Data in brief 25 (2019) 104203

Contents lists available at ScienceDirect

Data in brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Genome-wide copy number variant data for inflammatory bowel disease in a caucasian population

Svetlana Frenkel [a], Charles N. Bernstein [b], Yong Won Jin [a], Michael Sargent [b], Qin Kuang [a], Wenxin Jiang [a, c], John Wei [d], Bhooma Thiruvahindrapuram [d], Stephen W. Scherer [d, e], Pingzhao Hu [a, c, f, *]

[a] Department of Biochemistry and Medical Genetics, The George and Fay Yee Centre for Healthcare Innovation, University of Manitoba, Winnipeg, MB, Canada
[b] Department of Internal Medicine, The University of Manitoba IBD Clinical and Research Centre, University of Manitoba, Winnipeg, MB, Canada
[c] Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[d] The Centre for Applied Genomics, Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada
[e] Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada
[f] Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB, Canada

ARTICLE INFO

Article history:
Received 9 May 2019
Received in revised form 20 June 2019
Accepted 24 June 2019
Available online 2 July 2019

ABSTRACT

Genome-wide copy-number association studies offer new opportunities to identify the mechanisms underlying complex diseases, including chronic inflammatory, psychiatric disorders and others. We have used genotyping microarrays to analyse the copy-number variants (CNVs) from 243 Caucasian individuals with Inflammatory Bowel Disease (IBD). The CNV data was obtained by using multiple quality control measures and merging the results of three different CNV detection algorithms: PennCNV, iPattern, and QuantiSNP. The final dataset contains 4,402 CNVs detected by two or three algorithms independently with high confidence. This paper provides a detailed description of the data generation and quality control

DOI of original article: https://doi.org/10.1016/j.ygeno.2019.05.001.
* Corresponding author. Department of Biochemistry and Medical Genetics, Room 308 - Basic Medical Sciences Building, 745 Bannatyne Avenue, University of Manitoba, Winnipeg, Manitoba, R3E 0J9, Canada.
E-mail addresses: sveta@frenkel-online.com (S. Frenkel), Charles.Bernstein@umanitoba.ca (C.N. Bernstein), jiny2@myumanitoba.ca (Y.W. Jin), Michael.Sargent@umanitoba.ca (M. Sargent), Qin.Kuang@umanitoba.ca (Q. Kuang), wx.jiang@mail.utoronto.ca (W. Jiang), wei@sickkids.ca (J. Wei), bthiruv@sickkids.ca (B. Thiruvahindrapuram), stephen.scherer@sickkids.ca (S.W. Scherer), pingzhao.hu@umanitoba.ca (P. Hu).

https://doi.org/10.1016/j.dib.2019.104203

steps. For further interpretation of the data presented in this article, please see the research article entitled '*Copy number variation-based gene set analysis reveals cytokine signalling pathways associated with psychiatric comorbidity in patients with inflammatory bowel disease*'.

Specifications table

Subject area	*Genetics*
More specific subject area	*Copy number variant analysis*
Type of data	*Tables*
How data was acquired	*Copy number variants (CNVs) were detected from the DNA microarray-based genotypes using three independent computational algorithms*
Data format	*Filtered and analyzed*
Experimental factors	*The blood samples were collected from 243 Caucasian individuals with Inflammatory Bowel Disease (IBD) enrolled to The Manitoba IBD Cohort Study*
Experimental features	*The blood samples were genotyped using Illumina Omni2.5M-8 microarray. CNVs were detected using three independent computational algorithms: PennCNV, iPattern, and QuantiSNP*
Data source location	*Manitoba, Canada*
Data accessibility	*The list of filtered stringent CNVs was deposited to dbVar database at NCBI as the following:* https://www.ncbi.nlm.nih.gov/dbvar/studies/nstd157/
Related research article	*S. Frenkel, C.N. Bernstein, M. Sargent, W. Jiang, Q. Kuang, W. Xu, P. Hu, Copy number variation-based gene set analysis reveals cytokine signalling pathways associated with psychiatric comorbidity in patients with inflammatory bowel disease, Genomics. (2019).* https://doi.org/10.1016/j.ygeno.2019.05.001.

Value of the data

- The IBD CNV data set provides a valuable resource for identifying potential causal genes for IBDs and its drug targets.
- It can be used as a baseline to compare and analyze the CNVs identified in other populations.
- These data will be useful to researchers to investigate the contribution of CNVs to IBD and its subtypes

1. Data

The presented report is a description of the CNVs identified in 243 IBD patients with Caucasian ethnicity enrolled in the Manitoba IBD Cohort Study [1]. We genotyped 269 individuals with IBD using the Illumina Omni2.5M – 8 microarray. After sample quality control and population stratification analysis, we initially selected 246 IBD patients of Caucasian ethnicity. Three different CNV detection algorithms were applied to analyze the data: PennCNV [2], iPattern [3], and QuantiSNP [4]. The detected CNVs were filtered under stringent quality control criteria for their size, probe content, and algorithm-specific quality score. The quality control workflow is presented in Fig. 1. The quality control criteria and corresponding number of disqualified samples are presented in Table 1. To obtain high-confidence calls, we removed the CNVs detected by only one of the three algorithms while the CNVs detected by two or three algorithms were merged by retaining the outer boundary [5]. Numbers of CNVs detected by different algorithms are presented in Fig. 2. The examples of merging of the results obtained by three algorithms are presented on Fig. 3. Three IBD samples with extremely large number of detected CNVs were removed, which left 243 IBD samples for the further analysis. Of the remaining data, CNVs with significant overlap with the repeat rich regions, such as centromeres and telomeres, segmental duplications, and immunoglobulin regions, were excluded [2,6]. Table 2 contains numbers

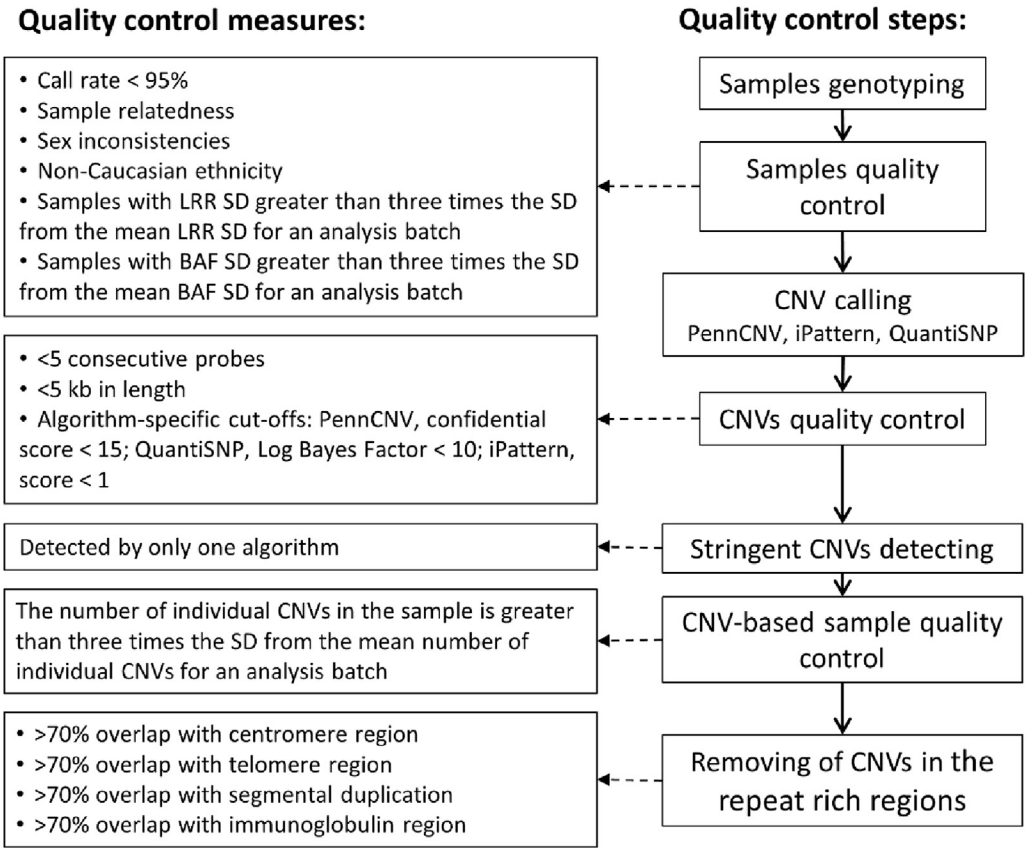

Fig. 1. Quality control and CNV detecting workflow. Stringent CNV calling was conducted using variants detected by two or three calling algorithms, which were of sizes greater than 5 kb and spanned at least five array probes. SD: Standard deviation; LRR: Log R ratio; BAF: B allele frequency.

of CNVs detected by each algorithm, corresponding numbers of disqualified CNVs, CNVs qualified for merging, and removed due to overlapping with the repeat-rich regions.

After the quality control and filtering, 4,402 stringent CNVs remained for the analysis; of those, 2,872 were deletions and 1,530 duplications. The chromosomal distribution of the stringent CNVs is presented on Table 3; the same data is visually presented on Fig. 4.

2. Experimental design, materials and methods

2.1. Study population

Individuals were enrolled in The Manitoba IBD Cohort Study – a population-based longitudinal study of patients with IBD [7,8]. At enrolment in the Cohort Study, participants were at least 18 years of age with a median disease duration of 4.3 years and maximal disease duration of 7 years. Participants were identified and recruited from a population-based registry, the University of Manitoba IBD Research Registry. The diagnosis of IBD was determined based on surgical, endoscopic, and histologic data. At the time of the cohort study recruitment, there were 3192 participants in the research registry. The Manitoba IBD Cohort

Table 1
QC criteria and their cutoffs for sample qualification and the corresponding number of disqualified samples.

QC criterion	QC cutoff for qualification	Number of disqualified samples
Call rate	>95%	3
SD for BAF	Between −0.009 and 0.074	
SD for LRR	Between 0.042 and 0.131	2
Population outliers	European ancestry	18
Number of CNVs	<145	3

Study was approved by the University of Manitoba Health Research Ethics Board, and participants provided written informed consent. Blood samples for genotyping were obtained from a total of 269 IBD patients.

2.2. Genotyping

Blood samples acquired from the 269 IBD patients in the cohort were genotyped using Illumina Infinium Omni2.5-8 microarray at The Centre for Applied Genomics (TCAG) in Toronto. Rigorous quality control (QC) procedures were performed on the resulting data. The Illumina Infinium Omni2.5-8 microarray contains a total of 2,372,784 markers for SNP and CNV analyses. Samples were processed using the manufacturer's recommended protocol; BeadChips were scanned on the Illumina BeadArray Reader using default settings. Analysis and intra-chip normalization were performed using Illumina's GenomeStudio software v.2011.1. Probes reclustering was conducted in the GenomeStudio using the project-specific samples to produce custom cluster file, which was applied to generate LogR ratios (LRR) and B Allele Frequencies (BAF) for the CNV detection.

2.3. Intensity quality control for CNV detection

Quality control (QC) for SNPs was performed at the individual SNP level (Table 1). Samples were excluded from the analysis if they had: i) array call rate <95%; ii) standard deviation (SD) for LRR and BAF values outside mean ± 3SD for SD of an analysis batch. Closely related samples (by identity-by-descent distance for each pair of individuals), duplicates, samples with gender mismatches (by X chromosome homozygosity rate) or Mendelian error rate >1% were excluded from the analysis batch.

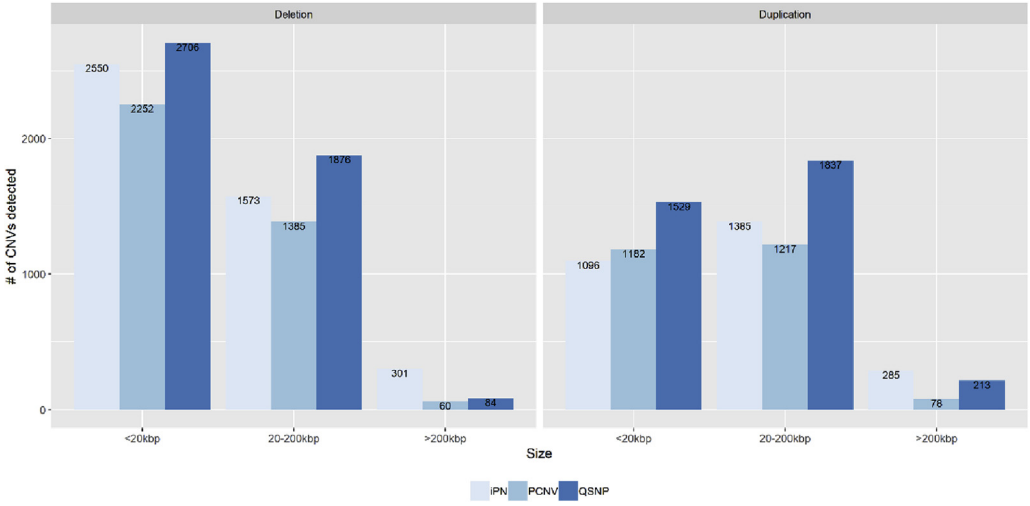

Fig. 2. **Number of CNVs detected by different algorithms for each size group.** Labels inside the bars indicate the exact number of CNVs detected for each category after sample and CNV quality control but before CNV merging. In the legends, iPN = iPattern; PCNV = PennCNV; and QSNP = QuantiSNP.

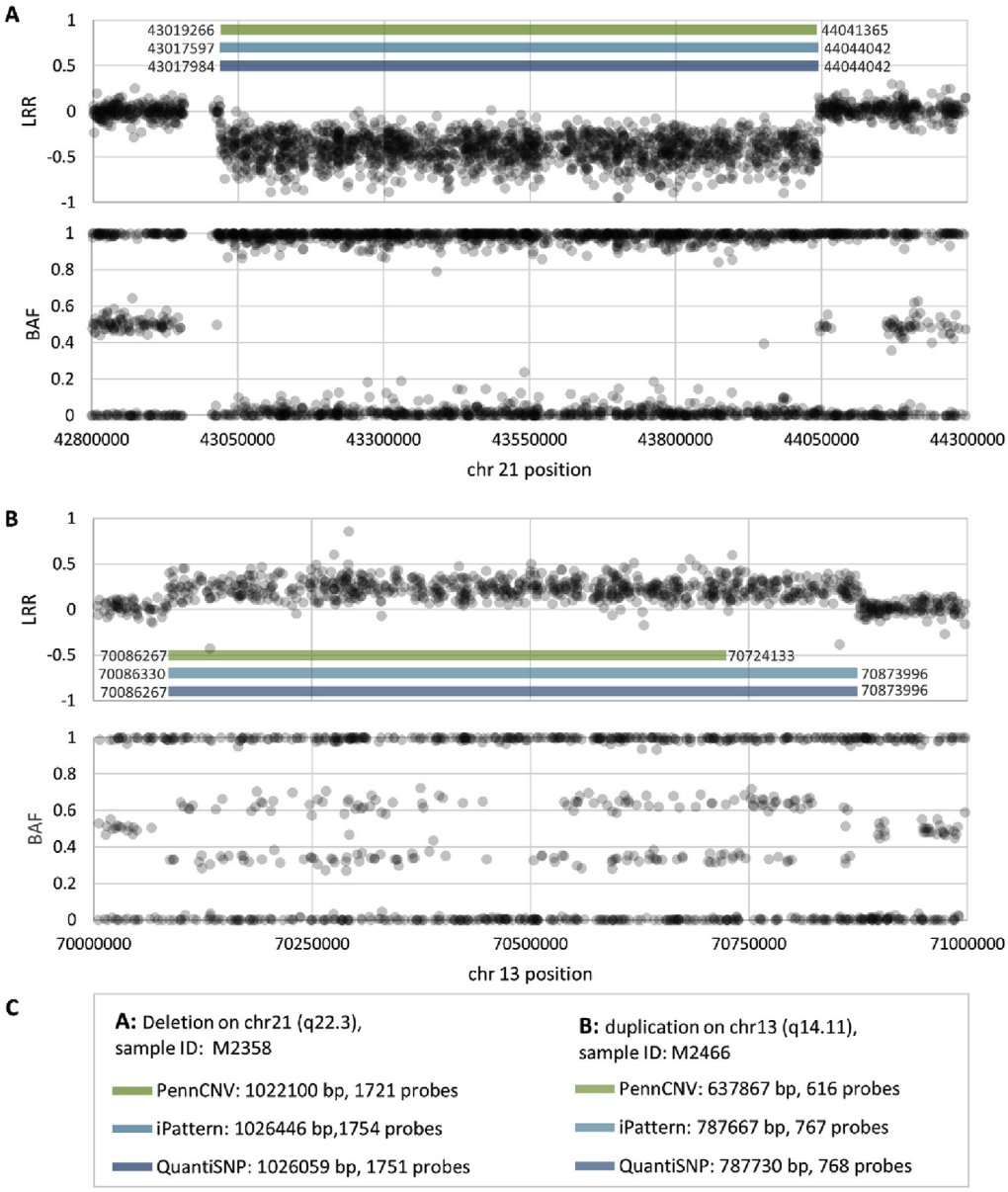

Fig. 3. Examples of CNV deletion and duplication detected by all three algorithms. CNV length and number of probes for each algorithm were provided. Start and end probe positions for each algorithm were presented on the corresponding bars. Outer start and end positions were used as start and end positions of the stringent CNV. **A**: An example of deletion found in chr21q22.3; **B**: An example of duplication found in chr13q14.11; **C**: the start and end positions of the CNV detected by each of the algorithms and corresponding number of probes. LRR: Log R ratio; BAF: B allele frequency.

2.4. Population stratification analysis

The reference population for population stratification analysis using multidimensional scaling (MDS) was obtained from Phase 3 data of 1000 Genomes Project [9]. Population stratification was

Table 2
Overview of the number of CNVs after each step of quality control. Numbers of CNVs (deletions and duplications separately, and together) detected by each of the three algorithms, number of disqualified and qualified CNVs, number of CNVs detected by each combination of algorithms, and number of CNVs removed due to overlapping with the regions of chromosomal instability.

	Duplications	Deletions	All CNVs
Initially detected			
PennCNV	3776	6763	10539
iPattern	3590	10199	13789
QuantiSNP	11194	18413	29607
Disqualified CNVs in QC			
PennCNV	1299	3066	4365
iPattern	824	5775	6599
QuantiSNP	7615	13747	21362
Qualified CNVs for merging			
PennCNV	2477	3697	6174
iPattern	2766	4424	7190
QuantiSNP	3579	4666	8245
Detected CNVs by two or three algorithms			
PennCNV, iPattern, QuantiSNP	1316	2682	3998
PennCNV, QuantiSNP	507	300	807
iPattern, QuantiSNP	330	603	933
PennCNV, iPattern	20	68	88
All stringent	2173	3653	5826
Removed CNVs due to overlapping with regions of chromosomal instability			
>70% overlap with segmental duplication	440	601	1041
>70% overlap with centromere, telomere or immunoglobulin region	225	228	453
Number of removed CNVs	643	781	1424
Final number of CNVs	1530	2872	4402

Table 3
Chromosomal distribution of the stringent CNVs. Number of deletions and duplications in three size categories, summary numbers of deletions and duplications and the total number of CNVs were presented for each chromosome excluding sex chromosomes. del: deletion and dupl: duplication.

Chromosome	<20 kbp, del/dupl	20–200 kbp, del/dupl	>200 kbp, del/dupl	all sizes, del/dupl	all CNV
1	79/36	38/52	4/5	121/93	214
2	126/88	50/29	20/4	196/121	317
3	141/47	79/83	3/4	223/134	357
4	110/54	53/40	1/5	164/99	263
5	100/36	56/39	3/2	159/77	236
6	84/18	98/42	2/8	184/68	252
7	138/44	47/28	17/6	202/78	280
8	137/34	45/23	2/4	184/61	245
9	111/44	30/28	2/2	143/74	217
10	72/24	29/28	3/12	104/64	168
11	125/62	15/13	2/4	142/79	221
12	149/12	20/61	0/1	169/74	243
13	54/12	37/10	0/2	91/24	115
14	53/10	105/12	2/23	160/45	205
15	122/2	10/12	0/3	132/17	149
16	63/50	41/45	9/3	113/98	211
17	53/20	16/84	2/4	71/108	179
18	80/17	12/4	0/0	92/21	113
19	18/28	59/54	3/2	80/84	164
20	57/19	36/8	1/1	94/28	122
21	6/16	19/29	1/0	26/45	71
22	11/4	8/26	3/8	22/38	60
All autosomes	1889/677	903/750	80/103	2872/1530	4402

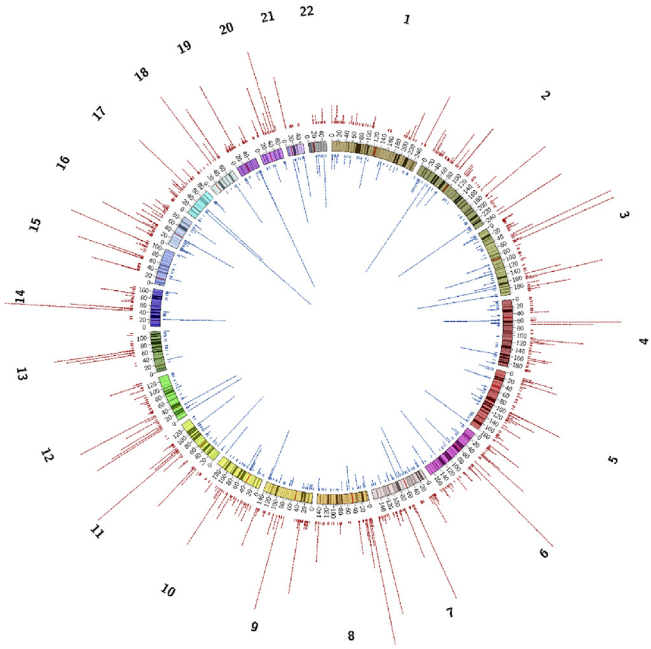

Fig. 4. Chromosome view of the detected CNVs. The CNVs were presented as tiles on the corresponding genomic positions. The colour of the tile indicated the CNV type: deletions are red, duplications are blue. The height of tile stack in each genomic region corresponds to the number of CNVs; if a genomic region contains more than 30 CNVs, only 30 tiles were presented. The figure was built using the Circos [12] tool.

conducted using PLINK [10] version 1.07. All ethnicity outliers were removed so that only samples of European ancestry were used in this study.

2.5. CNV calling algorithms

For comprehensive detection of CNVs in the IBD patients, we ran three CNV calling algorithms, namely, PennCNV [2], iPattern [3], and QuantiSNP [4]. The required data for CNV analysis, i.e. within-sample normalized fluorescence (i.e. X and Y normalized values), between-sample normalized fluorescence (i.e. Log R ratios (LRR) and B allele frequency (BAF) values) and genotypes for each sample, were exported directly from Illumina's Genomestudio software. Only autosomal probes were used in the CNV analysis. In summary, 10539, 13789 and 29607 CNVs were detected by PennCNV, iPattern and QuantiSNP, correspondingly (Table 2). We excluded the CNVs if they failed the following quality control criteria: <5 probes, <5000 bp in length and low algorithm-specific confidence score (PennCNV confidential score < 15, QuantiSNP Log Bayes Factor < 10 or iPattern score < 1). After this filtering, 6174, 7190 and 8245 CNVs were identified as high quality CNVs calls for PennCNV, iPattern and QuantiSNP, respectively. Each algorithm performed differently in calling CNVs of different sizes, with PennCNV being the most conservative in detection of CNVs, while QuantiSNP was least conservative except for large (>200 kbp) CNVs (Fig. 2).

2.6. CNV merging

To obtain stringent CNV calls, we merged high quality CNVs detected by at least two of the three algorithms using outer probe boundaries (Fig. 3). All CNVs detected by only one algorithm were excluded from the further analysis. As an additional step of sample QC, we excluded three samples with excessive number of stringent CNVs. We removed the samples with more than 145 CNVs (as the mean

number of CNVs plus 3 SD). After CNV merging, 5826 CNVs were considered as stringent. 2173 and 3653 of the CNVs are duplications and deletions, respectively (see Table 2).

2.7. CNV filtering

We further excluded CNVs that: 1) overlapped the centromere (100 kbp regions before and after centromeres) or the telomere (100 kb from the ends of the chromosome); 2) had > 70% of its length overlapping a segmental duplication using the entire segmental duplication dataset downloaded from the University of California, Santa Cruz (UCSC) Genome Browser website [11]; 3) had >70% overlap with immunoglobulin region (susceptible to somatic changes) [2,6].

The final CNV data set includes a total of 4402 CNVs, 2872 and 1530 of which were deletions and duplications, respectively (Fig. 4). The final list of stringent CNVs is available in the dbVar database at NCBI: ***nstd157***. Of all CNVs, 58.3% were smaller than 20 kbp, while 4.2% covered more than 200 kbp. Chromosomal distribution of the stringent deletions and duplications in three size categories (less than 20 kbp, 20–200 kbp and more than 200 kbp) were presented in the Table 3.

Acknowledgements

The authors wish to acknowledge Drs. Susan Walker and Mehdi Zarrei for their useful comments during the manuscript preparation. This work was supported in part by Health Sciences Centre Foundation, Mitacs, Manitoba Health Research Council and the University of Manitoba.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S. Frenkel, C.N. Bernstein, M. Sargent, W. Jiang, Q. Kuang, W. Xu, P. Hu, Copy number variation-based gene set analysis reveals cytokine signalling pathways associated with psychiatric comorbidity in patients with inflammatory bowel disease, Genomics (2019), https://doi.org/10.1016/j.ygeno.2019.05.001.

[2] K. Wang, M. Li, D. Hadley, R. Liu, J. Glessner, S.F.A. Grant, H. Hakonarson, M. Bucan, PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data, Genome Res. 17 (2007) 1665–1674, https://doi.org/10.1101/gr.6861907.

[3] D. Pinto, K. Darvishi, X. Shi, D. Rajan, D. Rigler, T. Fitzgerald, A.C. Lionel, B. Thiruvahindrapuram, J.R. Macdonald, R. Mills, A. Prasad, K. Noonan, S. Gribble, E. Prigmore, P.K. Donahoe, R.S. Smith, J.H. Park, M.E. Hurles, N.P. Carter, C. Lee, S.W. Scherer, L. Feuk, Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants, Nat. Biotechnol. 29 (2011) 512–520, https://doi.org/10.1038/nbt.1852.

[4] S. Colella, C. Yau, J.M. Taylor, G. Mirza, H. Butler, P. Clouston, A.S. Bassett, A. Seller, C.C. Holmes, J. Ragoussis, QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data, Nucleic Acids Res. 35 (2007) 2013–2025, https://doi.org/10.1093/nar/gkm076.

[5] D. Pinto, A.T. Pagnamenta, L. Klei, R. Anney, D. Merico, R. Regan, J. Conroy, T.R. Magalhaes, C. Correia, B.S. Abrahams, J. Almeida, E. Bacchelli, G.D. Bader, A.J. Bailey, G. Baird, A. Battaglia, T. Berney, N. Bolshakova, S. Bölte, P.F. Bolton, T. Bourgeron, S. Brennan, J. Brian, S.E. Bryson, A.R. Carson, G. Casallo, J. Casey, B.H.Y. Chung, L. Cochrane, C. Corsello, E.L. Crawford, A. Crossett, C. Cytrynbaum, G. Dawson, M. de Jonge, R. Delorme, I. Drmic, E. Duketis, F. Duque, A. Estes, P. Farrar, B.A. Fernandez, S.E. Folstein, E. Fombonne, C.M. Freitag, J. Gilbert, C. Gillberg, J.T. Glessner, J. Goldberg, A. Green, J. Green, S.J. Guter, H. Hakonarson, E.A. Heron, M. Hill, R. Holt, J.L. Howe, G. Hughes, V. Hus, R. Igliozzi, C. Kim, S.M. Klauck, A. Kolevzon, O. Korvatska, V. Kustanovich, C.M. Lajonchere, J.A. Lamb, M. Laskawiec, M. Leboyer, A. Le Couteur, B.L. Leventhal, A.C. Lionel, X.-Q. Liu, C. Lord, L. Lotspeich, S.C. Lund, E. Maestrini, W. Mahoney, C. Mantoulan, C.R. Marshall, H. McConachie, C.J. McDougle, J. McGrath, W.M. McMahon, A. Merikangas, O. Migita, N.J. Minshew, G.K. Mirza, J. Munson, S.F. Nelson, C. Noakes, A. Noor, G. Nygren, G. Oliveira, K. Papanikolaou, J.R. Parr, B. Parrini, T. Paton, A. Pickles, M. Pilorge, J. Piven, C.P. Ponting, D.J. Posey, A. Poustka, F. Poustka, A. Prasad, J. Ragoussis, K. Renshaw, J. Rickaby, W. Roberts, K. Roeder, B. Roge, M. L. Rutter, L.J. Bierut, J.P. Rice, J. Salt, K. Sansom, D. Sato, R. Segurado, A.F. Sequeira, L. Senman, N. Shah, V.C. Sheffield, L. Soorya, I. Sousa, O. Stein, N. Sykes, V. Stoppioni, C. Strawbridge, R. Tancredi, K. Tansey, B. Thiruvahindrapduram, A.P. Thompson, S. Thomson, A. Tryfon, J. Tsiantis, H. Van Engeland, J.B. Vincent, F. Volkmar, S. Wallace, K. Wang, Z. Wang, T.H. Wassink, C. Webber, R. Weksberg, K. Wing, K. Wittemeyer, S. Wood, J. Wu, B.L. Yaspan, D. Zurawiecki, L. Zwaigenbaum, J.D. Buxbaum, R.M. Cantor, E.H. Cook, H. Coon, M.L. Cuccaro, B. Devlin, S. Ennis, L. Gallagher, D.H. Geschwind, M. Gill, J.L. Haines, J. Hallmayer, J. Miller, A.P. Monaco, J.I. Nurnberger Jr., A.D. Paterson, M.A. Pericak-Vance, G.D. Schellenberg, P. Szatmari, A.M. Vicente, V.J. Vieland, E.M. Wijsman, S.W. Scherer, J.S. Sutcliffe, C. Betancur, Functional impact of global rare copy number variation in autism spectrum disorders, Nature 466 (2010) 368–372, https://doi.org/10.1038/nature09146.

[6] M. Zarrei, J.R. MacDonald, D. Merico, S.W. Scherer, A copy number variation map of the human genome, Nat. Rev. Genet. 16 (2015) 172–183, https://doi.org/10.1038/nrg3871.
[7] C.N. Bernstein, J.F. Blanchard, P. Rawsthorne, A. Wajda, Epidemiology of Crohn's disease and ulcerative colitis in a central Canadian province: a population-based study, Am. J. Epidemiol. 149 (1999) 916–924.
[8] L.A. Graff, J.R. Walker, L. Lix, I. Clara, P. Rawsthorne, L. Rogala, N. Miller, L. Jakul, C. McPhail, J. Ediger, C.N. Bernstein, The relationship of inflammatory bowel disease type and activity to psychological functioning and quality of life, Clin. Gastroenterol. Hepatol. 4 (2006) 1491–1501, https://doi.org/10.1016/j.cgh.2006.09.027.
[9] A. Auton, G.R. Abecasis, D.M. Altshuler, R.M. Durbin, G.R. Abecasis, D.R. Bentley, A. Chakravarti, A.G. Clark, P. Donnelly, E.E. Eichler, P. Flicek, S.B. Gabriel, R.A. Gibbs, E.D. Green, M.E. Hurles, B.M. Knoppers, J.O. Korbel, E.S. Lander, C. Lee, H. Lehrach, E.R. Mardis, G.T. Marth, G.A. McVean, D.A. Nickerson, J.P. Schmidt, S.T. Sherry, J. Wang, R.K. Wilson, R.A. Gibbs, E. Boerwinkle, H. Doddapaneni, Y. Han, V. Korchina, C. Kovar, S. Lee, D. Muzny, J.G. Reid, Y. Zhu, J. Wang, Y. Chang, Q. Feng, X. Fang, X. Guo, M. Jian, H. Jiang, X. Jin, T. Lan, G. Li, J. Li, Y. Li, S. Liu, X. Liu, Y. Lu, X. Ma, M. Tang, B. Wang, G. Wang, H. Wu, R. Wu, X. Xu, Y. Yin, D. Zhang, W. Zhang, J. Zhao, M. Zhao, X. Zheng, E.S. Lander, D.M. Altshuler, S.B. Gabriel, N. Gupta, N. Gharani, L.H. Toji, N.P. Gerry, A.M. Resch, P. Flicek, J. Barker, L. Clarke, L. Gil, S.E. Hunt, G. Kelman, E. Kulesha, R. Leinonen, W.M. McLaren, R. Radhakrishnan, A. Roa, D. Smirnov, R.E. Smith, I. Streeter, A. Thormann, I. Toneva, B. Vaughan, X. Zheng-Bradley, D.R. Bentley, R. Grocock, S. Humphray, T. James, Z. Kingsbury, H. Lehrach, R. Sudbrak, M.W. Albrecht, V.S. Amstislavskiy, T.A. Borodina, M. Lienhard, F. Mertes, M. Sultan, B. Timmermann, M.-L. Yaspo, E.R. Mardis, R.K. Wilson, L. Fulton, R. Fulton, S.T. Sherry, V. Ananiev, Z. Belaia, D. Beloslyudtsev, N. Bouk, C. Chen, D. Church, R. Cohen, C. Cook, J. Garner, T. Hefferon, M. Kimelman, C. Liu, J. Lopez, P. Meric, C. O'Sullivan, Y. Ostapchuk, L. Phan, S. Ponomarov, V. Schneider, E. Shekhtman, K. Sirotkin, D. Slotta, H. Zhang, G.A. McVean, R.M. Durbin, S. Balasubramaniam, J. Burton, P. Danecek, T.M. Keane, A. Kolb-Kokocinski, S. McCarthy, J. Stalker, M. Quail, J.P. Schmidt, C.J. Davies, J. Gollub, T. Webster, B. Wong, Y. Zhan, A. Auton, C.L. Campbell, Y. Kong, A. Marcketta, R.A. Gibbs, F. Yu, L. Antunes, M. Bainbridge, D. Muzny, A. Sabo, Z. Huang, J. Wang, L.J.M. Coin, L. Fang, X. Guo, X. Jin, G. Li, Q. Li, Y. Li, Z. Li, H. Lin, B. Liu, R. Luo, H. Shao, Y. Xie, C. Ye, C. Yu, F. Zhang, H. Zheng, H. Zhu, C. Alkan, E. Dal, F. Kahveci, G.T. Marth, E.P. Garrison, D. Kural, W.-P. Lee, W. Fung Leong, M. Stromberg, A.N. Ward, J. Wu, M. Zhang, M.J. Daly, M.A. DePristo, R.E. Handsaker, D.M. Altshuler, E. Banks, G. Bhatia, G. del Angel, S.B. Gabriel, G. Genovese, N. Gupta, H. Li, S. Kashin, E.S. Lander, S.A. McCarroll, J.C. Nemesh, R.E. Poplin, S.C. Yoon, J. Lihm, V. Makarov, A.G. Clark, S. Gottipati, A. Keinan, J.L. Rodriguez-Flores, J.O. Korbel, T. Rausch, M.H. Fritz, A.M. Stütz, P. Flicek, K. Beal, L. Clarke, A. Datta, J. Herrero, W.M. McLaren, G.R.S. Ritchie, R.E. Smith, D. Zerbino, X. Zheng-Bradley, P.C. Sabeti, I. Shlyakhter, S.F. Schaffner, J. Vitti, D.N. Cooper, E.V. Ball, P.D. Stenson, D.R. Bentley, B. Barnes, M. Bauer, R. Keira Cheetham, A. Cox, M. Eberle, S. Humphray, S. Kahn, L. Murray, J. Peden, R. Shaw, E.E. Kenny, M.A. Batzer, M.K. Konkel, J.A. Walker, D.G. MacArthur, M. Lek, R. Sudbrak, V.S. Amstislavskiy, R. Herwig, E.R. Mardis, L. Ding, D.C. Koboldt, D. Larson, K. Ye, S. Gravel, A. Swaroop, E. Chew, T. Lappalainen, Y. Erlich, M. Gymrek, T. Frederick Willems, J.T. Simpson, M.D. Shriver, J.A. Rosenfeld, C.D. Bustamante, S.B. Montgomery, F.M. De La Vega, J.K. Byrnes, A.W. Carroll, M.K. DeGorter, P. Lacroute, B.K. Maples, A.R. Martin, A. Moreno-Estrada, S.S. Shringarpure, F. Zakharia, E. Halperin, Y. Baran, C. Lee, E. Cerveira, J. Hwang, A. Malhotra, D. Plewczynski, K. Radew, M. Romanovitch, C. Zhang, F.C.L. Hyland, D.W. Craig, A. Christoforides, N. Homer, T. Izatt, A.A. Kurdoglu, S.A. Sinari, K. Squire, S.T. Sherry, C. Xiao, J. Sebat, D. Antaki, M. Gujral, A. Noor, K. Ye, E.G. Burchard, R.D. Hernandez, C.R. Gignoux, D. Haussler, S.J. Katzman, W. James Kent, B. Howie, A. Ruiz-Linares, E.T. Dermitzakis, S.E. Devine, G.R. Abecasis, H. Min Kang, J. M. Kidd, T. Blackwell, S. Caron, W. Chen, S. Emery, L. Fritsche, C. Fuchsberger, G. Jun, B. Li, R. Lyons, C. Scheller, C. Sidore, S. Song, E. Sliwerska, D. Taliun, A. Tan, R. Welch, M. Kate Wing, X. Zhan, P. Awadalla, A. Hodgkinson, Y. Li, X. Shi, A. Quitadamo, G. Lunter, G.A. McVean, J.L. Marchini, S. Myers, C. Churchhouse, O. Delaneau, A. Gupta-Hinch, W. Kretzschmar, Z. Iqbal, I. Mathieson, A. Menelaou, A. Rimmer, D.K. Xifara, T.K. Oleksyk, Y. Fu, X. Liu, M. Xiong, L. Jorde, D. Witherspoon, J. Xing, E.E. Eichler, B.L. Browning, S.R. Browning, F. Hormozdiari, P.H. Sudmant, E. Khurana, R.M. Durbin, M.E. Hurles, C. Tyler-Smith, C.A. Albers, Q. Ayub, S. Balasubramaniam, Y. Chen, V. Colonna, P. Danecek, L. Jostins, T.M. Keane, S. McCarthy, K. Walter, Y. Xue, M.B. Gerstein, A. Abyzov, S. Balasubramanian, J. Chen, D. Clarke, Y. Fu, A.O. Harmanci, M. Jin, D. Lee, J. Liu, X. Jasmine Mu, J. Zhang, Y. Zhang, Y. Li, R. Luo, H. Zhu, C. Alkan, E. Dal, F. Kahveci, G.T. Marth, E.P. Garrison, D. Kural, W.-P. Lee, A.N. Ward, J. Wu, M. Zhang, S.A. McCarroll, R.E. Handsaker, D.M. Altshuler, E. Banks, G. del Angel, G. Genovese, C. Hartl, H. Li, S. Kashin, J.C. Nemesh, K. Shakir, S.C. Yoon, J. Lihm, V. Makarov, J. Degenhardt, J.O. Korbel, M.H. Fritz, S. Meiers, B. Raeder, T. Rausch, A.M. Stütz, P. Flicek, F. Paolo Casale, L. Clarke, R.E. Smith, O. Stegle, X. Zheng-Bradley, D.R. Bentley, B. Barnes, R. Keira Cheetham, M. Eberle, S. Humphray, S. Kahn, L. Murray, R. Shaw, E.-W. Lameijer, M.A. Batzer, M.K. Konkel, J. A. Walker, L. Ding, I. Hall, K. Ye, P. Lacroute, C. Lee, E. Cerveira, A. Malhotra, J. Hwang, D. Plewczynski, K. Radew, M. Romanovitch, C. Zhang, D.W. Craig, N. Homer, D. Church, C. Xiao, J. Sebat, D. Antaki, V. Bafna, J. Michaelson, K. Ye, S.E. Devine, E.J. Gardner, G.R. Abecasis, J.M. Kidd, R.E. Mills, G. Dayama, S. Emery, G. Jun, X. Shi, A. Quitadamo, G. Lunter, G.A. McVean, K. Chen, X. Fan, Z. Chong, T. Chen, D. Witherspoon, J. Xing, E.E. Eichler, M.J. Chaisson, F. Hormozdiari, J. Huddleston, M. Malig, B.J. Nelson, P.H. Sudmant, N.F. Parrish, E. Khurana, M.E. Hurles, B. Blackburne, S.J. Lindsay, Z. Ning, K. Walter, Y. Zhang, M.B. Gerstein, A. Abyzov, J. Chen, D. Clarke, H. Lam, X. Jasmine Mu, C. Sisu, J. Zhang, Y. Zhang, R.A. Gibbs, F. Yu, M. Bainbridge, D. Challis, U.S. Evani, C. Kovar, J. Lu, D. Muzny, U. Nagaswamy, J.G. Reid, A. Sabo, J. Yu, X. Guo, W. Li, Y. Li, R. Wu, G.T. Marth, E.P. Garrison, W. Fung Leong, A.N. Ward, G. del Angel, M.A. DePristo, S.B. Gabriel, N. Gupta, C. Hartl, R. E. Poplin, A.G. Clark, J.L. Rodriguez-Flores, P. Flicek, L. Clarke, R.E. Smith, X. Zheng-Bradley, D.G. MacArthur, E.R. Mardis, R. Fulton, D.C. Koboldt, S. Gravel, C.D. Bustamante, D.W. Craig, A. Christoforides, N. Homer, T. Izatt, S.T. Sherry, C. Xiao, E.T. Dermitzakis, G.R. Abecasis, H. Min Kang, G.A. McVean, M.B. Gerstein, S. Balasubramanian, L. Habegger, H. Yu, P. Flicek, L. Clarke, F. Cunningham, I. Dunham, D. Zerbino, X. Zheng-Bradley, K. Lage, J. Berg Jespersen, H. Horn, S.B. Montgomery, M.K. DeGorter, E. Khurana, C. Tyler-Smith, Y. Chen, V. Colonna, Y. Xue, M.B. Gerstein, S. Balasubramanian, Y. Fu, D. Kim, A. Auton, A. Marcketta, R. Desalle, A. Narechania, M.A. Wilson Sayres, E.P. Garrison, R.E. Handsaker, S. Kashin, S.A. McCarroll, J.L. Rodriguez-Flores, P. Flicek, L. Clarke, X. Zheng-Bradley, Y. Erlich, M. Gymrek, T. Frederick Willems, C.D. Bustamante, F.L. Mendez, G. David Poznik, P.A. Underhill, C. Lee, E. Cerveira, A. Malhotra, M. Romanovitch, C. Zhang, G.R. Abecasis, L. Coin, H. Shao, D. Mittelman, C. Tyler-Smith, Q. Ayub, R. Banerjee, M. Cerezo, Y. Chen, T.W. Fitzgerald, S. Louzada, A. Massaia, S. McCarthy, G.R. Ritchie, Y. Xue, F. Yang, R.A. Gibbs, C. Kovar, D. Kalra, W. Hale, D. Muzny, J.G. Reid, J. Wang, X. Dan, X. Guo, G. Li, Y. Li, C. Ye, X. Zheng, D.M. Altshuler, P. Flicek, L. Clarke, X. Zheng-Bradley, D.R. Bentley, A. Cox, S. Humphray, S. Kahn, R. Sudbrak, M.W. Albrecht, M. Lienhard, D. Larson, D.W. Craig, T. Izatt, A.A. Kurdoglu, S.T. Sherry, C. Xiao, D. Haussler, G.R. Abecasis, G.A. McVean, R.M. Durbin, S. Balasubramaniam, T.M. Keane, S. McCarthy, J. Stalker, A. Chakravarti, B.M. Knoppers,

G.R. Abecasis, K.C. Barnes, C. Beiswanger, E.G. Burchard, C.D. Bustamante, H. Cai, H. Cao, R.M. Durbin, N.P. Gerry, N. Gharani, R.A. Gibbs, C.R. Gignoux, S. Gravel, B. Henn, D. Jones, L. Jorde, J.S. Kaye, A. Keinan, A. Kent, A. Kerasidou, Y. Li, R. Mathias, G. A. McVean, A. Moreno-Estrada, P.N. Ossorio, M. Parker, A.M. Resch, C.N. Rotimi, C.D. Royal, K. Sandoval, Y. Su, R. Sudbrak, Z. Tian, S. Tishkoff, L.H. Toji, C. Tyler-Smith, M. Via, Y. Wang, H. Yang, L. Yang, J. Zhu, W. Bodmer, G. Bedoya, A. Ruiz-Linares, Z. Cai, Y. Gao, J. Chu, L. Peltonen, A. Garcia-Montero, A. Orfao, J. Dutil, J.C. Martinez-Cruzado, T.K. Oleksyk, K.C. Barnes, R.A. Mathias, A. Hennis, H. Watson, C. McKenzie, F. Qadri, R. LaRocque, P.C. Sabeti, J. Zhu, X. Deng, P.C. Sabeti, D. Asogun, O. Folarin, C. Happi, O. Omoniwa, M. Stremlau, R. Tariyal, M. Jallow, F. Sisay Joof, T. Corrah, K. Rockett, D. Kwiatkowski, J. Kooner, T. Tịnh Hiền, S.J. Dunstan, N. Thuy Hang, R. Fonnie, R. Garry, L. Kanneh, L. Moses, P.C. Sabeti, J. Schieffelin, D.S. Grant, C. Gallo, G. Poletti, D. Saleheen, A. Rasheed, L.D. Brooks, A.L. Felsenfeld, J.E. McEwen, Y. Vaydylevich, E.D. Green, A. Duncanson, M. Dunn, J.A. Schloss, J. Wang, H. Yang, A. Auton, L.D. Brooks, R.M. Durbin, E.P. Garrison, H. Min Kang, J.O. Korbel, J.L. Marchini, S. McCarthy, G.A. McVean, G.R. Abecasis, A global reference for human genetic variation, Nature 526 (2015) 68–74, https://doi.org/10.1038/nature15393.

[10] C.C. Chang, C.C. Chow, L.C. Tellier, S. Vattikuti, S.M. Purcell, J.J. Lee, S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. Ferreira, D. Bender, B. Browning, S. Browning, B. Howie, P. Donnelly, J. Marchini, A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, P. Danecek, A. Auton, G. Abecasis, C. Albers, E. Banks, M. DePristo, H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, J. Yang, S. Lee, M. Goddard, P. Visscher, V. Lee, C. Kim, J. Chhugani, M. Deisher, D. Kim, A. Nguyen, I. Haque, V. Pande, W. Walters, H. Hardy, J. Wigginton, D. Cutler, G. Abecasis, S. Guo, E. Thompson, C. Mehta, N. Patel, D. Clarkson, Y. Fan, H. Joe, F. Requena, N.M. Ciudad, S. Lydersen, M. Fagerland, P. Laake, J. Graffelman, V. Moreno, J. Wall, J. Pritchard, S. Gabriel, S. Schaffner, H. Nguyen, J. Moore, J. Roy, B. Blumenstiel, J. Barrett, B. Fry, J. Maller, M. Daly, W. Hill, T. Gaunt, S. Rodríguez, I. Day, D. Taliun, J. Gamper, C. Pattaro, J. Friedman, T. Hastie, H. Höfling, R. Tibshirani, S. Vattikuti, J. Lee, C. Chang, S. Hsu, C. Chow, V. Steiß, T. Letschert, H. Schäfer, R. Pahl, X. Wan, C. Yang, Q. Yang, H. Xue, X. Fan, N. Tang, M. Ueki, H. Cordell, G. Abecasis, L. Cardon, W. Cookson, W. Ewens, M. Li, R. Spielman, Z. Su, J. Marchini, P. Donnelly, Y. Xu, Y. Wu, C. Song, H. Zhang, D. Defays, B. Browning, S. Browning, B. Browning, P. Loh, M. Baym, B. Berger, F. Sambo, B. Di Camillo, G. Toffolo, C. Cobelli, Second-generation PLINK: rising to the challenge of larger and richer datasets, GigaScience 4 (2015) 7, https://doi.org/10.1186/s13742-015-0047-8.

[11] D. Karolchik, G.P. Barber, J. Casper, H. Clawson, M.S. Cline, M. Diekhans, T.R. Dreszer, P. a Fujita, L. Guruvadoo, M. Haeussler, R. a Harte, S. Heitner, A.S. Hinrichs, K. Learned, B.T. Lee, C.H. Li, B.J. Raney, B. Rhead, K.R. Rosenbloom, C. a Sloan, M.L. Speir, A.S. Zweig, D. Haussler, R.M. Kuhn, W.J. Kent, The UCSC Genome Browser database: 2014 update, Nucleic Acids Res. 42 (2014) D764–D770, https://doi.org/10.1093/nar/gkt1168.

[12] M. Krzywinski, J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S.J. Jones, M. A Marra, Circos: an information aesthetic for comparative genomics, Genome Res. 19 (2009) 1639–1645, https://doi.org/10.1101/gr.092759.109.