) (for literature on CNV, see an online bibliography [4]). Local duplication and deletion events occurring at kb (10^3 bases) or Mb (10^6 bases) scales are the cause of CNV. If these events occurred in prior generations, CNV can be treated as a genetic
marker whose transmission might be traced in studying
the disease-status correlation. These events can also occur
in the current generation, as de novo mutations.

Similar duplication and deletion events also occur in
somatic cells, leading to copy number alteration (CNA).
Besides the link between CNA and cancers studied before
[5], an early CNV-disease association was reported on
Charcot-Marie-Tooth disease [6], in inherited neurologi-
cal disorder. In the past year or two, the number of reports
on association of CNV with human diseases increased
dramatically, especially for psychiatric disorders such as
Schizophrenia [7-9], bipolar [10], and for brain develop-
mental disorder such as Autism [11-15]. These diseases
have long been evading genetic dissection, and the CNV
link offers new optimism for our ultimate understanding
of these diseases.

The technology for CNV detection evolves from Mb-level
comparative genomic hybridization (CGH) to higher-res-
olution array-based CGH [16]. Genotyping array whose
original goal is to genotype individual SNPs, has increas-
ingly been used for CNV detection [17-20]. There are two
relevant pieces of information from a genotyping array
data for the purpose of CNV detection.

The first is the ratio of intensity reading of alleles for a
sample to that from a reference group of normal samples.
If the ratio is larger than 1, there are more copies of piece
of DNA in the sample than normal (which is 2 copies). If
the ratio is less than 1, it indicates a deletion. The second
signal is the genotype. Deletion of one of the chromo-
somes leads to a run of homozygosity for all SNPs in the
region, though run of homozygosity can also be due to
inbreeding [21,22]. The homozygosity property of one-
copy deletion is well exploited in detecting loss-of-hetero-
yzogosity in CNA of cancer cells [23].

CNV detection using genotype microarray data relies on
these two sequences: if the intensity ratio deviates from
the normal value of 1 for a chromosome region with a
consistent value, it can be a CNV region. Similarly, if a run
of homozygosity is observed in a region, it could indi-
rectly indicate a copy-number deletion. A CNV region
detection is more convincing if CNV signals exhibited by
both sequences overlap in a common region.

Methods for calling CNV regions can be roughly classified
into two types. The first type is straightforward: a CNV
detection is claimed when the log-ratio value is signifi-
cantly deviated from 0 [24]. The problem with this
method is that the threshold for calling CNV varies greatly
from platform to platform, from study to study, and a
comparative investigation is urgently needed [16]. The
second type uses hidden Markov models (HMM), where
the underlying CNV status is the hidden variable, and the
log-ratio and genotype sequences are the two observed
variables [25-30]. One advantage of the HMM framework
is that it can incorporate information from both
sequences at once.

When the parameter settings in a HMM are fixed, HMM is
a stationary (homogeneous) process along a chromo-
some. There is one parameter in HMM which controls the
transition probability from the (hidden) CNV state to
non-CNV state. That parameter can also be transformed to
the characteristic size for CNV region [28]. What if the
CNV regions do not have a characteristic size, or equiva-
ently, the length distribution is not exponential? In that
case, CNV-calling methods that do not require stationarity
are preferred.

The guanine-cytosine content (GC%) in DNA sequences
has been a focus of non-stationary, non-Markov, long-
range-correlated modeling for more than twenty years
[31-33]. It is well acknowledged that the hierarchical pat-
tern of GC%-domains within GC%-domains is possible
[34,35]. In order to detect both small and large GC-homo-
geneous domains, one applies methods that do not preset
a characteristic scale. One such method is the recursive
segmentation that adopts a divide-and-conquer approach
[36]. Another is the cumulative plot.

Cumulative plot is a graphic display of sequence informa-
tion such that trend in a region becomes more visible and
obvious. It is a window-less method because no character-
istic scale needs to be specified, although a window can be
imposed on a plot when all patterns within certain length
scale are to be ignored. In DNA sequence context, such
cumulative plots were called “DNA walk” [37,38] or “Z
curve” [39,40]. The cumulative plot has also been widely
used for detection of replication origin [41,42]. To our
knowledge, cumulative plots have not been applied to
CNV/CNA detection. The purpose of this paper is not to
provide a comprehensive comparison of various CNV/
CNA-calling methods, but limited to the presentation and
illustration of this new approach.

Results and discussion
Since our method applies equally to CNV and CNA data,
here we examine the CNA pattern in a cancer patient with
chronic lymphocytic leukemia (CLL) [43]. DNA samples
from the patient’s normal cell and that from the cancer
cell are obtained and genotyped with 317,000 SNPs
genomewide. Figure 1 shows the log-ratio and θ
sequences (see Methods) for chromosome 13, where a 9
Mb CNA region (deletion) in the cancer cell is clearly vis-
ible. A deletion region is characterized by a drop in log-
ratio value, and an absence of heterozygosity. Our goal is
to capture the same information using cumulative plots.
The left panel of Figure 2 shows the two cumulative plots corresponding to log-ratio sequence and homozygosity indicator sequence $h$, respectively. In the simplest version, at each new SNP, the curve moves up or down by an amount equal to the log-ratio value of that SNP, or by the presence of a homozygote (+1) and a heterozygote (-1). For a deletion region, the log-ratio value is consistently negative, and the first cumulative plot shows a drop; and genotype is consistently homozygous (also called run of homozygosity (ROH)), and the second cumulative plot shows a jump. However, from Figure 2 (left), even outside the CNA region, the first (second) cumulative plot continues to go down (up), reflecting a global abundancy of negative log-ratio over positive one (homozygotes over heterozygotes).

To remove the global or chromosome-wide average, we redraw a detrended cumulative plot (right panel of Figure 2) where the linear trend from the normal cell is subtracted from the two cumulative plots. If the difference of global trends between the cancer and normal cell is an artifact, e.g., the poor DNA quality in cancer cell that leads to higher missing rate for genotype calling, thus seemingly lower heterozygote frequency, then the normal and cancer cumulative plots should be detrended separately. Without such an evidence, we use the linear trend in normal cell to detrend both samples to highlight the difference between the two.

Cumulative plots can be customized to pick any specially defined signal. Suppose we are mainly interested in regions with copy number equal to 1, i.e., hemizygous deletion. Such deletion region should exhibit two features: (1) log-ratio is equal to $\log(1/2) = -0.693147$ (as versus $\log(2/2) = 0$ in the normal situation); (2) homozygosity indicator equal to 1 (as versus to a mixture of -1's and 1's). For a SNP, we then define a "one deletion" indicator variable whose value is 1 if $-2 < \text{log-ratio} < -0.34657$ (mid-point between -0.693147 and 0) and if its genotype is a homozygote and the value is -1 otherwise.

Figure 3 shows the cumulative plot for "one deletion" indicator variable, without or with detrending (by the lin-
ear trend in the normal sample). In both versions, the hemizygous deletion region can be seen clearly. Not only the cumulative plot detects the CNA region easily, but also it delineates the border of the deletion region accurately.

When deletion occurs in both chromosomes, called homozygous deletion, the copy number is equal to zero. For homozygous deletions, both A- and B-channel intensity (see Methods) is close to zero, and the log(r) is a large negative value. Because in the A- and B-channel plane (see Methods), these SNPs are near the origin, the angle $\theta$ can not be determined unambiguously. This leads to a broad distribution of $\theta$ values between 0 and 1, as can be seen from Figure 1 (top).

We define a "two deletions" indicator variable whose value is 1 if the log-ratio is < -2; and the value is -1 otherwise. Note that the genotype information is not used. Figure 4 shows the cumulative plot for the "two deletions" indicator variable for chromosome 13. One homozygous deletion region with ~1 Mb is clearly identified immediately adjacent to the 9 Mb hemizygous deletion region.

The 9 Mb deletion on chromosome 13 in our CLL sample, which was one of the known common deletions for this disease [44], represents an example of easy detection of CNA/CNV region, because the difference between the normal and cancer cell for both log-ratio and genotype sequence is already obvious from the raw data (Figure 1). The advantage of cumulative plot is perhaps its ability to detect CNA/CNV region of smaller sizes.

Figure 5 shows the example of chromosome 6 of our sample where there is no large-scaled CNA region. The log-
ratio and genotype sequence look almost identical between the normal and the cancer cell. The cumulative plot for the "one deletion" indicator variable shows that there are +400 more SNPs in the cancer cell than in the normal cell to have the one deletion signal (the "two deletion" cumulative plot is not shown because the signal is mostly absent along the chromosome). However, these SNPs are distributed throughout the chromosome, instead of forming clusters, and we still do not have strong evidence that the cancer sample has more micro deletion regions as compared to the normal sample. In order to explore the possible existence of smaller CNA regions, we pick the longest ROH region (roughly 4 Mb) and view it with cumulative plots. Figure 6 (left) shows the undetrended cumulative plot for the one-deletion indicator variable in this region. A clear hemizygous deletion region should show up as a jump in the cumulative plot. However, the tendency within this ROH region is downward instead of upward. In other words, although all genotypes in this region are homozygous, the log-ratio mostly fails the <-0.34657 criterion.

The failure in detecting hemizygous deletion at the Mb scale does not necessarily prevent its possible existence at a smaller length scale. The right panel of Figure 6 shows a 200 kb sub-region (marked in Figure 6 (left)) that contains a 36 kb region with an upward trend in the cumulative plot. A zooming into any small region in a cumulative plot enables it to detect CNA/CNV regions with ever smaller sizes.

It was previously suggested that run of homozygosity can be a sequence feature that is associated with certain human diseases [45]. We see here that ROH is only a partial indicator for a CNA/CNV region. The longest ROH on chromosome 6 in our sample only shows some weak evidence in a much narrower region for one-deletion CNA. Considering both ROH and log-ratio sequence is clearly better than considering ROH alone. Although ROH may still be biologically meaningful, as it could reflect a copy-neutral loss-of-heterozygosity event, one has to obtain extra evidence to exclude population genetics events such as inbreeding as the true cause.
The pairing of the normal and the cancer sample is not essential to our method. In Figures 3, 4, 6, the CNA regions can be identified by cancer sample (the blue curve) alone. However, the comparison with the normal sample provides supporting evidence that deletion only occurs in the cancer cell and not in the normal cell. When SNPs along a chromosome are not evenly distributed, it may not be appropriate to move one step per SNP in the cumulative plot. For example, if multiple SNPs are in strong linkage disequilibrium in a densely typed region, the indicator variable values are positively correlated, and a sequence of +1 values is partially a consequence of their correlation, not as a series of independent evidences for CNA/CNV. We can adjust for this correlation by calculating the probability ratio $\alpha$ (see Method) in favor for concordant genotypes between neighboring SNPs, as compared to the average. If $\alpha > 1$, we discount a +1 (or -1) movement by dividing the $\alpha$ value. For the chromosome 13 data, $\alpha$ is in a very narrow range of (0.9921, 1.0002). Because the probability ratio in favor of concordant homozygotes is so close to 1, the adjusted cumulative plot is indistinguishable from the original cumulative plot.

So far the delineation of an upward trend in the cumulative plot is determined by visual inspection. Segmentation programs can be developed to carry out the delineation automatically. In particular, one may move along the cumulative plot, calculate the slope from the start point to the moving position, then from the moving position to the end point. The position that maximizes the difference of the two slopes is chosen, leading to the first segmentation. This segmentation can be carried out recursively similar to the method described in [36].

Finally, for case-control analysis using CNV, one deals with two groups of samples [46]. In this situation, cumulative plot can be first applied to each individual person to identify the CNV/CNA region. Then, chromosomes can be partitioned into equal-sized windows and the frequency

---

**Figure 4**

**Cumulative plot of the homozygous deletion indicator variable.** Cumulative plot (top) and detrended cumulative plot (bottom) for the 1 Mb homozygous deletion region on chromosome 13, using the “two deletions” indicator variable.
of CNV/CNA-containing window in the case group is compared to that in the control group for a statistical test.

**Conclusion**
We have shown here that cumulative plots of an indicator variable derived from the log-ratio and SNP genotype sequence can easily identify CNV or CNA regions. We illustrate the procedure for hemizygous deletion (copy number equal to 1) and homozygous deletion (copy number equal to 0) using samples taken from a chronic lymphocytic leukemia patient. Although CNV/CNA regions at the Mb scale can also be detected by viewing the raw data, cumulative plot is able to delineate the borders with higher degree of accuracy. Another advantage of cumulative plot is perhaps in detecting smaller CNV/CNA regions, such as those in the range of 10 kb–100 kb, as it is a scale-free approach that does not require a fixing of the window size. Cumulative plot is simple enough that no special-purpose program is needed for its use except a graphic routine: for example, all results shown here are obtained by the general statistical package R [47].

**Methods**

**log-ratio and genotype data**
In a two-channel (two-color) SNP genotyping microarray, the A- and B-channel (A- and B-allele) intensity reading is recorded. These two intensities are normalized by reference intensity values which are obtained by averaging many normal samples. Each SNP can be represented by a point in the (x, y) plane where x, y are the normalized A- and B-channel intensity. The polar coordinate of the point...
$r = \sqrt{x^2 + y^2}$ and $\theta = \tan^{-1}(y/x)$ \cite{48}. Log(r) is the "log ratio" value that provides a copy-number information, and $\theta$ provides a genotype information, where $\theta = 0, 1$ correspond to two homozygotes, and $\theta = 0.5$ corresponds to the heterozygote. Note: (1) $r$ value depends on a group-averaged reference level, and this information is provided by the array-maker company. (2) Although $r$ and $\theta$ is in principle independent, there could be weak correlation between them. Our starting point are the two sequences of log(r) and discretized $\theta$ values (i.e. genotype) along a chromosome.

Cumulative plots for log-ratio and homozygosity sequence

The $r$ and $\theta$ variable is transformed by: log-ratio = log(r) and homozygosity indicator $h = 4 \times |\theta - 0.5| - 1$. For heterozygotes, $h$ is close to -1, and for two homozygotes, $h$ is close to 1. Denote the $i$-th SNP’s log-ratio and homozygosity indicator as $\log(r_i)$ and $h_i$. The (original) cumulative plots of these two sequences are:

$$
cumu.log.ratio_j = \sum_{i=1}^{j} \log(r_i)$$

$$
cumu.h_j = \sum_{i=1}^{j} (h_i)$$

A cumulative plot can be detrended such that the first and the last SNP are on the same horizontal line. The purpose of this detrending is to remove the chromosome-wide bias so that regional deviations are highlighted. In our normal and cancer cell from the same individual example, we detrend the normal sample by subtracting the linear function $a + bx_i$, where $x_i$ is the Mb position of the $i$th SNP, $N$ is the number of SNPs, and
\[
b = \frac{\text{cumu.log.ratio}_N - \text{cumu.log.ratio}_1}{x_N - x_1},
\]
\[
a = \frac{\text{cumu.log.ratio}_N - bx_N}.
\]

To highlight the difference between the cancer cell and the normal cell, we use the \(a\) and \(b\) obtained from the normal cell to detrend the cumulative plot for the cancer cell.

**Cumulative plots corrected by spacing between neighboring SNPs**

When SNPs are not distributed evenly along a chromosome, one may consider correcting the effect of inhomogeneous correlation between neighboring SNPs. We first calculate the probability of a neighboring SNP of a homozygous SNP to be also homozygous due to the correlation between them. This calculation is carried out by the Haldane’s map [49].

Haldane’s map relates the number of recombinations within a chromosomal interval \(M\) and the probability of observing a recombinant between the two end points \(R\):

\[
R = \frac{1 - \exp(-2M)}{2}.
\]

The unit of \(M\) is Morgan, which is roughly equal to 100 Mb (or 1 centi Morgan is equal to 1 Mb [50]). The probability of observing a non-recombinant is 1 - \(R\).

Denote \(p_{\text{same}}\) the probability that one homozygous SNP is followed by another homozygous SNP that is \(M\) genetic distance apart. Since Haldane formula is applicable to haplotype, or a single copy of a chromosome, for two copies of a chromosome, \(p_{\text{same}} = (1 - R)^2 \approx 1 - 2R = e^{-2M}\).

Suppose the average spacing between two neighboring SNPs is \(\bar{M}\). For a neighboring SNP pair whose spacing \(M < \bar{M}\), it is more likely for both SNPs to be homozygous than the average, by a probability ratio of

\[
\alpha = \frac{p_{\text{same}}}{\bar{M}_{\text{same}}} = e^{-2(M-\bar{M})},
\]

and the cumulative plot for the homozygosity indicator variable can be adjusted by dividing that ratio:

\[
cumu_h_j = \sum_{i=1}^{j} h_i / \alpha_i = \sum_{i=1}^{j} h_i e^{2(M_{i-1} - \bar{M})}.
\]

We assume that \(p_{\text{same}}\) is calculated in the same way as for other indicator variables, meaning CNV/CNA of a particular type is maintained at the neighboring SNP by the same probability \(e^{-2M}\), and the above formula can be used to correct other cumulative plots. Note that transition probability from one genotype in a SNP to another genotype in the neighboring SNP can also be estimated from the HapMap data.

**List of abbreviations used**

CGH: comparative genomic hybridization; CLL: chronic lymphocytic leukemia; CNA: copy number alterations; CNV: copy number variations; GC%: guanine and cytosine contents; HMM: hidden Markov models; ROH: run of homozygosity; SNP: single nucleotide polymorphism

**Competing interests**

The authors declare that they have no competing interests.

**Authors’ contributions**

W.L. designed the method, carried out the analysis, and wrote the manuscript; A.L. genotyped the samples; P.K.G. proposed the CNV study of chronic lymphocytic leukemia.

**Acknowledgements**

We thank Nick Chiorazzi for providing the CLL sample, and Pedro Ber-naola-Galván, José Oliver for discussions on segmentation methods.

This article has been published as part of BMC Bioinformatics Volume 10 Supplement 1, 2009: Proceedings of The Seventh Asia Pacific Bioinformatics Conference (APBC) 2009. The full contents of the supplement are available online at http://www.biomedcentral.com/1471-2105/10/S1/S67.

**References**

1. Rosenberg N, Li L, Ward R, Pritchard J: Informativeness of genetic markers for inference of ancestry. Am J Hum Genet 2003, 73:1402-1422.
2. Botstein D, Risch N: Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nature Genet 2003, 33:218-237.
3. Li W: Three lectures on case-control genetic association analysis. Brief Bioinform 2008, 9:1-13.
4. An online bibliography on copy number variations [http://www.nslij-genetics.org/cnv/]
5. Kallioniemi A, Kallioniemi O, Sudar D, Rutovitz D, Gray J, Waldman F, Pinkel D: Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science 1992, 258:818-821.
6. Lupski J, de Oca-Luna R, Slagdenhout P, Pentao L, Guzzetta V, Trask B, Savceno-Cardenas O, Barker D, Killian J, Garcia C, Chakravarti A, Patel P: DNA duplication associated with Charcot-Marie-Tooth disease type IA. Cell 1991, 66:219-232.
7. Sutrall S, Gaossens D, Williams N, Heyrman L, Adolfsson R, Norton N, Buckland P, Del-Favero J: Gene copy number variation in schizophrenia. Schiz Res 2003, 96:1-3.
8. Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, Nord AS, Kusenda M, Malhotra D, Bhandari A, Stray SM, Rippey CF, Rocanova P, Markov V, Larbsmi B, Fidling RL, Sikich L, Stromberg T, Merriman B, Gogtay N, Butler P, Eckstrand K, Noory L, Gohmann P, Long R, Chen Z, Davis S, Baker C, Melitzer EE, Nelson PS, Singleton AB, Lee MK, Rapoport JL, King MC, Sejz J: Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Cell 2008, 130(5875):539-543.
9. Xu B, Roos J, Levy S, van Rensburg E, Gogos J, Karayiorgou M: 2008. Nature Genet 2008, 40:880-885.
10. Wilson G, Filibote S, Chopra V, Melnyk B, Honer WG, Holt R: DNA copy-number analysis in bipolar disorder and schizophrenia
reveals aberrations in genes involved in glutamate signaling. Hum Mol Genet 2006, 15:734-749.

11. Sebat N, Lakritz R, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J, Loetta A, Pai D, Zhang R, Lee Y, Hicks J, Spence S, Lee A, Puura K, Lehtimaki T, Ledbetter D, Gregersen P, Bregman J, Sutcliffe J, Jobanputra V, Chung W, Warburton D, King M, Skuse D, Geschwind D, Gilliam T, Ye K, Wigler M: Strong association of de novo copy number mutations with autism. Science 2007, 316:445-449.

12. Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, Liu XQ, Vincent JB, Saugy TP, Samson D, Levesque M, Lévesque P, Chiasson J, Béland D, Simard V, Giguère F, Gillanders D, Glass J, Sandlund J, Johnson B, Hixson L, McDonald C, Lin J, Tomer J, Karchmer M, Belzile D, Gregersen P, Bregman J, Sutcliffe J, Jobanputra V, Chung W, Warburton D, King M, Skuse D, Geschwind D, Gilliam T, Ye K, Wigler M: Strong association of de novo copy number mutations with autism. Science 2007, 316:445-449.

12. Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, Liu XQ, Vincent JB, Saugy TP, Samson D, Levesque M, Lévesque P, Chiasson J, Béland D, Simard V, Giguère F, Gillanders D, Glass J, Sandlund J, Johnson B, Hixson L, McDonald C, Lin J, Tomer J, Karchmer M, Belzile D, Gregersen P, Bregman J, Sutcliffe J, Jobanputra V, Chung W, Warburton D, King M, Skuse D, Geschwind D, Gilliam T, Ye K, Wigler M: Strong association of de novo copy number mutations with autism. Science 2007, 316:445-449.

13. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, Saemundsdottir E, Stefansson H, Ferreira MA, Green TD, Dyer M, Jernvall J,Reviewer: Yes, please review this question.

14. Yang H, Weickhardt P, Feuer A, Wold HA, Li W, Kaneko K: High-confidence copy number variation using DNA microarrays. Proc Natl Acad Sci U S A 2006, 103:13621-13626.

15. Eichler E, Zimmerman AW: A hot spot of genetic instability in Autism. New Eng J Med 2008, 358:677-678.

16. Bernaola-Galván P, Roman-Roldán R, Oliver J: Are isochore sequences homogeneous? Protoplasma 2002, 220:173-186.

17. Hallmayer J, Lotspeich L, Sutcliffe JS, Haines JL, Folstein SE, Piven J, Gilliam TC, Herbert M, Lajonchere C, Ledbetter DH, Lese-Martin C, Miller J, Nelson S, Samango-Sprouse CA, Haines JL, Folstein SE, Piven J, Gilliam TC, Herbert M, Lajonchere C, Ledbetter DH, Lese-Martin C, Miller J, Nelson S, Samango-Sprouse CA, Haines JL, Folstein SE, Piven J, Gilliam TC, Herbert M, Lajonchere C, Ledbetter DH, Lese-Martin C, Miller J, Nelson S, Samango-Sprouse CA, Haines JL, Folstein SE, Piven J, Gilliam TC, Herbert M, Lajonchere C, Ledbetter DH, Lese-Martin C, Miller J, Nelson S, Samango-Sprouse CA, Haines JL, Folstein SE, Piven J, Gilliam TC, Herbert M, Lajonchere C, Ledbetter DH, Lese-Martin C, Miller J, Nelson S, Samango-Sprouse CA.
vival in chronic lymphocytic leukemia. New Eng J Med 2000, 343:1910-1916.

45. Lenz T, Lambert C, DeRosse P, Burdick K, Morgan T, Kane J, Kucherlapati R, Malhotra A: Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. Proc Natl Acad Sci 2007, 104:19942-19947.

46. Barnes C, Plagnol V, Fitzgerald T, Redon R, Marchini J, Clayton D, Hurles ME: A robust statistical method for case-control association testing with copy number variation. Nature Genet 2008, 40:1245-1252.

47. The R project for statistical computing [http://www.r-project.org/]

48. Peiffer D, Gunderson K: SNP-CGH technologies for genomic profiling of LOH and copy number. Clinical Lab International 2006.

49. Ott J: Analysis of Human Genetic Linkage 3rd edition. Baltimore: The Johns Hopkins University Press; 1999.

50. Ulgen A, Li W: Comparing single-nucleotide-polymorphism marker-based and microsatellite marker-based linkage analyses. BMC Genet 2005, 6(suppl 1):S13.