

RESEARCH ARTICLE

Open Access



Copy number variation of *CCL3L1* among three major ethnic groups in Malaysia

Jalilah Jamaluddin, Nur Khairina Mohd Khair, Shameni Devi Vinodamane, Zulkefley Othman and Suhaili Abubakar* 

Abstract

Background: C-C motif Chemokine Ligand 3 Like 1 (*CCL3L1*) is a multiallelic copy number variable, which plays a crucial role in immunoregulatory and hosts defense through the production of macrophage inflammatory protein (MIP)-1 α . Variable range of the *CCL3L1* copies from 0 to 14 copies have been documented in several different populations. However, there is still lack of report on the range of *CCL3L1* copy number exclusively among Malaysians who are a multi-ethnic population. Thus, this study aims to extensively examine the distribution of *CCL3L1* copy number in the three major populations from Malaysia namely Malay, Chinese and Indian. A diploid copy number of *CCL3L1* for 393 Malaysians (Malay = 178, Indian = 90, and Chinese = 125) was quantified using Paralogue Ratio Tests (PRTs) and then validated with microsatellites analysis.

Results: To our knowledge, this is the first report on the *CCL3L1* copy number that has been attempted among Malaysians and the Chinese ethnic group exhibits a diverse pattern of *CCL3L1* distribution copy number from the Malay and Indian ($p < 0.0001$). The *CCL3L1* ranged from 0 to 8 copies for both the Malay and Indian ethnic groups while 0 to 10 copies for the Chinese ethnic. Consequently, the *CCL3L1* copy number among major ethnic groups in the Malaysian population is found to be significantly varied when compared to the European population ($p < 0.0001$). The mean/median reported for the Malay, Chinese, Indian, and European are 2.759/2.869, 3.453/3.290, 2.437/1.970 and 2.001/1.940 respectively.

Conclusion: This study reveals the existence of genetic variation of *CCL3L1* in the Malaysian population, and suggests by examining genetic diversity on the ethnicity, and specific geographical region could help in reconstructing human evolutionary history and for the prediction of disease risk related to the *CCL3L1* copy number.

Keywords: Copy number variation, CNV, *CCL3L1*, Malaysian population, Paralogue ratio test, PRT

Background

Malaysia is geographically located in Southeast Asia with a total land area of approximately 328,000 km², separated by the South China Sea into two regions namely Peninsular Malaysia and East Malaysia (Labuan, Sabah, and Sarawak) [1]. The population of Malaysian is nearly 32 million which is equivalent to 0.42% of the world's populations. Uniquely, Malaysia is populated by different ethnic groups which largely comprised Malay (68.8%), followed by Chinese (23.2%), Indian (7.0%), and other ethnic groups (1%) [2]. Delineation of ethnic groups in Malaysia based on genetic variation has been carried out on several genes such as

UGT1A1 and *CYP2*. *UGT1A1**28 was found to be high in frequency among the Malay and Indian as compared to the Chinese ethnic group [3]. Furthermore, the presence of *CYP2B6**6 among the Malay is also reported in a higher frequency compared to the other ethnic groups [4]. *CYP2D6**10 is recently observed in both Malay and Chinese while *CYP2D6**4 in the Indian ethnic group has been observed among breast cancer patients [5]. Besides, genome-wide of single nucleotide polymorphism (SNP) analysis also reported that the Malay group genetically differs from Chinese and Indian [6, 7]. This study suggested population-specific reference is important to explore unobserved common SNPs particularly for the Malay group which the data is limited in the database [6]. In addition, the gene with copy number variable is also less

* Correspondence: suhaili_ab@upm.edu.my

Department of Biomedical Science, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, UPM, 43400 Serdang, Selangor, Malaysia



described among Malaysian, particularly in relation to the Malay who represent the largest populations of the Malay Archipelago- mainly present in Southeast Asia [8, 9].

Copy number variation (CNV) is a repetitive element in population diversity which results from the losses or gains of DNA segments of more than 1 kilobase which leads to different copies being presented [10, 11]. Among the reported genes with CNV, C Chemokine Ligand 3 Like 1 (*CCL3L1*) has been identified as one of the prominent genes with CNV in the current research field [12]. *CCL3L1* which was previously known as *LD78β*, is clustered on chromosome 17q12. This gene showed involvement in segmental duplication which was enriched with the immune response gene. According to the *Homo sapiens* Annotation Release 108, *CCL3L1* harboured at the base pairs of 250,577 to 252,466 on chromosome 17q21. The repeat unit of *CCL3L1* is approximately 90 kb in size and within the same repeat unit, lies together *CCL4L1*. Both *CCL3L1* and *CCL4L1* share 95% similarity with their own non-copy variable paralogues, *CCL3* and *CCL4*, respectively. Therefore, *CCL4L1* frequently acts as an additional parameter, allowing accurate determination of *CCL3L1* copy number [13–15]. *CCL3L1* encodes for LD78β, the isoforms of macrophage inflammatory protein 1α (MIP-1α) which only differs by three amino acids of the *CCL3* mature protein (LD78α) [16, 17]. The MIP-1α products play an important part in anti-tumour immunity and participate in the immunoregulatory and inflammatory processes [18, 19]. The truncated-2 form of LD78β produced by *CCL3L1* leads to high binding affinity to the CC chemokine receptor (CCR) 1 and 5 [16, 20]. This gene expression is in a different dosage which will eventually lead to the diversity of ethnic phenotypic and may contribute to ethnic-specific diseases [21, 22].

Interestingly, different ranges of *CCL3L1* diploid copy number have been observed among different populations such as 0 to 4 copies in the United Kingdom [23, 24], 0 to 11 copies in the United States of America [25], 0 to 14 copies in Africa [12, 26], 0 to 10 copies in China [27], 1 to 9 copies in Korea [28], and 0 to 10 copies in Japan [29]. While the other populations have been extensively sampled, Southeast Asian population have rarely been included. Although *CCL3L1* copy number has been proven to vary in different populations, such information is yet to be reported in Malaysia. Thus, this study aims to investigate the distribution of *CCL3L1* copy number among three major ethnicities in the Malaysian population. Indeed, the characterization of Malay genetic structure would be beneficial data in Southeast Asia which is poorly explored in genetic studies. Additionally, it may provide a starting base for fundamental insight into the association of *CCL3L1* with related diseases.

In order to report the copy number, there are varieties of platforms used such as real-time qPCR [12]. However,

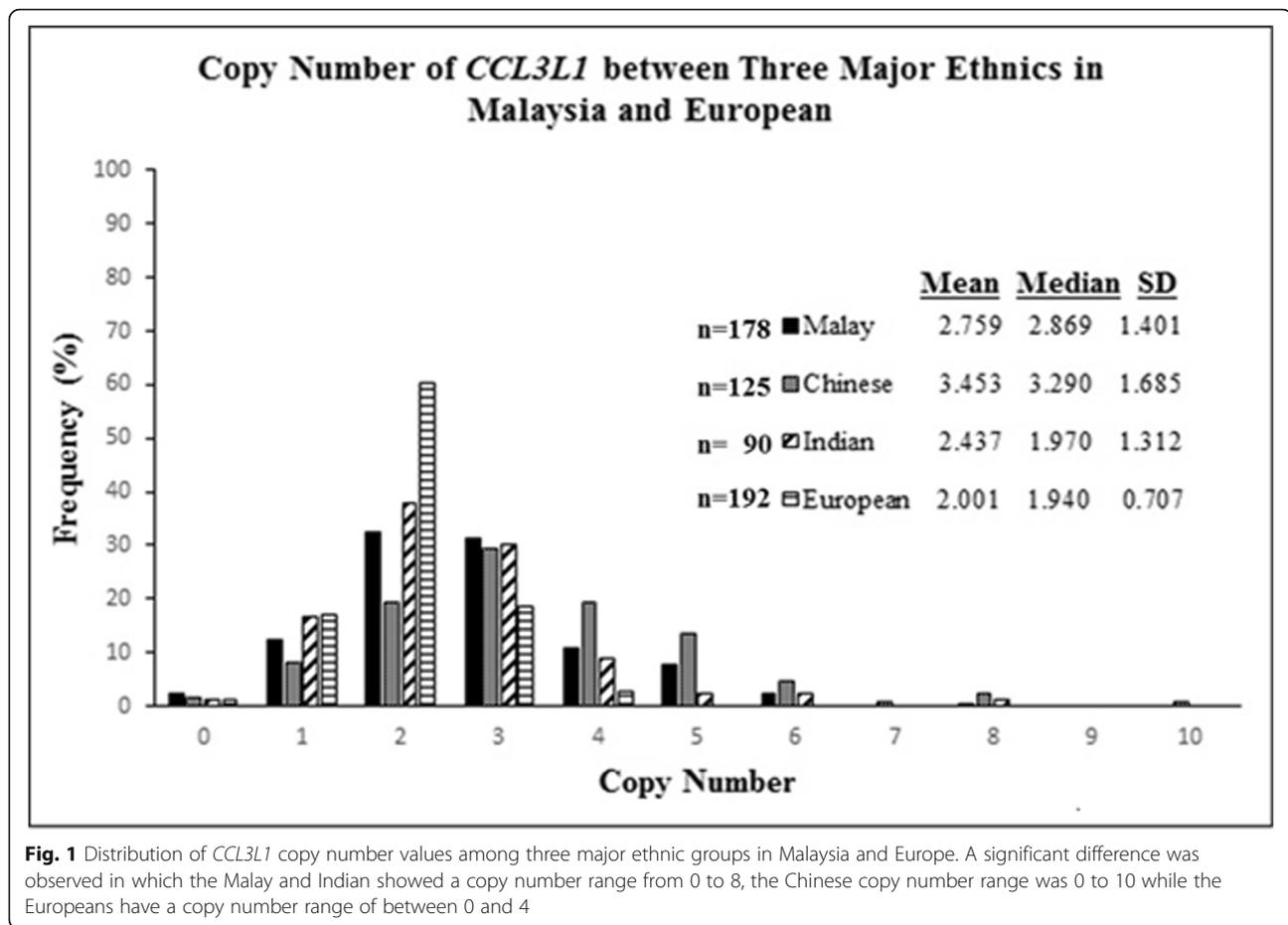
in this study, we adapted the Parologue Ratio Test (PRT), a PCR-based technique which was designed specifically to simultaneously amplify two products with a set of primer; one which yields from the reference locus and another one from the test locus [24, 30]. PRT has shown a high accuracy level in measuring the copy number and the amount of DNA previously used was as little as 10 ng [24]. Furthermore, PRT is a system that offers a very simple, rapid, yet inexpensive method for quantifying copy number which is compatible with the application to genotype more than thousands of samples [30].

Results

Distribution of *CCL3L1* copy number among three major ethnics in Malaysia

CCL3L1 was found in variable copies where the Malay and Indian shared a similar range of 0 to 8 copies while the Chinese exhibit a wider range of 0 to 10 copies (Fig. 1). However, the Chinese ($n = 125$) and Malay ($n = 178$) shared a common copy number of 3 whereas a copy number of 2 was commonly found in Indian only ($n = 90$). The mean values were followed by the median values for each ethnic group comprising the Malay, Chinese and Indian are 2.759/2.869, 3.453/3.290 and 2.437/1.970 respectively. The histogram of the unrounded *CCL3L1* copy number obtained from major Malaysian ethnic groups was plotted to show the distribution of each group's copy number (see Additional file 2). It is clearly showed that there was a grouping of data around each integer copy number even though there was no clear gap between each copy number's division; rather, there was an overlap with each division at the end of the left and right sides for each copy number group. One-way ANOVA analysis revealed that the distribution of the *CCL3L1* copy number among those three major ethnic groups was significantly different with the p -value < 0.0001 .

As we were emphasizing on the Malay CNV data as the major population, the Chinese and Indian ethnic groups were included for comparison purposes. Since the characteristics of the *CCL3L1* copy number among the Malay, Chinese and Indian groups have been established, the comparison to the European population was directly carried out. The *CCL3L1* copy number used for the European population in this comparison was obtained from a previous report by Walker et al. [24]. The distribution clearly showed a different pattern of *CCL3L1* copy number among the four groups (Fig. 1). The mean value of *CCL3L1* copy number reported for the European population were 2 copies. An analysis by the One-way ANOVA strongly showed a significant difference in the *CCL3L1* copy number between the three major ethnic groups in Malaysia when they were compared to the European population (p -value of < 0.0001). Table 1 showed the comparison of the copy number for three major ethnic groups in Malaysia and we had also



included European, Norwegian and South African population as well. Result analysed by X^2 showed significant difference of *CCL3L1* copy number range in Malaysian population between Malay and Chinese groups ($X^2 = 18.72$; p -value = 0.0277). In addition, X^2 analysis proved that the *CCL3L1* copy number distribution of Malay group is different when compared to European (p -value of < 0.0001), Norwegian (p -value of < 0.0001) and South African groups (p -value of < 0.0001) without consideration of the methodological approaches used.

Discussion

To the best of our knowledge, this is the first report on *CCL3L1* copies in Malaysia. In this study, we had investigated the distribution of the *CCL3L1* copies among the three major ethnic groups: Malay, Indian and Chinese using the PCR-based analysis (PRTs and microsatellites). Genetic diversity of the *CCL3L1* copy number among the Malaysian population is in agreement with the preceding findings [4, 6, 34]. The *CCL3L1* copy number range generally fall within the Asian regions which include Korea 1–9 copies [28], Japan 0–10 copies [29], China 0–10 copies [27], and India 0–7 copies [35].

According to a previous study, it was reported that the haplotype harbouring was more than 90% among the East Asia population (including Japan and Korea) could also be found in the population of Southeast Asia and South Asia including Malaysia and India [36]. However, further comparative analysis of the mentioned population could not be carried out due to insufficient data.

In the other regions, the *CCL3L1* copy number was reported in diverse ranges such as (0–4 copies) for Europeans [23, 24], (1–11 copies) for African American [25], (0–6 copies) for South American (Peruvian) [33], and (0–14 copies) for Africans [12, 26]. Thus, this data suggests that different population from different geographical regions has their own unique *CCL3L1* copy number range and that could act as a source of genetic diversity. The distribution of the *CCL3L1* copy number could be predominantly classified into three groups which include European, African and East Asian. This finding is in accordance with the recent report of the principal component analysis (PCA) plot on the global population. PCA has been widely performed to identify the genetic ancestry difference of individuals from various populations in order to generate the population structure [37]. These

Table 1 Comparison of the copy number followed by frequency (%) calculated in bracket between the ethnic groups in Malaysia and other reported populations

Copy number	Malaysia			European [24]	UK Biobank [31]	Norwegian [32]	South African [33]
	Malay	Chinese	Indian				
0	4 (2.25%)	2 (1.60%)	1 (1.11%)	2 (1.04%)	127 (2.55%)	11 (1.22%)	4 (0.81%)
1	22 (12.36%)	10 (8.00%)	15 (16.67%)	33 (17.19%)	1046 (21.00%)	159 (17.57%)	17 (3.46%)
2	58 (32.58%)	24 (19.20%)	34 (37.78%)	116 (60.42%)	2806 (56.33%)	527 (58.23%)	64 (13.0%)
3	56 (31.46%)	37 (29.60%)	27 (30.00%)	36 (18.75%)	853 (17.13%)	172 (19.01%)	131 (24.63%)
4	19 (10.67%)	24 (19.20%)	8 (8.89%)	5 (2.60%)	128 (2.57%)	35 (3.87%)	128 (26.02%)
5	14 (7.87%)	17 (13.60%)	2 (2.22%)	–	21 (0.42%)	1 (0.11%)	76 (15.45%)
6	4 (2.25%)	6 (4.80%)	2 (2.22%)	–	–	–	44 (8.94%)
7	–	1 (0.80%)	–	–	–	–	21 (4.27%)
8	1 (0.56%)	3 (2.40%)	1 (1.11%)	–	–	–	7 (1.42%)
9	–	–	–	–	–	–	–
10	–	1 (0.80%)	–	–	–	–	–
11	–	–	–	–	–	–	–
12	–	–	–	–	–	–	1 (0.20%)
Total	178	125	90	192	4981	905	493
X ²	–	18.72	5.35	53.26	363.70	136.50	85.31
P-value	–	0.0277	0.6177	< 0.0001	< 0.0001	< 0.0001	< 0.0001

might strengthen the existence of different *CCL3L1* copy number between the populations. However, this should be further investigated to provide a strong conclusion on the relationship between *CCL3L1* diversity with different populations.

In addition, the significant difference of the *CCL3L1* copy number distribution among ethnicities found within the population (Malaysia) was in line with a study from the African population which suggested that ethnic background might influence the genetic variations [12, 38]. A particular reason behind copy number divergence between the populations and the ethnic groups remained unknown. Emphasizing the Malay in the Malaysian population, we noted a valuable finding when compared to other ethnic groups. The present result exhibited a different range of *CCL3L1* copy numbers between the Malay and Chinese groups in contrast to a previous study which reported that the Malay is genetically similar to the Chinese [6]. However, another study reported that *CCL3L1* is greater among Han-Chinese from Zhejiang province of China up to ten copies [27]. In addition the descendants of Chinese in Malaysia are predominantly from Southern China [39]. Therefore, extensive analysis including replicate or more studies is crucial in order to conclude the origin of Chinese in Malaysia. On the other hand, the analysed data displayed a sharing distribution of *CCL3L1* copy number between the Malay and Indian. Our analysis supports a study that revealed the presence of genetic influx from Indian to Malay as a result of admixture during the early existence of the Malay group [40]. This may provide a plausible

explanation of the similar range of distribution of the copy number shared between these two groups as reported in this study.

Nevertheless, a whole genome sequencing study by Wong et al. (2013) stated that the Malay are significantly distinct from the Indian and Chinese [6, 34]. Intriguingly, it has been reported that Malay was a product of admixture between East Asian, South Asian, Austro-nesian, and Southeast Asian aboriginal people which lead to complex ancestry of Malaysian genetic data [34, 37, 41, 42]. Notably, the admixture event was also encountered within the subgroups of the Malay, observed among Malay in Peninsular Malaysia. A distinct genetic difference was reported between the Malay from the south and north of Peninsular Malaysia. This was probably due to the prolonged isolation followed by subsequent gene flow and adaptation, leading to changes in the genetic architectures of the Malay in the Malaysian population [34]. Therefore, a complex genetic diversity can be observed among the Malay in the Malaysian population. Unfortunately, genetic diversity data of the Malay and other ethnic groups in the Malaysian population is still insufficient to draw a concrete conclusion, particularly on the genes with CNV. This suggests that intensive investigations is needed in order to explore these ethnic groups further. Despite the fact that the single gene-focus in this study could not estimate the population's history, it is a profoundly valuable contribution to the literature on *CCL3L1* copy number across the Malaysian population especially the Malay. Besides, the data is valuable in the efforts to investigate the genetic

heterogeneity of *CCL3L1* among other ethnic groups in Malaysia and its association with related diseases.

Different types of platform or technique used for the determination of the *CCL3L1* copy number might influence the obtained range. In this study, a combination of PCR-based approaches includes PRTs and microsatellites were used although qPCR was frequently employed in the previous studies [12, 29, 43–45], and MLPA and ddPCR were recently used more [46, 47]. Therefore, the comparison between the copy number obtained between previous studies may not be as constant as it might generate more or less copy number than PRT. As for the PRTs assay, it has previously been shown to have a high degree of accuracy in predicting the integer copy number by identifying the discrete clusters with a consistency of microsatellite data [23, 24]. Even though there was a possibility of amplification on the pseudogene that might influence the final copy number reported, particularly for a high copy but the primers were specifically designed to amplify within intron 1, which was majority absent in pseudogene [22]. Meanwhile, it has been proven that qPCR tends to lead to false positive results as reported by Field et al. (2010) [48]. This is because qPCR primers amplify exon 3 within *CCL3L1*, and the pseudogene which resulted in extra copies of *CCL3L1* obtained [47]. Another finding also showed that qPCR might potentially lead to false-positive calls. Therefore, any studies of CNV association using only qPCR analysis have been suggested to perform a validation calling of the copy number [49, 50].

Until now, variable copies of the *CCL3L1* have yet to be investigated in the Asian populations compared to those from the European or American populations. Studies of *CCL3L1* copy number and its association with the susceptibility of related diseases have been found in European, Japanese, African, and Korean populations and a similar study could be conducted among Malaysians as well. Thus, more information about the CNV database particularly involving the Malay group is required as it represents one of the largest populations in Southeast Asia. The present finding had increased the CNV status data on the Malay along with the understanding of human genetic variations in relation to the different populations. It has also provided a fundamental implication in genetic studies. The impact of this study may offer a starting point for diseases susceptibility studies and other CNV diversity among the populations. The investigation of diseases susceptibility in association with *CCL3L1* variability in Malaysia is highly recommended.

Several limitations were identified in this present study. We clearly did not study on the association of *CCL3L1* copy number with any related diseases, thus we could not provide any association between those two. Besides, analysis on larger number of the sample size

from each ethnic groups is highly suggested to increase the statistical power. This initiative might avoid any bias in mapping the genetic structure of Malaysian.

Conclusion

In brief, this comprehensive study had allowed us to provide further insight into the distribution of *CCL3L1* copy number in the Malaysian population where a significant scattering has been ascertained among the three major ethnic groups. This present study offers a valuable genetic frequency in Malaysian population which currently underrepresented in population genetic studies. Future human evolutionary and association studies in relation to diseases will appreciate the availability of this data. Apart from that, ethnicity is another highly necessary variable for any research of genetic mapping.

Methods

Ethics approval, study design and subject recruitment

Ethical approval was obtained from the Ethics Committee of Universiti Putra Malaysia (UPM/TNCPI/RMC/1.4.18.1 (JKEUPM)/F2). All the subjects were given an explanation in regards to the study through the Respondent Information Sheet, and their written consent was obtained through the Respondent Consent Form. Whole blood samples were withdrawn in the range of 1 to 3 mL. A total of 430 unrelated volunteers comprising three major ethnics in Malaysia; Malay ($n = 189$), Chinese ($n = 144$) and Indian ($n = 97$) was successfully collected and quantified for the *CCL3L1* copy number. The subjects were students and staffs recruited from Universiti Putra Malaysia (UPM) Serdang, Selangor. Volunteers must fulfil the inclusion criteria as Malaysian, age between 18 and 55 years old, no mixed ethnicity for at least up to three (3) generations due to interracial marriage hence could be grouped as Malay, Chinese and Indian, and no clinical history of HIV or any autoimmune diseases. The determination of the ethnic groups was self-identified by the volunteers.

DNA extraction

QIAamp DNA Mini Kit (QIAGEN, Germany) was used to extract the DNA from 200 μ l of peripheral blood collected in BD Vacutainer® K2 EDTA tubes. Then, the concentration and purity of the extracted DNA were quantified using Nanodrop ND-1000 spectrophotometer (Thermo, USA).

Quantification and validation of *CCL3L1* copy number

CCL3L1 copies were first quantified using three Paralog Ratio Tests (PRTs); *CCL3A*, *CCL4A*, and *LTR61A*, as described in the previous study [24]. In principle, PRT is a PCR-based method which only uses one pair of primer to simultaneously amplify two different loci, test

(copy number variable) and reference (copy number invariable) [30]. The first system, CCL3A uses a set of primers designed to amplify both *CCL3L1* (test locus) and *CCL3* (reference locus) simultaneously on chromosome 17. The second system, CCL4A is developed to support the clarification of the *CCL3L1* copy number quantified by the CCL3A system where the primer for the CCL4A amplifies *CCL4L1* (test locus) and *CCL4* (reference locus). LTR61A is the third system, which also assists in verifying the copy number obtained which harbours on the long terminal repeat sequence located between *CCL3L1* and *CCL4L1* on chromosome 17 against the unlinked reference locus on chromosome 10 [24]. The three primer sets used to perform the capillary electrophoresis were fluorescently-labelled; HEX-CCL3A Forward (TCATAGTGGGTTCTCTGTTTC) with CCL3 Reverse (ATCCAGGGCTGCTTACTT), HEX-LTR61A Forward (AGTTTTCTCTGCC TAGC) with LTR61A Reverse (TATTTATTTAAGG TGTGCAC), and FAM-CCL4A Forward (GAGTCTGCTTCCAGTGCT) with CCL4A Reverse (GAGGAGTCCTGAGTATGGAG). The determination of the copy number was carried out by initially calculating the ratio obtained from the test to reference loci for each system.

To validate the integer of copy number calling by PRTs, two microsatellites lying within the *CCL3L1* region (TATC17 and TTAT17) were analysed in this study [24]. The amplification of the TATC17 region was carried out using non-labelled Forward primer TATC17 (CTTAGGGGTCCTCTTGTC) and HEX-labelled Reverse primer TATC17 (CCAAAATCTGGATTAGT-CAG). Meanwhile, for TTAT17, the amplification was carried out using FAM-labelled TTAT17 Forward primer (TCAGTTTTGCAAACGACCA) and non-labelled Reverse primer TTAT17 (GAAACTGGAAGGTGGAGATG). The amplified microsatellite products were resolved by capillary electrophoresis. The peak ratios of all the alleles to the smallest allele were calculated to infer the copy number, and the calculated total allelic ratio is considered as the integer copy number [24, 51].

Quality control

European Collection of Authenticated Cell Cultures (ECACC) Human Random Control DNA samples known for their *CCL3L1* copy number from previous typing [24] were used as the internal control (reference) to infer the copy number for the unknown samples by deriving a linear regression equation from each PRT systems. The samples were C0075 (cn = 1), C0150 (cn = 2), C0007 (cn = 3) and C0877 (cn = 4). In addition, the typing missingness for the reference samples were also analysed to estimate the error rate of per-test. There was a total of 96 typing for all four (4) samples, and the incorrect calling copy numbers were 19 times, which mainly involved sample represented for three copy number (C0007). This analysis suggested

that the error rates for the whole systems together were 20%, divided by three for three PRT systems (CCL3A, CCL4A and LTR61A) results in approximately 7% for each system. Armour et al. stated that in the determination of *DEFB* copy number, a single PRT test might contribute to approximately 8% of the error rate in determining the copy number [30]. Thus, the copy number typing of the reference samples from the whole set of experiments were considered as reproducible. Kolmogorov-Smirnov normality test was also performed on these samples in order to estimate the probability of the copy number group misclassification. All four copy number groups showed no significant difference, suggesting that the samples were normally distributed.

CCL3L1 copy number for the unknown copy number samples showed 76% concordant either between two or three PRTs systems ($n = 328$). However, the agreement of the copy number calling between the PRTs system and microsatellite analyses found that (91%) 393 concordant out of 430 samples, leaving 37 samples being discordant. All the 37 samples with discordance were excluded from the statistical analysis due to the inconsistency of the copy number in all the assays (PRTs and microsatellites) (see Additional file 1). Thus, based on the agreement of the integer values between PRTs and microsatellites, only 393 samples were included for further statistical analysis.

Statistical analyses

One-way ANOVA was used to calculate the significant difference of the *CCL3L1* copy number between three major ethnics as well as in comparison between the ethnic groups in Malaysia and the European population. Chi-square (X^2) test of homogeneity was performed to identify the significance different in the frequency of *CCL3L1* copy number distribution among Malaysian as well as to the other populations. A statistical value is considered significant when the *p-value* is less than 0.05. The descriptive analyses between the population and the major ethnic groups in Malaysia were also calculated and described.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12863-019-0803-3>.

Additional file 1. Agreement between PRT and microsatellites. Samples with discordancy of copy number were excluded from further statistical analysis.

Additional file 2. Distribution of unrounded copy number among three major ethnics. Overlapping of copy number observed at the end (right and left sides) for each copy number group.

Abbreviations

CCL3L1: Chemokine ligand 3-like 1; CNV: copy number variation; PRT: Paralogue ratio test

Acknowledgements

We thank all the staff of Molecular Biology and Bioinformatics Laboratory, UPM for providing sufficient equipment to allow us to perform the experiments.

Authors' contributions

Conceived and designed the experiments: SA. Performed the sample recruitment and copy number typing: JJ, NKMK, SDV. Conducted the statistical analysis: JJ, ZO and SA. Provided the reagents/ materials/ analysis tools to carry out the experiments: SA. Wrote the paper: JJ, ZO and SA. All authors read and approved the final manuscript.

Funding

This study was funded by the Ministry of Higher Education (FRGS-2013(2)5524413), JJ received a postgraduate scholarship from the School of Graduate Studies (Graduate Research Fellowship, Universiti Putra Malaysia) and Ministry of Higher Education Malaysia (MyBrain15). The FRGS-2013(2)5524413 had contributed to the study design and samples collection, data analysis and interpretation, and in writing the manuscript. The Graduate Research Fellowship, Universiti Putra Malaysia, and Ministry of Higher Education Malaysia (MyBrain15) had no role in the design, samples and data collection, and analysis of the study.

Availability of data and materials

The data sets of this article are included within the article and its additional files.

Ethics approval and consent to participate

This study protocol had obtained the approval from the Ethics Committee of Universiti Putra Malaysia (UPM/TNCPI/RMC/1.4.18.1 (JKEUPM)/F2). All subjects were explained regarding the study through the Respondent Information Sheet, and their written consent was obtained from the Respondent Consent Form.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 6 May 2018 Accepted: 17 December 2019

Published online: 03 January 2020

References

- Halim-Fikri H, Etemad A, Latif AZA, Merican AF, Baig AA, Annuar AA, et al. The first Malay database toward the ethnic-specific target molecular variation. *BMC Res Notes*. 2015;8:176.
- Department of Statistics Malaysia. Press Release: Current Population Estimates, Malaysia, 2016–2017 [Internet]. Available from: https://www.dosm.gov.my/v1/index.php?r=column/cthemebByCat&cat=397&bul_id=SmtYsJf1eXpQTIrWh3Wkx5N1pyQT09&menu_id=L0pheU43NWjwRWw5ZkIWdzQ4TihUUT09. Accessed 2018.
- Teh LK, Hashim H, Zakaria ZA, Salleh MZ. Polymorphisms of UGT1A1*6, UGT1A1*27 & UGT1A1*28 in three major ethnic groups from Malaysia. *Indian J Med Res*. 2012;136(2):249–59.
- Musa N, Ismail R, Fauzi H, Talib N, Mohamad N, Zulkafli M. Haplotypes frequencies of CYP2B6 in Malaysia. *J Postgrad Med*. 2012;58:235.
- Chin FW, Chan SC, Abdul Rahman S, Noor Akmal S, Rosli R. CYP2D6 genetic polymorphisms and phenotypes in different ethnicities of Malaysian breast Cancer patients. *Breast J*. 2016;22:54–62.
- Wong LP, Ong RTH, Poh WT, Liu X, Chen P, Li R, et al. Deep whole-genome sequencing of 100 southeast Asian Malays. *Am J Hum Genet*. 2013;92:52–66.
- Norhalifah HK, Syaza FH, Chambers GK, Edinur HA. The genetic history of peninsular Malaysia. *Gene*. 2016;586:129–35.
- Deng L, Hoh B-P, Lu D, Saw W-Y, Tzee-Hee Ong R, et al. Dissecting the genetic structure and admixture of four geographical Malay populations. *Nat Publ Gr*. 2015;5:14375.
- Sarji A. The encyclopedia of Malaysia; The Rulers of Malaysia. 2011. (Edition Millet).
- Pietiläinen OPH, Rehnström K, Jakkula E, Service SK, Congdon E, Tilgmann C, et al. Phenotype mining in CNV carriers from a population cohort. *Hum Mol Genet*. 2011;20:2686–95.
- de Smith AJ, Walters RG, Froguel P, Blakemore AI. Human genes involved in copy number variation: mechanisms of origin, functional effects and implications for disease. *Cytogenet Genome Res*. 2008;123:17–26.
- Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science*. 2005;307:1434–40.
- Modi WS. CCL3L1 and CCL4L1 chemokine genes are located in a segmental duplication at chromosome 17q12. *Genomics*. 2004;83:735–8.
- Mamtani M, Rovin B, Brey R, Camargo JF, Kulkarni H, Herrera M, et al. CCL3L1 gene-containing segmental duplications and polymorphisms in CCR5 affect risk of systemic lupus erythematosus. *Ann Rheum Dis*. 2008;67:1076–83.
- Colobran R, Pedrosa E, Carretero-Iglesias L, Juan M. Copy number variation in chemokine superfamily: the complex scene of CCL3L-CCL4L genes in health and disease. *Clin Exp Immunol*. 2010;162:41–52.
- Nakao M, Nomiyama H, Shimada K. Structures of human genes coding for cytokine LD78 and their expression. *Mol Cell Biol*. 1990;10:3646–58.
- Carpenter D, McIntosh RS, Pleass RJ, Armour JAL. Functional effects of CCL3L1 copy number. *Genes Immun*. 2012;13:374–9.
- Kouno J, Nagai H, Nagahata T, Onda M, Yamaguchi H, Adachi K, et al. Up-regulation of CC chemokine, CCL3L1, and receptors, CCR3, CCR5 in human glioblastoma that promotes cell growth. *J Neuro-Oncol*. 2004;70:301–7.
- Colobran R, Comas D, Faner R, Pedrosa E, Anglada R, Pujol-Borrell R, et al. Population structure in copy number variation and SNPs in the CCL4L chemokine gene. *Genes Immun*. 2008;9:279–88.
- Nibbs RJB, Yang J, Landau NR, Mao JH, Graham GJ. LD78B, a non-allelic variant of human MIP-1a (LD78a), has enhanced receptor interactions and potent HIV suppressive activity. *J Biol Chem*. 1999;274:17478–83.
- Townson JR, Barcellos LF, Nibbs RJB. Gene copy number regulates the production of the human chemokine CCL3L1. *Eur J Immunol*. 2002;32:3016–26.
- Li J, Yang T, Wang L, Yan H, Zhang Y, Guo Y, et al. Whole genome distribution and ethnic differentiation of copy number variation in Caucasian and Asian populations. *PLoS One*. 2009;4(11):e7958.
- Carpenter D, Walker S, Prescott N, Schalkwijk J, Armour JAL. Accuracy and differential bias in copy number measurement of CCL3L1 in association studies with three auto-immune disorders. *BMC Genomics*. 2011;12:418.
- Walker S, Janyakhantikul S, Armour JAL. Multiplex Paralogue ratio tests for accurate measurement of multiallelic CNVs. *Genomics*. 2009;93:98–103.
- Shao W, Tang J, Song W, Wang C, Li Y, Wilson CM, et al. CCL3L1 and CCL4L1: variable gene copy number in adolescents with and without human immunodeficiency virus type 1 (HIV-1) infection. *Genes Immun*. 2007;8:224–31.
- Akillu E, Odenthal-Hesse L, Bowdrey J, Habtewold A, Ngaimisi E, Yimer G, et al. CCL3L1 copy number, HIV load, and immune reconstitution in sub-Saharan Africans. *BMC Infect Dis*. 2013;13:536.
- Li H, Xie HY, Zhou L, Feng XW, Wang WL, Liang TB, et al. Copy number variation in CCL3L1 gene is associated with susceptibility to acute rejection in patients after liver transplantation. *Clin Transpl*. 2012;26:314–21.
- Kim HE, Kim JJ, Han MK, Lee KY, Song MS, Lee HD, et al. Variations in the number of CCL3L1 gene copies and Kawasaki disease in Korean children. *Pediatr Cardiol*. 2012;33:1259–63.
- Nakajima T, Ohtani H, Naruse T, Shibata H, Mimaya JI, Terunuma H, et al. Copy number variations of CCL3L1 and long-term prognosis of HIV-1 infection in asymptomatic HIV-infected Japanese with hemophilia. *Immunogenetics*. 2007;59:793–8.
- Armour JAL, Palla R, Zeeuwen PLJM, Den Heijer M, Schalkwijk J, Hollox EJ. Accurate, high-throughput typing of copy number variation using paralogue ratios from dispersed repeats. *Nucleic Acids Res*. 2007;35(3):e19.
- Adewoye AB, Shrine N, Odenthal-Hesse L, Welsh S, Malarstig A, Jelicinsky S, et al. Human CCL3L1 copy number variation, gene expression, and the role of the CCL3L1-CCR5 axis in lung function. *Wellcome Open Res*. 2018;3:13.
- Nordang G, Carpenter D, Viken M, Kvien T, Armour J, Lie B. Association analysis of the CCL3L1 copy number locus by paralogue ratio test in Norwegian rheumatoid arthritis patients and healthy controls. *Genes Immun*. 2012;13:579–82.
- Carpenter D, Taype C, Goulding J, Levin M, Eley B, Anderson S, et al. CCL3L1 copy number, CCR5 genotype and susceptibility to tuberculosis. *BMC Med Genet*. 2014;15:5.
- Deng L, Hoh BP, Lu D, Fu R, Phipps ME, Li S, et al. The population genomic landscape of human genetic structure, admixture history and local adaptation in peninsular Malaysia. *Hum Genet*. 2014;133:1169–85.

35. Nakajima T, Kaur G, Mehra N, Kimura A. HIV-1/AIDS susceptibility and copy number variation in CCL3L1, a gene encoding a natural ligand for HIV-1 co-receptor CCR5. *Cytogenet Genome Res.* 2008;123:156–60.
36. Abdulla MA, Ahmed I, Assawamakin A, Bhak J, Brahmachari SK, Calacal GC, et al. Mapping human genetic diversity in Asia. *Science.* 2009;326:1541–5.
37. Fu R, Mokhtar SS, Phipps ME, Hoh BP, Xu S. A genome-wide characterization of copy number variations in native populations of peninsular Malaysia. *Eur J Hum Genet.* 2018:1–12.
38. Picton ACP, Paximadis M, Tiemessen CT. Contribution of variable CCL3L1 copy number to CCL3 protein production in two ethnically divergent south African populations. *Infect Genet Evol.* 2013;14:347–56.
39. Barbie J. Chinese culture and demographic trends in Thailand and Malaysia. *Chin J Popul Sci.* 1992;4:139–48.
40. Hatin WI, Nur-Shafawati AR, Etemad A, Jin W, Qin P, Xu S, et al. A genome wide pattern of population structure and admixture in peninsular Malaysia Malays. *HUGO J.* 2014;8:1–18.
41. Hatin WI, Nur-Shafawati AR, Zahri M-K, Xu S, Jin L, Tan S-G, et al. Population genetic structure of peninsular Malaysia Malay sub-ethnic groups. *PLoS One.* 2011;6(4):e18312.
42. Gan GG, Lee MTM, Subramaniam R, Lu LS, Tai MC, Phipps ME. Allele and genotype frequencies of VKORC1 -1639g > a polymorphism in three different ethnic groups in Malaysia. *Asia-Pacific J Mol Biol Biotechnol.* 2012; 20:19–23.
43. Burns JC, Shimizu C, Gonzalez E, Kulkarni H, Patel S, Shike H, et al. Genetic variations in the receptor-ligand pair CCR5 and CCL3L1 are important determinants of susceptibility to Kawasaki disease. *J Infect Dis.* 2005;192: 344–9.
44. Mamtani M, Mummidi S, Ramsuran V, Pham MH, Maldonado R, Begum K, et al. Influence of variations in CCL3L1 and CCR5 on tuberculosis in a northwestern Colombian population. *J Infect Dis.* 2011;203:1590–4.
45. Ahuja SK, Kulkarni H, Catano G, Agan BK, Camargo JF, He W, et al. CCL3L1-CCR5 genotype influences durability of immune recovery during antiretroviral therapy of HIV-1-infected individuals. *Nat Med.* 2009;14:413–20.
46. Ben, Kilani MS, Achour Y, Perea J, Cornelis F, Bardin T, Chaudru V, et al. Characterization of copy number variants for CCL3L1 gene in rheumatoid arthritis for French trio families and Tunisian cases and controls. *Clin Rheumatol.* 2016;35:1917–22.
47. Cantsilieris S, Western PS, Baird PN, White SJ. Technical considerations for genotyping multi-allelic copy number variation (CNV), in regions of segmental duplication. *BMC Genomics.* 2014;15:329.
48. Field SF, Howson JMM, Maier LM, Walker S, Walker NM, Smyth DJ, et al. Experimental aspects of copy number variant assays at CCL3L1. *Nat Med.* 2009;15(10):1115–7.
49. Haridan US, Mokhtar U, Machado LR, Aziz ATA, Shueb RH, Zaid M, et al. A comparison of assays for accurate copy number measurement of the low-affinity FC gamma receptor genes FCGR3A and FCGR3B. *PLoS One.* 2015;10:1–13.
50. Aldhous MC, Bakar SA, Prescott NJ, Palla R, Soo K, Mansfield JC, et al. Measurement methods and accuracy in copy number variation: failure to replicate associations of beta-defensin copy number with Crohn's disease. *Hum Mol Genet.* 2010;19:4930–8.
51. Hollox EJ, Armour JAL, Barber JCK. Extensive normal copy number variation of a beta-defensin antimicrobial-gene cluster. *Am J Hum Genet.* 2003;73: 591–600.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

