Received: 31 July 2015 | Revised: 9 August 2016 | Accepted: 21 September 2016

DOI: 10.1002/gepi.22019

RESEARCH ARTICLE

WILEY Genetic Epidemiology

OFFICIAL JOURNAL INTERNATIONAL GENETIC EPIDEMIOLOGY SOCIETY www.geneticepi.org

Exome copy number variation detection: Use of a pool of unrelated healthy tissue as reference sample

Stephane Wenric[1]* | Tiberio Sticca[1]* | Jean-Hubert Caberg[3] | Claire Josse[1] | Corinne Fasquelle[1] | Christian Herens[3] | Mauricette Jamar[3] | Stéphanie Max[4] | André Gothot[4] | Jo Caers[2,5] | Vincent Bours[1,3]

[1]Laboratory of Human Genetics, GIGA-Research, University of Liège, Liège, Belgium

[2]Laboratory of Haematology, GIGA-Research, University of Liège, Liège, Belgium

[3]Department of Human Genetics, University Hospital (CHU), Liège, Belgium

[4]Department of Haematology and Immunohaematology, University Hospital (CHU), Liège, Belgium

[5]Department of Clinical Haematology, University Hospital (CHU), Liège, Belgium

Correspondence
Vincent Bours, GIGA-Research, Laboratory of Human Genetics, University of Liège, Domaine Universitaire, CHU, Sart Tilman, 4000 Liège, Belgium.
Email: vbours@ulg.ac.be

*These authors contributed equally to this work.

ABSTRACT

An increasing number of bioinformatic tools designed to detect CNVs (copy number variants) in tumor samples based on paired exome data where a matched healthy tissue constitutes the reference have been published in the recent years. The idea of using a pool of unrelated healthy DNA as reference has previously been formulated but not thoroughly validated. As of today, the gold standard for CNV calling is still aCGH but there is an increasing interest in detecting CNVs by exome sequencing. We propose to design a metric allowing the comparison of two CNV profiles, independently of the technique used and assessed the validity of using a pool of unrelated healthy DNA instead of a matched healthy tissue as reference in exome-based CNV detection. We compared the CNV profiles obtained with three different approaches (aCGH, exome sequencing with a matched healthy tissue as reference, exome sequencing with a pool of eight unrelated healthy tissue as reference) on three multiple myeloma samples. We show that the usual analyses performed to compare CNV profiles (deletion/amplification ratios and CNV size distribution) lack in precision when confronted with low LRR values, as they only consider the binary status of each CNV. We show that the metric-based distance constitutes a more accurate comparison of two CNV profiles. Based on these analyses, we conclude that a reliable picture of CNV alterations in multiple myeloma samples can be obtained from whole-exome sequencing in the absence of a matched healthy sample.

KEYWORDS

NGS, WES, CNV, aCGH, normalization, multiple myeloma, control, read count

1 | INTRODUCTION

Copy number variations (CNVs) are genomic modifications responsible of phenotypic diversity but are also involved in many pathologies like cardiovascular diseases, autoimmune diseases, neurodegenerative diseases, and cancers (Beroukhim et al., 2010; Kim et al., 2013). In cancers chromosomal alterations might lead to several specific genomic profiles which can be linked to prognosis or response to treatment, for example the amplification of the ERBB2 gene in breast cancer leads to its overexpression, and to sensitivity to treatment by trastuzumab (Robert et al., 2006).

Multiple myeloma is a hematological cancer characterized by a high level of CNV, implicating plasma cells. Some of them are linked to an adverse prognosis: del(17)(p), del(1)(p), dup(1)(q), and del(13) (Fonseca et al., 2004; Walker et al., 2010). On the other hand, hyper-diploidies involving odd chromosomes are rather associated with a favorable outcome (Smadja et al., 2001). CNV assessment during treatment course of these malignancies is also essential to evaluate disease progression (Avet-Loiseau et al., 2009; Chung, Mulligan, Fonseca, & Chng, 2013).

Traditionally, CNV detection has been performed with cytogenetic techniques such as fluorescent in situ hybridization (FISH). Comparative genomic hybridization arrays (aCGH) are currently considered as the reference technology to measure genomic alterations. However, next-generation sequencing (NGS) could soon become an essential tool for cancer study as it allows the detection of punctual mutations and insertions/deletions. Moreover, whole genome sequencing (WGS) can also be used for the detection of CNVs and displays a higher resolution than aCGH, down to 40 bp (Xi et al., 2011). However in the clinical field, WGS is too expensive and WES or targeted sequencing is more commonly considered. CNV are more easily computed from WGS data, as the entire genome is theoretically sequenced at constant

coverage and one does not have to take into account the inter-probe coverage variability that arises in WES (Hwang et al., 2015; Liu et al., 2013). That being said, WES focuses on a highly function-enriched subset of the genome and it requires smaller computational resources for processing and storage of the data than WGS. For these reasons, a number of dedicated computational algorithms have been developed to accurately retrieve segmental CNV from WES data (Guo et al., 2013; Tan et al., 2014).

Several factors are responsible for biases in CNV detection: GC rich fragments, variability of the fragmentation process during library preparation, or copy number polymorphisms. Most of the bioinformatic tools set-up for CNV detection in tumor by WES consider these potential biases and try to minimize them (Xi et al., 2011). Some of the algorithms designed to detect CNV on tumor samples also require a matched paired healthy tissue sampled from the same patient, as they use the read depth ratio between tumor and healthy sample to infer the copy number at each locus. This control sample needs to be compiled from the same technological platform. However, such paired reference tissue is very seldom available, especially in large epidemiology studies, and could theoretically be replaced by the use of a pool of unrelated healthy tissues from patients of the same ethnicity (Sathirapongsasuti et al., 2011). However, no data are currently available in the literature to state if this solution would allow the acquisition of comparable CNV results.

To evaluate if the replacement of the matched paired healthy tissue with a pool of unrelated healthy tissue confers the same results, we have compared the performances of these two reference types against results obtained by aCGH, considered as the gold standard.

The whole study was conducted on a multiple myeloma (MM) cohort. Malignant cells population was enriched by positive selection, and analyzed by WES (Nextera, Illumina) and aCGH (SureSelect, Agilent).

2 | MATERIALS

2.1 | Ethical concerns

Ethics approval was obtained from the Institutional Review Board (Ethical Committee of the Faculty of Medicine of the University of Liège) in compliance with the Declaration of Helsinki. All patients signed a written informed consent form. This work consisted of a prospective study and did not lead to any change in the treatment of enrolled patients.

2.2 | Patients and sample preparation

Bone marrow samples of 10 MM patients were obtained from CHU of Liège. CD138 human MicroBeads (Miltenyi Biotec) were used to positively select plasma cells and enrich malignant cell populations. Genomic DNA (gDNA) was extracted from enriched plasma cells using AllPrep DNA extraction kit (Qiagen) following manufacturer's instructions. Normal gDNA for three of these patients was collected and extracted from buccal cells with Gentra Puregene Buccal Cell Kit (Qiagen) following manufacturer's instructions. Eight additional normal DNA were also extracted using the same methodology and separately sequenced to constitute a pool of normal DNA.

2.3 | aCGH and CNV analysis

Plasma cells of the whole MM cohort were analyzed with the SurePrint G3 Human CGH Microarray Kit 8 × 60K (Agilent Technologies) according to manufacturer's instructions, and results were interpreted using the Cytogenomics software (Agilent Technologies). The arrays were scanned with a G2565CA microarray scanner (Agilent Technologies) and the images were extracted and analyzed with CytoGenomics software v2.0 (Agilent Technologies). An ADM-2 algorithm (cut-off 6.0), followed by a filter to select regions with three or more adjacent probes and a minimum average log2 ratio of ±0.25, was used to detect copy number changes. The quality of each experiment was assessed by the measurement of the derivative log ratio spread with CytoGenomics software v2.0. Genomic positions were based on the UCSC human reference sequence (hg19) (NCBI build 37 reference sequence assembly).

2.4 | Whole exome sequencing and CNV call

Fifty nanograms of double-stranded gDNA were used to prepare libraries with a Nextera Rapid Capture Expanded Exome Kit (Illumina) according to the manufacturer's instructions. Libraries were checked for integrity using Agilent High Sensitivity DNA Kit (Agilent Technologies) after tagmentation and after the last step of library preparation. Sequencing reactions were performed on a HiSeq2000 sequencer (Illumina).

3 | METHODS AND RESULTS

3.1 | Whole exome sequencing and CNV call

The raw sequencing data were aligned on the Human reference genome (NCBI build 137 hg19) with the BWA software (Li & Durbin, 2009). The resulting alignment BAM files went through several filtering and correcting steps (local realignment, base quality score recalibration, low quality reads filtering, and PCR duplicate reads removal) performed using the Genome Analysis Toolkit (McKenna et al., 2010) and the Picard software package (http://picard.sourceforge.net/).

A slightly modified version of the coverage files generated by the CalculateHsMetrics tool of the Picard software

package (using the *PER_TARGET_COVERAGE* software option) was used as input of the ExomeCNV software (version 1.4).

For three tumor samples for which matched normal tissue was available, two CNV profiles were called using the recommended parameters of ExomeCNV: one with the matched normal sample as control and the other one with a pool of unrelated healthy samples as control.

The ExomeCNV input file representing the pool of eight unrelated healthy samples is generated thanks to a Perl script that averages the *coverage* and *average_coverage* columns of the Exome CNV input file among all unrelated healthy samples.

The Perl scripts used to convert the output files of the CalculateHsMetrics tool to input files suitable for ExomeCNV and to generate the ExomeCNV input file for the pool are available as supplementary data.

3.2 | CNV profiles comparison

Several analyses were performed to compare the CNV profiles obtained through aCGH, Exome CNV with the matched normal sample as control and Exome CNV with the pool of eight unrelated healthy samples as control. Only autosomes were considered in this study.

For the sake of brevity, for each sample S_k, let us note S_kC, S_kM, S_kP, respectively, the CNV profile obtained through arrayCGH, the CNV profile obtained through the Exome CNV software with the matched normal sample as control and the CNV profile obtained through the Exome CNV software with the pool of eight unrelated healthy samples as control.

3.3 | Deletion/Amplification ratio

The deletion/amplification ratio has been determined for each CNV profile, as to detect possible method-specific biases. Amplifications and deletions with a |LRR| (|Log-R-Ratio|) smaller than 0.29 (corresponding to alterations whose copy number is approximately between 1.6 and 2.4) were considered to be inconsistent and were filtered out for all CNV profiles. The ratio is based on the total number of deleted and amplified bases, as this gives a more reliable information than a ratio based on the count of amplifications and deletions. As shown on Figure 1, both Samples 1 and 3 show close deletion/amplification ratios and absolute values for each of the three profiles. No specific bias in favor of amplification or deletion is found in the CNV profiles obtained through Exome CNV software with the pool of eight unrelated healthy samples as control. Interestingly, for Sample 2, the absolute values for the number of deleted bases are very similar, but the number of amplified bases varies. S_2P thus shows a deletion/amplification ratio much more similar to S_2C than S_2M. This is explained by the fact that most of the missing amplifications in S_2M are present but have in fact a low LRR and are filtered out. Due to their low LRR, these amplifications are undistinguishable from false positives.

3.4 | CNV size distribution

To know if the use of a pool as reference had an impact on the size of detected CNV, we determined the CNV size distribution for each profile (see Fig. 2). Amplifications and deletions with a |LRR| smaller than 0.29 were filtered out. Although the absolute count of very small CNVs (< 1 kb) is higher in profiles obtained through the use of a pool as reference, their relative contribution remains unchanged and insignificant (see Additional File 1).

3.5 | Confusion matrix

For both exome-based CNV profiles of each sample (S_kM and S_kP), TPR (true-positive rate), FPR (false-positive rate), TNR (true-negative rate), and FNR (false-negative rate) were determined separately for amplifications and deletions, as shown in Table 1. Amplifications and deletions with a |LRR| smaller than 0.29 were filtered out. Interestingly, CNV profiles obtained through the use of a pool of eight unrelated healthy individuals yield overall slightly better results. The low TPR value for the amplifications of S_2M can again be explained by LRR values not passing the aforementioned threshold.

3.6 | CNV profile distance metric

Each of the previous analyses highlights potential biases, similarities, and/or discrepancies between CNV profiles but, due to methodological specificities, none gives a global picture of the real distance between profiles, as even the outcomes of the confusion matrix do not take into accounts the variation of copy number in amplifications and deletions (e.g., if the reference contains a segment of copy-number 3 and the tested profile contains a segment of copy-number 4 at the same locus, both profiles are considered to contain an amplification, no difference penalty is taken into account and the confusion matrix values are the same as they would be if both segments shared the exact same copy-number).

We propose a new distance metric, designed to compare CNV profiles, which takes into account the exact LRR values, thus giving a more precise insight, independently of the technique used to obtain the profiles.

Each CNV profile is represented as a combination of sequences of LRR segments of fixed size, one sequence for each chromosome.

Let j, k be the indexes of two CNV profiles.

At each segment s, the difference in terms or LRR between the two profiles is noted as

$$\left| LRR_{sj} - LRR_{sk} \right|.$$

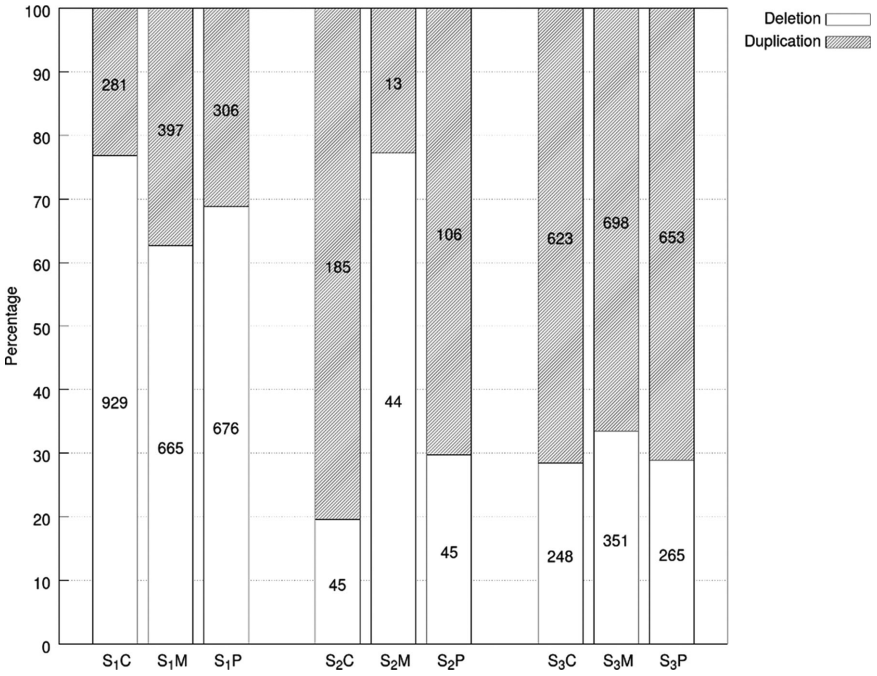

FIGURE 1 Amplification and deletion ratios for each sample. Bar heights represent the percentage of amplified or deleted bases. The total number of amplified or deleted megabases is written inside each bar

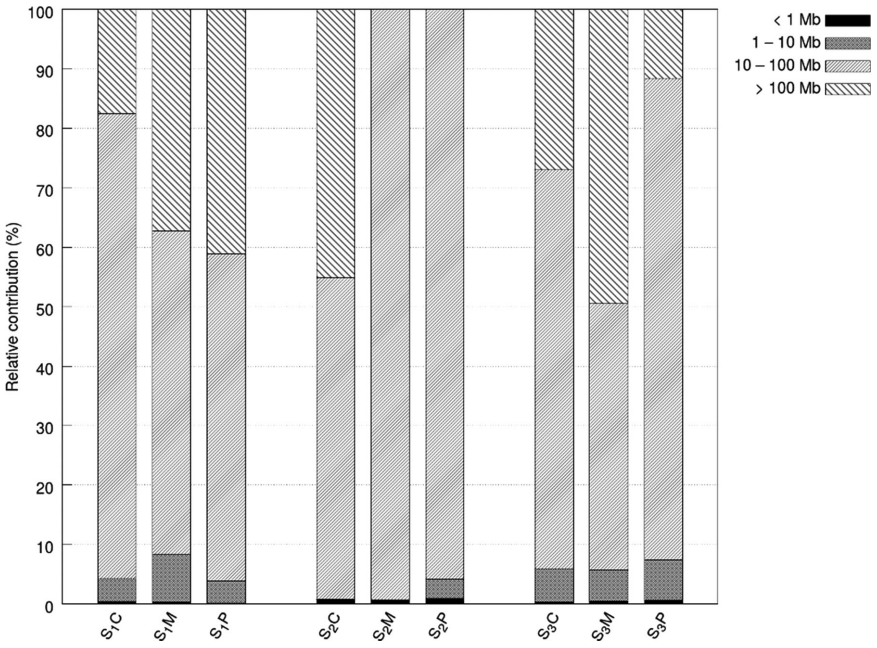

FIGURE 2 CNV size distribution. Bar heights represent the relative contribution of each group

From each LRR value, the corresponding copy-number for autosomes can be derived by

$$CN = 2 \times 2^{LRR}.$$

The difference in terms of copy number at each segment s can thus be expressed as

$$\left|CN_{sj} - CN_{sk}\right| = \left|2 \times \left(2^{LRR_{sj}} - 2^{LRR_{sk}}\right)\right|.$$

We defined the distance metric between two CNV profiles as the sum of distances between all segments divided by the total number of segments.

$$d\,(j,k) = \frac{1}{S} \times \sum_{s}^{S} \left|2 \times \left(2^{LRR_{sj}} - 2^{LRR_{sk}}\right)\right|,$$

where S is the total number of segments.

The relation between the genome size, segment size, and total number of segments is noted:

TABLE 1 Confusion Matrix

		S1_M	S1_P	S2_M	S2_P	S3_M	S3_P
AMP	TPR	97.29	96.58	12.31	99.24	99.96	97.01
	FPR	2.71	3.42	87.69	0.76	0.04	2.99
	TNR	63.83	82.14	99.98	97.79	89.23	92.63
	FNR	36.17	17.86	0.02	2.21	10.77	7.37
DEL	TPR	94.83	96.43	98.01	98.01	96.82	95.88
	FPR	5.17	3.57	1.99	1.99	3.18	4.12
	TNR	97.5	97.59	99.63	97.4	68.06	89.29
	FNR	2.5	2.41	0.37	0.6	31.94	10.71

TABLE 2 Distance Between Each CNV Profile

Sheet 1									
	S1_M	S1_P	S1_C	S2_M	S2_P	S2_C	S3_M	S3_P	S3_C
S1_M	0	1.4	1.9	3.5	3.5	3.5	4.8	4.7	4.1
S1_P	1.4	0	1.3	3.5	3.1	3.3	4.5	4.5	4.2
S1_C	1.9	1.3	0	2.6	2.5	2.3	4.6	4.4	4.2
S2_M	3.5	3.5	2.6	0	0.7	0.4	4	3.7	3.6
S2_P	3.5	3.1	2.5	0.7	0	0.5	3.8	3.6	3.7
S2_C	3.5	3.3	2.3	0.4	0.5	0	4	3.6	3.5
S3_M	4.8	4.5	4.6	4	3.8	4	0	1	2
S3_P	4.7	4.5	4.4	3.7	3.6	3.6	1	0	2
S3_C	4.1	4.2	4.2	3.6	3.7	3.5	2	2	0

$G = S \times L$, where G is the genome size and L is the segment size

A Perl script implementing this distance metric is available as supplementary data. For clarity, all distance values have been multiplied by 10.

Table 2 shows distance values computed for all possible combinations of the nine CNV profiles generated based on our cohort. Several observations can be made based on these results. For each sample, the smallest distance is always found between the two profiles obtained through the use of ExomeCNV. For each sample, the distance between the aCGH profile and the ExomeCNV profile using a pool of eight unrelated healthy individual as control is similar to the distance between the aCGH profile and the ExomeCNV profile using the matched paired healthy tissue as control. The intersample distance, whatever the technique, is always greater than the intrasample distance.

3.7 | Additional validation

The same analyses were performed on 7 MM samples for which no matched normal tissue was available. Here, only S_kP and S_kC (respectively the profile obtained through ExomeCNV with a pool, and the profile obtained through aCGH) were compared.

The proportion of deletion to amplification does not show any specific bias and the amplified and deleted bases counts are highly correlated between S_kP and S_kC (Pearson correlation coefficient of 0.975, see Additional file 2 for the count of deleted and amplified bases).

The confusion matrix values obtained when comparing S_kP to $S_kC$are relatively similar to the previous values obtained. The average values for the true-positive rate and the true negative rate for the amplifications are respectively 89.85% and 92.41%. The corresponding values for the deletions are respectively 97.04% and 74.97% (see Additional file 3 for the complete data).

The distance metric was computed for each pair of the 14 profiles. As previously, for each sample the intrasample distance is always smaller than all intersamples distances involving this sample. The average value for intrasample distance was 1.8 ± 0.2, while the average value for intersample distance was 3.25 ± 0.16. All distance values are shown in additional file 4.

4 | DISCUSSION

To date, several CNV detection tools catered to WES data exist, some of these tools make use of paired healthy DNA as references, while others use different methodologies and do not need such references. Paired methods that use the read depth or read count ratio are often more effective but inadequate for the analysis of a sample without corresponding healthy DNA.

Although ExomeCNV is a method based on read depth using paired healthy DNA as control, its authors suggested that a pool of unrelated healthy individuals could also be used as reference. Based on preliminary results, the authors also emit the hypothesis that the use of such a pool could lead to more reliable results thanks to a reduction in variance of depth-of-coverage (Sathirapongsasuti et al., 2011). No thorough analysis had previously been performed to assess the validity of these claims. Furthermore, we propose a new, better suited, way to compare CNV profiles, independently of the technique used to obtain said profile and tested this method on a small number of multiple myeloma samples.

Research and clinical application of WES for CNV detection are most useful in the cancer field. Indeed, many clinically actionable genetic changes have been described. These changes include CNV (deletions, amplifications) as well as punctual mutations. Their identification has an increasing clinical impact as they define the prognosis and can also predict treatment response or resistance, paving the way toward personalized medicine and the use of specific targeted treatments. The molecular diagnosis remains, however, difficult as it is presently based on limited amounts of DNA (from biopsies) and has to deal with the tumor heterogeneity. Moreover, large retrospective studies based on samples stored in biobanks are needed to validate genetic biomarkers in various cancers.

In our study we explored MM which is characterized by a high genomic instability. Indeed, alterations with clinical impact like monosomy 13 and trisomy of odd chromosomes are easily detected with this method while partial alterations of chromosome 1 and 17 sometimes show some approximation concerning the exact breakpoints. Even if the impact of punctual mutations in this type of cancer is still unclear, a few studies performed by NGS show a high level of mutations implicating genes frequently involved in cancers and coding for therapeutic targets (Chapman et al., 2011; Lohr et al., 2014). However, as it has been demonstrated that MM is characterized by a high level of clonal heterogeneity according to the stage of the disease, WES allows an evaluation of each clonal population proportion at the different stages of the disease (Walker et al., 2014). It could therefore be helpful for the follow-up of patients to evaluate clonal evolution in response to treatment at relapse. A simple method identifying point mutations and CNVs is certainly required for such a clinical application.

In conclusion, our data indicate that a reliable picture of CNV alterations in MM samples could be obtained from WES in the absence of a matched healthy sample. As our data were obtained on a very low number of MM samples, they will need to be confirmed on larger cohorts of other cancer types. If this can be done, it would considerably facilitate genomic studies on biobank material as well as in the clinical setting as the collection, study and data storage for matched normal DNA is expensive and generates cancer-unrelated incidental findings.

ACKNOWLEDGMENTS

We thank the patients. We also thank the GIGA-genotranscriptomic platform. Funding was obtained from the following institutions: National Fund for Scientific Research (PDR, Télévie, and FRIA); Centre-Anti-Cancéreux (ULg, Liège); Fond d'Investissement à la recherche Scientifique (CHU, Liège); Région Wallonne; and University of Liège (Fonds spéciaux pour la recherche).

All the authors declare that they have no conflict of interest.

REFERENCES

Avet-Loiseau, H., Li, C., Magrangeas, F., Gouraud, W., Charbonnel, C., Harousseau, J. L., …, Minvielle, S. (2009). Prognostic significance of copy-number alterations in multiple myeloma. *Journal of Clinical Oncology*, *27*(27), 4585–4590.

Beroukhim, R., Mermel, C. H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., …, Meyerson, M. (2010). The landscape of somatic copy-number alteration across human cancers. *Nature*, *463*(7283), 899–905.

Chapman, M. a, Lawrence, M. S., Keats, J. J., Cibulskis, K., Sougnez, C., Schinzel, A. C., …, Golub, T. R. (2011). Initial genome sequencing and analysis of multiple myeloma. *Nature*, *471*(7339), 467–472.

Chung, T. H., Mulligan, G., Fonseca, R., & Chng, W. J. (2013). A novel measure of chromosome instability can account for prognostic difference in multiple myeloma. *Plos One*, *8*(6), 1–8.

Fonseca, R., Barlogie, B., Bataille, R., Bastard, C., Bergsagel, P. L., Chesi, M., …, Avet-Loiseau, H. (2004). Genetics and cytogenetics of multiple myeloma: A workshop report. *Cancer Research*, *64*, 1546–1558.

Guo, Y., Sheng, Q., Samuels, D. C., Lehmann, B., Bauer, J. A., Pietenpol, J., & Shyr, Y. (2013). Comparative study of exome copy number variation estimation tools using array comparative genomic hybridization as control. *BioMed Research International*, *2013*, 1–7.

Hwang, M. Y., Moon, S., Heo, L., Kim, Y. J., Oh, J. H., Kim, Y.-J., …, Kim, B.-J. (2015). Combinatorial approach to estimate copy number genotype using whole-exome sequencing data. *Genomics*, *105*(3), 145–149.

Kim, T. M., Xi, R., Luquette, L. J., Park, R. W., Johnson, M. D., & Park, P. J. (2013). Functional genomic analysis of chromosomal aberrations in a compendium of 8000 cancer genomes. *Genome Research*, *23*(2), 217–227.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, *25*(14), 1754–1760.

Liu, B., Morrison, C. D., Johnson, C. S., Trump, D. L., Qin, M., Conroy, J. C., …, Liu, S. (2013). Computational methods for detecting copy number variations in cancer genome using next generation sequencing: Principles and challenges. *Oncotarget*, *44*(11), 1868–1881.

Lohr, J., Stojanov, P., Carter, S., Cruz-Gordillo, P., Lawrence, M., Auclair, D., …, Golub, T. R. (2014). Widespread genetic heterogeneity in multiple myeloma: Implications for targeted therapy. *Cancer Cell*, *25*, 91–101.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., …, DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303.

Robert, N., Leyland-Jones, B., Asmar, L., Belt, R., Llegbodu, D., Loesch, D., …, Slamon, D. (2006). Randomized phase III study of trastuzumab, paclitaxel, and carboplatin compared with trastuzumab and paclitaxel in women with HER-2 - Overexpressing metastatic breast cancer. *Journal of Clinical Oncology*, *24*(18), 2786–2792.

Sathirapongsasuti, J. F., Lee, H., Horst, B. A. J., Brunner, G., Cochran, A. J., Binder, S., …, Nelson, S. F. (2011). Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. *Bioinformatics*, *27*(19), 2648–2654.

Smadja, N. V. (2001). Hypodiploidy is a major prognostic factor in multiple myeloma. *Blood*, *98*(7), 2229–2238.

Tan, R., Wang, Y., Kleinstein, S. E., Liu, Y., Zhu, X., Guo, H., …, Zhu, M. (2014). An evaluation of copy number variation detection tools from whole-exome sequencing data. *Human Mutation*, *35*(7), 899–907.

Walker, B. A., Leone, P. E., Chiecchio, L., Dickens, N. J., Jenner, M. W., Boyd, K. D., …, Morgan, G. J. (2010). A compendium of myeloma-associated chromosomal copy number abnormalities and their prognostic value. *Blood*, *116*, 56–65.

Walker, B. A., Wardell, C., Melchor, L., Brioli, A., Johnson, D., Kaiser, M. F., …, Morgan, G. J. (2014). Intraclonal heterogeneity is a critical early event in the development of myeloma and precedes the development of clinical symptoms. *Leukemia*, *28*, 384–390.

Xi, R., Hadjipanayis, A. G., Luquette, L. J., Kim, T.-M., Lee, E., Zhang, J., …, Tibshirani, R. (2011). Copy number variation detection in whole-genome sequencing data using the Bayesian information criterion. *PNAS*, *108*, 1128–1136.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.