Human *CCL3L1* copy number variation, gene expression, and the role of the CCL3L1-CCR5 axis in lung function

Adeolu B Adewoye (1)*, Nick Shrine (2)*, Linda Odenthal-Hesse (1), Samantha Welsh (3), Anders Malarstig (4), Scott Jelinsky (5), Iain Kilty (5), Martin D Tobin (2,6), Edward J Hollox (1)†, Louise V Wain (2,6)†

1. Department of Genetics and Genome Biology, University of Leicester, Leicester, UK
2. Department of Health Sciences, University of Leicester, Leicester, UK
3. UK Biobank, Stockport, UK
4. Pfizer Worldwide Research and Development, Stockholm, Sweden.
5. Pfizer Worldwide Research and Development, Cambridge, Massachusetts, USA
6. National Institute of Health Research Biomedical Research Centre, University of Leicester, Leicester, UK

* † These authors contributed equally to this work.

Corresponding authors: Edward J Hollox (ejh33@leicester.ac.uk), Louise V Wain (lvw1@leicester.ac.uk)

Keywords: copy number variation, lung function, CCL3L1, CCR5, CNV, UK Biobank

Abstract

The CCL3L1-CCR5 signaling axis is important in a number of inflammatory responses, including macrophage function, and T-cell-dependent immune responses. Small molecule CCR5 antagonists exist, including the approved antiretroviral drug maraviroc, and therapeutic monoclonal antibodies are in development. Repositioning of drugs and targets into new disease areas can accelerate the availability of new therapies and substantially reduce costs. As it has been shown that drug targets with genetic evidence supporting their involvement in the disease are more likely to be successful in clinical development, using genetic association studies to identify new target repurposing opportunities could be fruitful. Here we investigate the potential of perturbation of the CCL3L1-CCR5 axis as treatment for respiratory disease. Europeans typically carry between 0 and 5 copies of *CCL3L1* and this multi-allelic variation is not detected by widely used genome-wide single nucleotide polymorphism studies. We directly measured the complex structural variation of *CCL3L1* using the Paralogue Ratio Test (PRT) and imputed (with validation) CCR5del32 genotypes in 5,000 individuals from UK Biobank, selected from the extremes of the lung function distribution, and analysed DNA and RNAseq data for *CCL3L1* from the 1000 Genomes Project. We confirmed the gene dosage effect of *CCL3L1* copy number on *CCL3L1* mRNA expression levels. We found no evidence for association of *CCL3L1* copy number or CCR5del32 genotype with lung function suggesting that repositioning CCR5 antagonists is unlikely to be successful for the treatment of airflow obstruction.

Introduction

Genome-wide association studies have identified thousands of disease-gene associations leading to new disease insight and potential new approaches to treatment. It has been shown that drug targets supported by genetic studies have an increased chance of success in clinical development [1]. Even so, only a subset of candidate drugs will make it through to the clinic. Identifying opportunities for repositioning existing drugs and targets is therefore an appealing prospect and using genetic studies to define alternative indications for an already-approved drug is a promising approach.

The Mip1alpha (encoded by *CCL3* and *CCL3L1*)-CCR5 signaling axis is important in a number of inflammatory responses, including macrophage function, and T-cell-dependent immune responses [2]. It is perturbed by CCR5 antagonists such as Pfizer's maraviroc, the only CCR5 antagonist to be approved by the United States Food and Drug Administration [3, 4]. Identification of a genetic association of variants within the genes involved (*CCR5* and *CCL3*/*CCL3L1*) would strongly support the potential use of CCR5 antagonists in the treatment of respiratory conditions [5].

In mice, MIP1alpha is implicated in virus-mediated inflammation of the lung, pulmonary eosinophilia following paramyxovirus infection, clearance of pulmonary infections [6, 7], and in the response to respiratory syncytial virus infection [8-10]. In humans Mip1apha controls the recruitment of immune cells to inflammatory foci, and increased levels of Mip1alpha mRNA are found in bronchial epithelial cells of COPD patients [11], and increased protein levels in the sputum of COPD patients [12] where increased macrophage and neutrophil infiltration in the lung is a key pathology.

The *CCR5* gene in humans has a 32bp exonic deletion allele (rs333, *CCR5*d32) with a minor allele frequency of between 5-15% in Europeans [13]. This allele causes a translational frameshift and abrogates expression of the receptor at the cell surface, such that homozygotes for the deletion allele lack any functional CCR5 receptor [14, 15]. This variant has been strongly and repeatedly associated with resistance to HIV infection and slower HIV progression, as CCR5 is a common coreceptor for HIV entry into T-lymphocytes [16]. The CCR5d32 allele has been suggested to confer a reduced risk of asthma in children in one study [17] although this has not been replicated [18, 19].

In humans, there are two isoforms of Mip1alpha, the LD78a isoform encoded by the CCL3 gene and the LD78b isoform encoded by the paralogous *CCL3L1* gene [20, 21]. The two isoforms differ by three amino acids, but only one of these small changes, a serine to proline change at position 2 of the mature protein, alters the affinity to the cell surface receptor CCR5, with the beta isoform (*CCL3L1*) having approximately six-fold greater affinity [22] for CCR5 than the alpha isoform (*CCL3*).

The *CCL3L1* gene is part of a complex structurally variable region, although the *CCL3* gene is not. The *CCL3L1* gene and the neighboring *CCL4L1* gene are tandemly repeated with the total diploid

copy number ranging from 0 copies to 6 copies in Europeans [23, 24]. Higher copy numbers are observed elsewhere, for example 10 in Tanzanians [25] and 14 in Ethiopians [26]. Previous studies have shown evidence of a gene dosage effect, with *CCL3L1* gene dose reflected in mRNA levels as well as in the ability to chemoattract monocytes [27, 28].

Measuring *CCL3L1* multiallelic copy number variation has been challenging [29]. Early studies used qPCR assays with a low signal:noise ratio [23, 30, 31], but assays based on the paralogue ratio test (PRT), allowed more accurate estimation of diploid copy number [24, 32]. Because of the challenges in measuring *CCL3L1* copy number in sufficiently large and well-powered sample sizes, the effect of structural variation of the genes encoding the Mip1alpha-CCR5 ligand-receptor pair has not been adequately explored.

In this study, we set out to confirm previous reports that *CCL3L1* copy number is associated with CCL3L1 gene expression, then measure *CCL3L1* copy number and *CCR5d32* genotype in 5000 individuals from UK Biobank, and finally test for association with lung function. Furthermore, we validated our copy number typing approach and observed copy number frequencies using publicly available sequence data from the 1000 Genomes Project. For *CCL3L1* copy number measurement in the 5000 individuals from UK Biobank, we used a triplex paralogue ratio test (PRT) which is considered to be the gold standard approach for measurement of this copy number variation [24, 29]. For genotyping of *CCR5*d32 in UK Biobank, we used a standard genotype imputation approach with additional PCR validation. We tested for association with extremes of Forced Expired Volume in 1 second (FEV_1) as a binary trait. This study is the largest analysis of the effect of CCL3L1 copy number and CCR5d32 genotypes on lung function undertaken to date.

Methods

Sample selection

Individuals were selected from the UK BiLEVE [33, 34] subset of UK Biobank. In brief, 502,682 individuals were recruited to UK Biobank of whom 275,939 were of self-reported European-ancestry, and had two or more measures of Forced Expired Volume in 1s (FEV_1) and Forced Vital Capacity (FVC) measures (Vitalograph Pneumotrac 6800, Buckingham, UK) passing ATS/ERS criteria [35]. Based on the highest available FEV_1 measurement, 50,008 individuals with extreme low (n=10,002), near-average (n=10,000) and extreme high (n=5,002) % predicted FEV_1 were selected from amongst never-smokers (total n=105,272) and heavy-smokers (mean 35 pack-years of smoking, total n=46,758), separately. For this study, we selected 2500 age-matched European-ancestry heavy smokers from the extreme high and extreme low % predicted FEV_1 subsets defined for the UK BiLEVE study (Figure 1, Table 1). DNA samples for these 5000 individuals were prepared by UK Biobank and provided back to the University of Leicester with new identification codes such that typing of *CCL3L1* copy number and *CCR5*d32 was blinded to lung function status. Positive control samples for the copy number typing were from the Human Random Control panel from Public Health England.

CCL3L1 copy number estimation in UK Biobank and 1000 Genomes Project samples using the paralogue ratio test (PRT)

CCL3L1 copy number was determined using a triplex paralogue ratio test (PRT) assay as used previously [24, 26]. Briefly, PRT is a comparative PCR method that amplifies a test and reference locus using the same pair of primers, followed by capillary electrophoresis and quantification of the two products [32, 36]. The triplex assay produced three independent estimates of copy number per test, of which the average was taken as a representative copy number value. The three values were consistent in 95% of samples, however, for 5% of samples the value from the LTR61A PRT assay was significantly lower than the other two PRT values, and an average of the two consistent PRTs was taken in these 5% of samples. For each typing experiment, 4 positive controls of known copy number were also included, as previously [26, 37]. The copy number values clustered about integer copy numbers, and a Gaussian mixture model was fitted to allow assignment of individuals to an integer copy number call using CNVtools [38]. For the 5000 individuals from UK Biobank, 58 individuals were selected by UK Biobank investigators as blind spiked duplicates as part of the quality control check to ensure genotyping accuracy. Copy numbers from UK Biobank samples are available from UK Biobank at http://www.ukbiobank.ac.uk/data-showcase/.

Gene Expression levels in 1000 genomes project lymphoblastoid cell lines

Matched RNAseq data that is publically available for the 1000 genomes samples were grouped based on *CCL3L1* copy number and analysed for their differential expression using Cufflinks v2.1.1 [39]. This allows measurement of the effect of genomic copy number of *CCL3L1* on gene expression levels. The analyses were all performed on ALICE High Performance Computing Facility at the University of Leicester. The RNAseq data were generated by (Lappalainen et al. 2013) and deposited in EBI ArrayExpress (accessions E-GEUV-1, E-GEUV-2, E-GEUV-3). Using Cufflinks, the fragments per kilobase of transcript per million fragments mapped (FPKM) values were estimated by applying a statistical model that normalises the mapped reads by length and

their abundance. Briefly, the fragment reads are divided by transcript size and the total number of reads and then adjusted to 1 kb and 1 million reads.

Genotyping of CCR5d32 polymorphism

Imputation to 1000 Genomes Project Phase 1+UK10K reference panel [40] and PCR were used to genotype the *CCR5*del32 polymorphism (rs333) in the 5000 UK Biobank individuals. Phasing and imputation were undertaken with SHAPEIT v2.r790 [41] and IMPUTE2 v2.3.1 [42]. For individuals with imputation posterior probability <0.95 (431 samples), and an additional 20 samples that were imputed as homozygous for the minor del32 allele, we validated the imputation results using direct PCR genotyping. Duplicates of a random selection of 28 of individuals were included as a quality control check for genotyping reproducibility (genotyping was also blinded to duplicate status). Genotypes from UK Biobank samples are available from UK Biobank at http://www.ukbiobank.ac.uk/data-showcase/.

CCL3L1 copy number estimation from sequencing data for 1000 Genomes Project individuals

1000 genomes phase 3 whole genome aligned Bam files generated from Illumina platforms available from the European Bioinformatics Institute (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/) were downloaded and the genomic region including *CCL3L1* (hg19:chr17:33670000-35670000) was analysed using CNVrd2 [43]. Using 500bp window sequence read depth, the sequence read depth was calculated across the region for all 2502 genomes from 26 populations, and standard deviation/quantile calculated for each window. The segmentation scores obtained from this analysis were clustered into different groups using a Gaussian mixture model. A prior information for all populations was estimated using the expectation maximisation (EM) algorithm on a population group with clear clusters of segmentation scores. The prior information (means, standard deviations and proportions of the mixture components) was fed into Bayesian model to infer *CCL3L1* integer copy number in all populations. Copy number estimates are available from dbVar (https://www.ncbi.nlm.nih.gov/dbvar) under study accession number nstd155.

Association analysis

We tested for association of *CCL3L1* copy number and CCR5d32 genotype separately with lung function extremes (as a binary trait) using logistic regression with pack-years of smoking and the first ten principal components (obtained previously using full genome-wide SNP genotyping data to adjust for fine-scale population structure as covariates [33]. For *CCR5*d32, a genotypic genetic model was assumed for the primary analysis. We then fitted a full linear regression model that included *CCR5*d32 genotype (genotypic mode), *CCL3L1* copy number, pack years, 10 principal components and a term for the interaction of *CCR5d32* and *CCL3L1*.

Results

Using CNVrd2, we typed *CCL3L1* copy number from whole genome sequence alignments for 2502 individuals from the 1000 Genomes project (Figure 2a). The data were grouped into large superpopulations, as defined by the 1000 Genomes Project [44], and our analysis confirmed previous observations that Europeans have the lowest *CCL3L1* diploid copy number, ranging between 0 and 5 with a mean copy number of 1.97, and sub-Saharan Africans have the highest diploid copy number, ranging between 1 and 9 with a mean of 4.19, which is more than twice as high as Europeans (Table 2)[24, 25].

For 144 individuals from the CEU (n=96) and YRI (n=48) populations of the 1000 Genomes project, we also determined *CCL3L1* copy number using the PRT approach (Figure 2b). There was strong concordance between results, with discrete clusters of raw data, representing individual integer copy numbers, formed, particularly at low copy number. For the range seen in Europeans (copy numbers 0 to 5), there are seven clear discrepancies which gives an joint error rate of 5%.

To confirm previous studies that reported an association between *CCL3L1* copy number and *CCL3L1* mRNA levels, we compared the 1000 Genomes Project *CCL3L1* copy numbers with transcript levels of *CCL3L1* and its non-copy number variable paralogue *CCL3*, as generated by RNAseq of the corresponding B-lymphoblastoid cell lines (Figures 3a, 3b). Comparison with transcript level estimates using RNAseq data showed a clear positive correlation between *CCL3L1* copy number and expression level (figure 3b, r^2=0.25, $p<2x10^{-16}$). We used the specific sequence changes between *CCL3L* and *CCL3* to distinguish transcripts from either gene, and confirmed this by showing that *CCL3* expression has no relationship with *CCL3L1* copy number (figure 3a, r^2=0.006, p=0.087), as well as showing that individuals with zero copies of *CCL3L1* show no transcripts from *CCL3L1* (figure 3b).

We confirmed an increase of one to two orders of magnitude for *CCL3* transcript levels compared to *CCL3L1* transcript levels in B-lymphoblast cells. Following normalization of the *CCL3L1* expression levels to CCL3 expression levels, we show that *CCL3L1* transcript levels are closely correlated with gene copy number (Figure 3c, r^2=0.5, $p<2x10^{-16}$).

Having confirmed a relationship between gene copy number and transcript levels of *CCL3L1,* we investigated the relationships between CCL3L1 copy numbers, CCR5d32 genotype and lung function in individuals selected from the extremes of the lung function distribution in UK Biobank. We typed 5000 UK Biobank samples using PRT, with 19 failures. The results showed a clear mixture of Gaussian distributions centered on each integer copy number (Figure 2c). All 58 duplicates were consistently typed, resulting in an error rate between 0% and 4.7%. We observed clear distances between the clusters, further suggesting that the measurement error rate for this cohort is likely to be low.

We estimated *CCL3L1* integer copy numbers in all the samples using Gaussian mixture modelling (Table 3). The copy number range was consistent with previous observations in UK

population [24], and with our estimation from the 1000 Genomes project samples. The two copy genotype was the most frequent with a frequency of 0.563. The *CCL3L1* zero copy null genotype is uncommon, with a frequency of 2.5% in the UK. 4993 of the 5000 UK Biobank samples were genotyped for *CCR5*d32 by imputation with the genotypes for 474 individuals validated using direct PCR analysis. There was no evidence that the genotype frequencies departed from Hardy-Weinberg equilibrium (chi-squared test, p=0.35) and the observed *CCR5*d32 deletion allele frequency was 0.11, consistent with previous estimates [13].

A total of 4975 UK Biobank individuals had both *CCL3L1* copy number and *CCR5*d32 genotypes measured (2486 high and 2489 low FEV_1, Table 4). There was no evidence of an association between *CCL3L1* copy number and *CCR5*d32 genotype (chi-squared test p=0.803).

We fitted a full model with both CCR5 genotypes (genotypic model) and *CCL3L1* copy number and an interaction term as described above. This was undertaken in order to identify whether particular combinations of *CCL3L1* copy number and *CCR5*d32 genotype were differentially associated with lung function. Pack years of smoking and 10 principal components were included as covariates. No associations were significant (Table 5).

Discussion

Our study provides robust large-scale confirmation of a gene dosage effect of CCL3L1 copy number on CCL3L1 mRNA levels, and also emphasises the strong dependence of *CCL3L1*:CCL3 mRNA ratio on copy number, with *CCL3L1* copy number accounting for 50% of total variation. Although it is clear that *CCL3L1* is expressed at much lower levels than *CCL3*, the MIP1alpha isoform encoded by *CCL3L1* (LD78beta) has a much stronger affinity to the CCR5 receptor than MIP1alpha isoform *CCL3* (LD78alpha). It therefore seems likely that the *CCL3L1* copy number variation mediates a biological effect *in vivo*. It should be noted that the expression data are from transformed B-lymphoblastoid cell lines, but a gene dosage effect is consistent with a study using fresh monocytes from 55 different individuals stimulated with bacterial lipopolysccharide [28].

Our analysis provides evidence that there is no effect of either *CCL3L1* copy number or *CCR5*d32 genotype, or any combinations of genotypes at the two loci, on lung function. This suggests that, although the Mip1alpha-CCR5 signaling axis can be disrupted by artificial CCR5 antagonists, there is no evidence that this axis has a functional effect on lung function and that development of new drugs to target this axis, or repurposing of existing drugs, might be of little or no therapeutic benefit in treating COPD.

We analysed approximately 5000 individuals. Whilst this represents a large sample size for labour-intensive PRT assays, it is a modest sample size in comparison with those employed in GWAS. That said, power was boosted by selecting from the extremes of the lung function distribution in the very large (n~500K) UK Biobank.

We reported PRT error rates of 2.5% for the 144 1000 Genomes Project samples and between 0% and 4.75% for the 4981 UK Biobank participants. A previous study using this PRT approach estimated an error rate of less than 0.1% [24], which suggests that much of the joint error rate for the PRT and sequence read depth could be due to errors in the sequence read depth approach.

The exact boundaries of the *CCL3L1* CNV have yet to be determined with precision but it is known to include the *CCL4L1* gene which encodes MIP1β [24]. The human genome assembly GRCh38 shows a single copy *CCL3L1*/*CCL4L1* repeat unit, and also includes the *TBC1D3* gene, encoding TBC1 Domain Family Member 3 [45-47]. The GRCh38 alternative assembly chr17_KI270909v1_alt shows two repeat units, both including *TBC1D3*. However an earlier assembly shows a complete contig with two repeat units carrying *CCL3L1*/*CCL4L1*, only one of which carries *TBC1D3*. ArrayCGH and fiber-FISH both confirm this is real heterogeneity by showing that the *TBC1D3* gene is included in some, but not all, tandemly repeated units in some individuals, together with *CCL3L1* and *CCL4L1* [26, 48]. Throughout this paper, and in most of the literature, *CCL3L1* CNV is used as a shorthand to describe the CNV of this complex repeat unit.

Given the gene content of this repeat unit, we would expect a gene dosage effect for *CCL4L1* and *TBC1D3*, in addition to *CCL3L1*, but this has not yet been confirmed. Our data do, however, show

no effect of *CCL3L1* copy number on expression levels of its close paralogue, *CCL3*, which is immediately proximal to the CNV. This difference shows that the considerable variation in genome structure distal to the *CCL3* gene does not affect overall levels of *CCL3* expression.

In summary, we selected individuals from the extremes of the lung function distribution of a very large general population cohort. We found no association of *CCL3L1* copy number, nor of the CCR5d32 variant with lung function, as defined by FEV_1.

Acknowledgements

This research has been conducted using the UK Biobank Resource under Application Number 7140. This work was supported by funding from Pfizer, Inc to MDT, LVW and EJH. This article presents independent research funded partially by the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. Single nucleotide polymorphism genotyping of the UK BiLEVE subset of UK Biobank was funded by a Medical Research Council (MRC) strategic award to MDT, IPH, DPS and LVW (MC_PC_12010) (UK Biobank Application Number 648). LVW holds a GSK / British Lung Foundation Chair in Respiratory Research. MDT holds a Wellcome Trust Investigator Award. This research used the ALICE and SPECTRE High Performance Computing Facilities at the University of Leicester. We would like to thank Gurdeep Lall and Amelia Veselis for technical support.

328

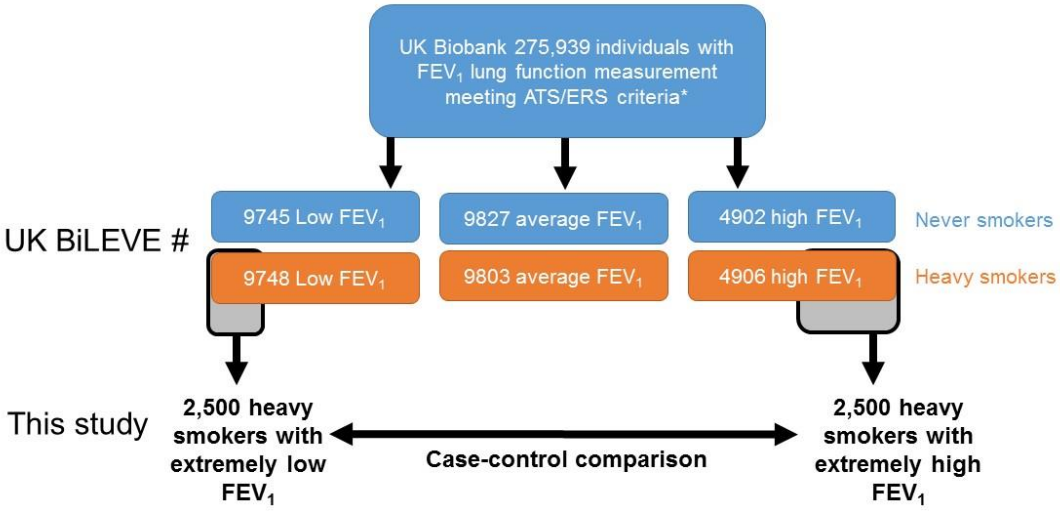
329 **Figure 1 – Study design**

330 FEV_1 is percent predicted FEV_1.

331 *Lung function measurement quality control defined previously [33]

332 # Final numbers after quality control [33]

333



334

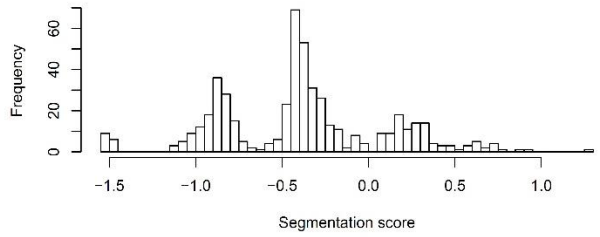
Figure 2 – *CCL3L1* Copy number typing

a) Histogram of raw copy number estimates of 1000 Genomes Project samples from sequence read depth represented as segmentation scores on the x axis, generated by CNVrd2, with higher scores reflecting higher copy number.

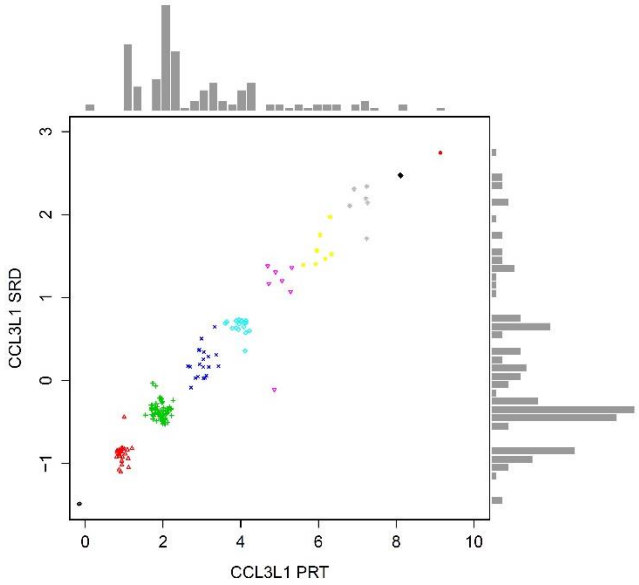
b) Validation of 144 1000 Genomes Project samples using PRT (x axis) against estimates made from sequence read depth. Colours/symbols in the scatterplot represent different integer copy numbers inferred from PRT clusters.

c) Histogram of raw copy number estimates using PRT for the UK Biobank cohort.

a)



b)



c)

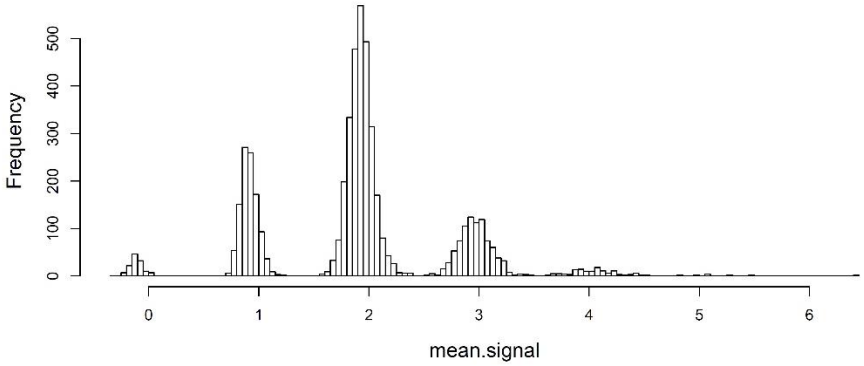


Figure 3 – Copy number and expression level of *CCL3L1* and *CCL3* in lymphoblastoid cell lines

a) *CCL3* mRNA level (FPKM units) across different *CCL3L1* copy numbers.
b) *CCL3L1* mRNA level (FPKM units) across different *CCL3L1* copy numbers.
c) *CCL3L1*:*CCL3* mRNA ratio across different *CCL3L1* copy numbers.

Individual data points are shown, with red bars indicating median and interquartile ranges.

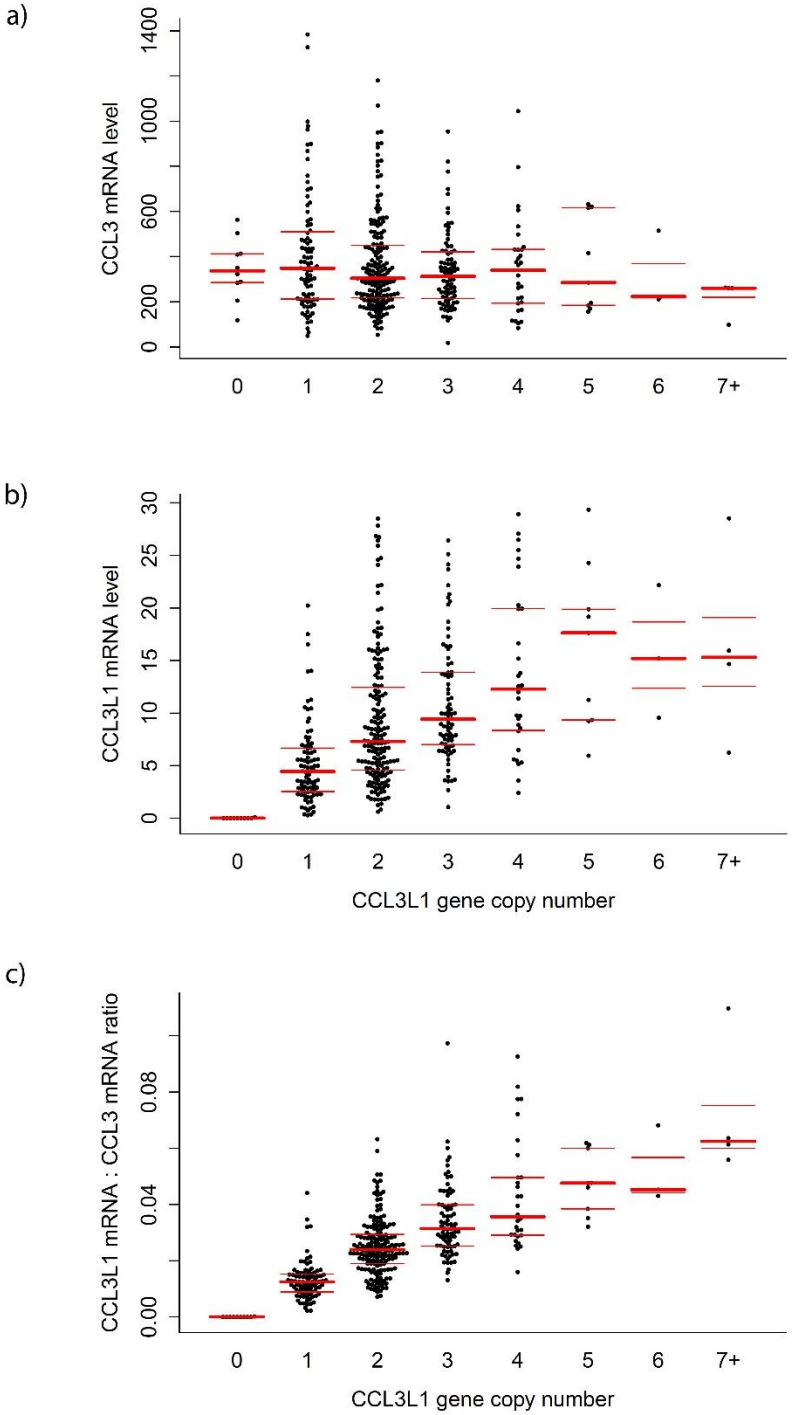


Table 1 – Demographics of selected UK Biobank cohort

	Low FEV_1 (n=2500)	High FEV_1 (n=2500)
n (%) male	1250 (50%)	1250 (50%)
Age	56.9 / 7.9 (40, 70)	56.9 / 7.9 (40, 70)
Pack-years	40.6 / 22.5 (10.8, 301.0)	29.37 / 13.4 (10.5, 134.0)
Pack-years as a proportion of lifespan.	0.96 / 0.47 (0.42, 7.00)	0.70 / 0.29 (0.42, 3.03)
FEV_1 (litres)	1.50 / 0.47 (0.36, 3.38)	3.64 / 0.73 (2.02, 6.72)
Percent predicted FEV_1	51.4 / 11.0 (14.9, 74.5)	123.3 / 8.2 (112.8, 205.7)

Values are Mean / SD (range), unless stated.

Table 2 -- *CCL3L1* copy number frequency distributions in 1000 Genomes data

superpopulation	n	average copy number	minimum copy number	maximum copy number
AFR (Sub-Saharan African)	661	4.19	1	9
AMR (Admixed American)	347	2.71	0	8
EAS (East Asian)	504	3.52	0	9
EUR (European)	501	1.97	0	5
SAS (South Asian)	489	2.39	0	7

374 **Table 3 *CCL3L1* copy number counts in UK Biobank data**
375

CCL3L1 diploid copy number	Number of samples	Frequency
0	127	0.025
1	1046	0.210
2	2806	0.563
3	853	0.171
4	128	0.026
5	21	0.004
Sum	4981	0.999

376
377
378 **Table 4 CCR5d32 genotype counts by CCL3L1 copy number in UK Biobank data**
379

	CCR5d32 genotype		
CCL3L1 copy number	ref/ref	del32/ref	del32/del32
0	92	33	2
1	826	203	16
2	2197	574	31
3	662	181	9
4	99	28	1
5	15	6	0
Sum	3891	1025	59

380
381
382

Table 5 - Association analysis of *CCR5* genotype and *CCL3L1* copy number with high vs low FEV_1

	OR (95% CI)	P value
CCR5d32 deletion heterozygote main effect	0.84 (0.57-1.23)	0.38
CCR5d32 deletion homozygote main effect	0.29 (0.07-1.30)	0.11
CCL3L1 copy number main effect	1.00 (0.92-1.09)	0.97
CCR5d32 deletion heterozygote interaction with CCL3L1 copy number	1.11 (0.93-1.32)	0.27
CCR5d32 deletion homozygote interaction with CCL3L1 copy number	1.74 (0.83-3.64)	0.14

2486 samples with high FEV_1 and 2489 samples with low FEV_1
Covariates: smoking pack-years, 10 principal components of SNP genetic variation.

References

1. Nelson MR, Johnson T, Warren L, Hughes AR, Chissoe SL, Xu CF, et al. The genetics of drug efficacy: opportunities and challenges. Nat Rev Genet. 2016;17(4):197-206. Epub 2016/03/15. doi: 10.1038/nrg.2016.12. PubMed PMID: 26972588.
2. Menten P, Wuyts A, Van Damme J. Macrophage inflammatory protein-1. Cytokine Growth Factor Rev. 2002;13(6):455-81. Epub 2002/10/29. PubMed PMID: 12401480.
3. Carter PH. Chemokine receptor antagonism as an approach to anti-inflammatory therapy: 'just right' or plain wrong? Current Opinion in Chemical Biology. 2002;6(4):510-25. doi: https://doi.org/10.1016/S1367-5931(02)00351-4.
4. Lieberman-Blum SS, Fung HB, Bandres JC. Maraviroc: a CCR5-receptor antagonist for the treatment of HIV-1 infection. Clin Ther. 2008;30(7):1228-50. Epub 2008/08/12. PubMed PMID: 18691983.
5. Koelink PJ, Overbeek SA, Braber S, de Kruijf P, Folkerts G, Smit MJ, et al. Targeting chemokine receptors in chronic inflammatory diseases: an extensive review. Pharmacol Ther. 2012;133(1):1-18. Epub 2011/08/16. doi: 10.1016/j.pharmthera.2011.06.008. PubMed PMID: 21839114.
6. Cook DN, Beck MA, Coffman TM, Kirby SL, Sheridan JF, Pragnell IB, et al. Requirement of MIP-1 alpha for an inflammatory response to viral infection. Science. 1995;269(5230):1583-5. Epub 1995/09/15. PubMed PMID: 7667639.
7. Lindell DM, Standiford TJ, Mancuso P, Leshen ZJ, Huffnagle GB. Macrophage inflammatory protein 1alpha/CCL3 is required for clearance of an acute Klebsiella pneumoniae

pulmonary infection. Infect Immun. 2001;69(10):6364-9. Epub 2001/09/13. doi: 10.1128/iai.69.10.6364-6369.2001. PubMed PMID: 11553580; PubMed Central PMCID: PMCPMC98771.

8. Domachowske JB, Bonville CA, Gao JL, Murphy PM, Easton AJ, Rosenberg HF. MIP-1alpha is produced but it does not control pulmonary inflammation in response to respiratory syncytial virus infection in mice. Cell Immunol. 2000;206(1):1-6. Epub 2001/02/13. doi: 10.1006/cimm.2000.1730. PubMed PMID: 11161432.

9. Haeberle HA, Kuziel WA, Dieterich HJ, Casola A, Gatalica Z, Garofalo RP. Inducible expression of inflammatory chemokines in respiratory syncytial virus-infected mice: role of MIP-1alpha in lung pathology. J Virol. 2001;75(2):878-90. Epub 2001/01/03. doi: 10.1128/jvi.75.2.878-890.2001. PubMed PMID: 11134301; PubMed Central PMCID: PMCPMC113984.

10. Tregoning JS, Pribul PK, Pennycook AM, Hussell T, Wang B, Lukacs N, et al. The chemokine MIP1alpha/CCL3 determines pathology in primary RSV infection by regulating the balance of T cell populations in the murine lung. PLoS One. 2010;5(2):e9381. Epub 2010/03/03. doi: 10.1371/journal.pone.0009381. PubMed PMID: 20195359; PubMed Central PMCID: PMCPMC2827540.

11. Fuke S, Betsuyaku T, Nasuhara Y, Morikawa T, Katoh H, Nishimura M. Chemokines in bronchiolar epithelium in the development of chronic obstructive pulmonary disease. Am J Respir Cell Mol Biol. 2004;31(4):405-12. Epub 2004/06/29. doi: 10.1165/rcmb.2004-0131OC. PubMed PMID: 15220136.

12. Ravi AK, Khurana S, Lemon J, Plumb J, Booth G, Healy L, et al. Increased levels of soluble interleukin-6 receptor and CCL3 in COPD sputum. Respir Res. 2014;15:103. Epub 2014/09/04. doi: 10.1186/s12931-014-0103-4. PubMed PMID: 25183374; PubMed Central PMCID: PMCPMC4156958.

13. Martinson JJ, Chapman NH, Rees DC, Liu YT, Clegg JB. Global distribution of the CCR5 gene 32-basepair deletion. Nat Genet. 1997;16(1):100-3. Epub 1997/05/01. doi: 10.1038/ng0597-100. PubMed PMID: 9140404.

14. Liu R, Paxton WA, Choe S, Ceradini D, Martin SR, Horuk R, et al. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. Cell. 1996;86(3):367-77. Epub 1996/08/09. PubMed PMID: 8756719.

15. Samson M, Libert F, Doranz BJ, Rucker J, Liesnard C, Farber CM, et al. Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature. 1996;382(6593):722-5. Epub 1996/08/22. doi: 10.1038/382722a0. PubMed PMID: 8751444.

16. Naranbhai V, Carrington M. Host genetic variation and HIV disease: from mapping to mechanism. Immunogenetics. 2017;69(8-9):489-98. Epub 2017/07/12. doi: 10.1007/s00251-017-1000-z. PubMed PMID: 28695282; PubMed Central PMCID: PMCPMC5537324.

17. Hall IP, Wheatley A, Christie G, McDougall C, Hubbard R, Helms PJ. Association of CCR5 delta32 with reduced risk of asthma. Lancet. 1999;354(9186):1264-5. Epub 1999/10/16. PubMed PMID: 10520641.

18. Mitchell TJ, Walley AJ, Pease JE, Venables PJ, Wiltshire S, Williams TJ, et al. Delta 32 deletion of CCR5 gene and association with asthma or atopy. Lancet. 2000;356(9240):1491-2. Epub 2000/11/18. doi: 10.1016/s0140-6736(00)03144-5. PubMed PMID: 11081537.

19. Song GG, Kim JH, Lee YH. The chemokine receptor 5 delta32 polymorphism and type 1 diabetes, Behcet's disease, and asthma: a meta-analysis. Immunol Invest. 2014;43(2):123-36. Epub 2013/11/01. doi: 10.3109/08820139.2013.847457. PubMed PMID: 24171669.
20. Irving SG, Zipfel PF, Balke J, McBride OW, Morton CC, Burd PR, et al. Two inflammatory mediator cytokine genes are closely linked and variably amplified on chromosome 17q. Nucleic Acids Res. 1990;18(11):3261-70. Epub 1990/06/11. PubMed PMID: 1972563; PubMed Central PMCID: PMCPMC330932.
21. Nakao M, Nomiyama H, Shimada K. Structures of human genes coding for cytokine LD78 and their expression. Mol Cell Biol. 1990;10(7):3646-58. Epub 1990/07/01. PubMed PMID: 1694014; PubMed Central PMCID: PMCPMC360801.
22. Nibbs RJ, Yang J, Landau NR, Mao JH, Graham GJ. LD78beta, a non-allelic variant of human MIP-1alpha (LD78alpha), has enhanced receptor interactions and potent HIV suppressive activity. J Biol Chem. 1999;274(25):17478-83. Epub 1999/06/11. PubMed PMID: 10364178.
23. Field SF, Howson JM, 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. Epub 2009/10/09. doi: 10.1038/nm1009-1115. PubMed PMID: 19812562; PubMed Central PMCID: PMC2873561.
24. Walker S, Janyakhantikul S, Armour JA. Multiplex Paralogue Ratio Tests for accurate measurement of multiallelic CNVs. Genomics. 2009;93(1):98-103.
25. Carpenter D, Färnert A, Rooth I, Armour JA, Shaw M-A. < i> CCL3L1</i> copy number and susceptibility to malaria. Infection, Genetics and Evolution. 2012;(12):1147-54.
26. Aklillu 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 infectious diseases. 2013;13(1):536.
27. Townson JR, Barcellos LF, Nibbs RJ. Gene copy number regulates the production of the human chemokine CCL3-L1. European journal of immunology. 2002;32(10):3016-26.
28. Carpenter D, McIntosh R, Pleass R, Armour J. Functional effects of CCL3L1 copy number. Genes and Immunity. 2012;13(5):374-9.
29. 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(1):329.
30. 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(5714):1434-40.
31. Urban TJ, Weintrob AC, Fellay J, Colombo S, Shianna KV, Gumbs C, et al. CCL3L1 and HIV/AIDS susceptibility. Nat Med. 2009;15(10):1110-2. doi: http://www.nature.com/nm/journal/v15/n10/suppinfo/nm1009-1110_S1.html.
32. 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 research. 2007;35(3):e19-e.
33. Wain LV, Shrine N, Miller S, Jackson VE, Ntalla I, Soler Artigas M, et al. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. Lancet Respir Med. 2015;3(10):769-81.

Epub 2015/10/02. doi: 10.1016/s2213-2600(15)00283-0. PubMed PMID: 26423011; PubMed Central PMCID: PMCPMC4593935.

34. Wain LV, Shrine N, Artigas MS, Erzurumluoglu AM, Noyvert B, Bossini-Castillo L, et al. Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. Nat Genet. 2017;49(3):416-25. Epub 2017/02/07. doi: 10.1038/ng.3787. PubMed PMID: 28166213; PubMed Central PMCID: PMCPMC5326681.

35. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, et al. Standardisation of spirometry. Eur Respir J. 2005;26(2):319-38. Epub 2005/08/02. doi: 10.1183/09031936.05.00034805. PubMed PMID: 16055882.

36. Hollox EJ. Analysis of Copy Number Variation Using the Paralogue Ratio Test (PRT). Methods in molecular biology (Clifton, NJ). 2017;1492:127-46. Epub 2016/11/09. doi: 10.1007/978-1-4939-6442-0_8. PubMed PMID: 27822860.

37. Carpenter D, Taype C, Goulding J, Levin M, Eley B, Anderson S, et al. CCL3L1 copy number, CCR5 genotype and susceptibility to tuberculosis. BMC medical genetics. 2014;15(1):5.

38. Barnes C, Plagnol V, Fitzgerald T, Redon R, Marchini J, Clayton D, et al. A robust statistical method for case-control association testing with copy number variation. Nature Genetics. 2008;40(10):1245-52.

39. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012;7(3):562-78. Epub 2012/03/03. doi: 10.1038/nprot.2012.016. PubMed PMID: 22383036; PubMed Central PMCID: PMCPMC3334321.

40. Huang J, Howie B, McCarthy S, Memari Y, Walter K, Min JL, et al. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. Nat Commun. 2015;6:8111. Epub 2015/09/15. doi: 10.1038/ncomms9111. PubMed PMID: 26368830; PubMed Central PMCID: PMCPMC4579394.

41. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods. 2013;10(1):5-6. Epub 2012/12/28. doi: 10.1038/nmeth.2307. PubMed PMID: 23269371.

42. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009;5(6):e1000529. Epub 2009/06/23. doi: 10.1371/journal.pgen.1000529. PubMed PMID: 19543373; PubMed Central PMCID: PMCPMC2689936.

43. Nguyen HT, Merriman TR, Black MA. The CNVrd2 package: measurement of copy number at complex loci using high-throughput sequencing data. Front Genet. 2014;5:248. Epub 2014/08/20. doi: 10.3389/fgene.2014.00248. PubMed PMID: 25136349; PubMed Central PMCID: PMCPMC4117933.

44. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68-74. Epub 2015/10/04. doi: 10.1038/nature15393. PubMed PMID: 26432245; PubMed Central PMCID: PMCPMC4750478.

45. Hodzic D, Kong C, Wainszelbaum MJ, Charron AJ, Su X, Stahl PD. TBC1D3, a hominoid oncoprotein, is encoded by a cluster of paralogues located on chromosome 17q12. Genomics.

2006;88(6):731-6. Epub 2006/07/26. doi: 10.1016/j.ygeno.2006.05.009. PubMed PMID: 16863688.

46. Wainszelbaum MJ, Charron AJ, Kong C, Kirkpatrick DS, Srikanth P, Barbieri MA, et al. The hominoid-specific oncogene TBC1D3 activates Ras and modulates epidermal growth factor receptor signaling and trafficking. J Biol Chem. 2008;283(19):13233-42. Epub 2008/03/06. doi: 10.1074/jbc.M800234200. PubMed PMID: 18319245; PubMed Central PMCID: PMCPMC2442359.

47. Frittoli E, Palamidessi A, Pizzigoni A, Lanzetti L, Garre M, Troglio F, et al. The primate-specific protein TBC1D3 is required for optimal macropinocytosis in a novel ARF6-dependent pathway. Mol Biol Cell. 2008;19(4):1304-16. Epub 2008/01/18. doi: 10.1091/mbc.E07-06-0594. PubMed PMID: 18199687; PubMed Central PMCID: PMCPMC2291429.

48. Perry GH, Yang F, Marques-Bonet T, Murphy C, Fitzgerald T, Lee AS, et al. Copy number variation and evolution in humans and chimpanzees. Genome Research. 2008;18(11):1698-710.