Inferring mechanisms of copy number change from haplotype structures at the human \textit{DEFA1A3} locus

Holly A Black, Fayeza F Khan, Jess Tyson and John AL Armour*

Abstract

\textbf{Background:} The determination of structural haplotypes at copy number variable regions can indicate the mechanisms responsible for changes in copy number, as well as explain the relationship between gene copy number and expression. However, obtaining spatial information at regions displaying extensive copy number variation, such as the \textit{DEFA1A3} locus, is complex, because of the difficulty in the phasing and assembly of these regions. The \textit{DEFA1A3} locus is intriguing in that it falls within a region of high linkage disequilibrium, despite its high variability in copy number (n = 3\textendash16); hence, the mechanisms responsible for changes in copy number at this locus are unclear.

\textbf{Results:} In this study, a region flanking the \textit{DEFA1A3} locus was sequenced across 120 independent haplotypes with European ancestry, identifying five common classes of \textit{DEFA1A3} haplotype. Assigning \textit{DEFA1A3} class to haplotypes within the 1000 Genomes project highlights a significant difference in \textit{DEFA1A3} class frequencies between populations with different ancestry. The features of each \textit{DEFA1A3} class, for example, the associated \textit{DEFA1A3} copy numbers, were initially assessed in a European cohort (n = 599) and replicated in the 1000 Genomes samples, showing within-class similarity, but between-class and between-population differences in the features of the \textit{DEFA1A3} locus. Emulsion haplotype fusion-PCR was used to generate 61 structural haplotypes at the \textit{DEFA1A3} locus, showing a high within-class similarity in structure.

\textbf{Conclusions:} Structural haplotypes across the \textit{DEFA1A3} locus indicate that intra-allelic rearrangement is the predominant mechanism responsible for changes in \textit{DEFA1A3} copy number, explaining the conservation of linkage disequilibrium across the locus. The identification of common structural haplotypes at the \textit{DEFA1A3} locus could aid studies into how \textit{DEFA1A3} copy number influences expression, which is currently unclear.

\textbf{Keywords:} \textit{DEFA1A3}, CNV, Defensin, Structural haplotype

\textbf{Background}

Copy number variation (CNV), involving a deletion or duplication of a region of DNA \(\geq 1\) kb in length, is a common feature of the human genome [1\textendash4]. At some loci, recurrent deletion and duplication events lead to a region being present in a highly variable number of copies; these are referred to as multiallelic copy number variants. Multiallelic CNV has been associated with a variety of disease phenotypes [5\textendash11]. However, an increase in copy number does not always result in increased mRNA levels and, in turn, an increased protein production [12]. In these circumstances, the knowledge of the structure of the locus can be essential for understanding the effect of CNV on phenotype. For example, in the case of the association of low \textit{FCGR3B} copy number with systemic lupus erythematosus (SLE), it is the presence of a zero-copy \textit{FCGR3B} haplotype, which leads to aberrant expression of a chimeric gene, \textit{FCGR2B}, in natural killer cells, that may be the key factor in an increased SLE risk, and not a reduced dosage of \textit{FCGR3B} [13,14]. Therefore, understanding the effect of copy number variation on phenotype comes from not only knowing the copy number of a region, but the spatial arrangement of the locus.
One locus exhibiting multiallelic CNV is the α-defensin DEFA1A3 locus on human chromosome 8p23.1 (Figure 1A) [15-18], with individuals having between 3–16 copies of DEFA1A3 [17-20]. SNPs are usually poor tags of copy number at multiallelic loci, due to the limited ability of a biallelic SNP to tag multiple different copy number states [21]. However, the SNP rs4300027 has been identified as a tag of DEFA1A3 copy number in populations with European ancestry, an association which has not been shown in other populations [18]. At the locus, each DEFA1A3 repeat unit can be occupied by one of two α-defensin genes, either DEFA1 or DEFA3, adding additional complexity. The two genes encode the human neutrophil peptides (HNP) 1–3; these are antimicrobial peptides involved in the innate immune response [22-25]. A recent GWAS found the SNP rs2738048, which falls within the same linkage disequilibrium block as DEFA1A3, to be associated with risk of IgA nephropathy in the Han Chinese population [26]. The basis of this association is unknown, but highlights a need to understand how variation at the DEFA1A3 locus influences HNP1-3 expression. There has only been a single small-scale study comparing DEFA1A3 copy number with HNP1-3 expression, which identified a positive correlation [19]. However, the spatial arrangement of the locus may influence expression.

Non-allelic homologous recombination (NAHR) is the predominant mechanism through which multiallelic copy number variants are formed, requiring segmental duplications (SDs, also referred to as low copy repeats, LCRs) of ≥10 kb in length with ≥95% sequence identity to mediate the rearrangements [4,27,28]. NAHR is a type of homologous recombination occurring between non-allelic copies of a region in different chromosomal positions, resulting in deletions and duplications and can lead to either gene conversion or chromosomal crossover [29-32]. However, there are SNPs either side of the DEFA1A3 locus that display high levels of linkage disequilibrium (LD) (Figure 1B). This suggests that crossover events across the DEFA1A3 region are rare, despite its high variability in copy number. In addition, it is unclear what features of the DEFA1A3 locus are shared between related haplotypes- for example, are haplotypes with the same DEFA1A3 copy number more closely related than haplotypes with different DEFA1A3 copy numbers? Therefore, the mechanisms responsible for generating variation at the DEFA1A3 locus are unclear.

---

**Figure 1** Structure of the DEFA1A3 locus. A) The DEFA1A3 locus consists of two single-copy partial repeats surrounding a variable number of full repeats. Each of the full repeats and the centromeric partial repeat contain a gene locus occupied by either DEFA1 or DEFA3. Symbols show the positions of the variant distinguishing DEFA1 from DEFA3, a 7 bp duplication in intron 1 of each copy of DEFA1A3 and a 5 bp Indel located upstream of each copy of DEFA1A3. The positions of the four SNPs tagging DEFA1A3 haplotype class are shown. Adapted from Khan et al. [18].

B) There are SNPs either side of the DEFA1A3 locus displaying high levels of linkage disequilibrium (D’ = 1), as shown by phased SNP genotype data for the HapMap CEU1 individuals, downloaded from the HapMap project (release #24, phase 1 and 2) [35,54]. D’ values are shown.

---
In this study, flanking sequence variation was used to identify related haplotypes at the \textit{DEFA1A3} locus, in order to determine the shared features of these haplotypes, such as \textit{DEFA1A3} copy number. Previously, we have demonstrated that emulsion haplotype fusion PCR (EHF-PCR) can be used to determine the relative positions of the \textit{DEFA1} and \textit{DEFA3} genes across a haplotype, providing spatial information at the \textit{DEFA1A3} locus [33]. This technique has now been applied to a larger number of haplotypes and to additional variants within the \textit{DEFA1A3} locus to generate more detailed structural haplotypes, allowing an inference of the mechanisms responsible for changes in \textit{DEFA1A3} copy number. This has provided a comprehensive understanding of the common allelic structures of the \textit{DEFA1A3} locus present in populations with European ancestry.

**Results**

**Gene conversion events at \textit{DEFA1A3}**

Our analysis of array-CGH data from Conrad et al. [34] appeared to identify a deletion polymorphism in the telomeric partial repeat of the \textit{DEFA1A3} locus. Further investigation in this study has demonstrated that this is not a deletion, but a replacement of the \textit{DEFA1A3} telomeric partial repeat sequence with sequence from the equivalent region of the \textit{DEFA1A3} full repeats, in a gene conversion event. This is the “Telomeric Replacement Polymorphism” (Figure 2A). The telomeric boundary of

---

**Figure 2** Gene conversion events at \textit{DEFA1A3}. A) The \textit{DEFA1A3} full and partial repeats are highly similar in their sequence, such that the partial repeats can be aligned with the full repeats (blue boxes). However, there are sequence differences between the \textit{DEFA1A3} full and partial repeats. Array CGH data and PCR analysis identified a sequence replacement in the \textit{DEFA1A3} telomeric partial repeat, in which the sequence from the equivalent region of the \textit{DEFA1A3} full repeats replaced the sequence in the telomeric partial repeat location. This occurred over approximately a 2.2 kb interval and is referred to as the “telomeric replacement polymorphism”. B) Sequence data for a 4.1 kb region within the \textit{DEFA1A3} centromeric partial repeat (red box) identified two additional gene conversion events- a ~140 bp event termed “Exchange 1” and a ~1990 bp event termed “Exchange 2”. 
the replacement is within the interval GRCh37/hg19 chr8: 6825864–6825878 and the centromeric boundary within the interval chr8: 6828055–6828082; therefore, it covers a region of approximately 2.2 kb. 16 of the 120 HapMap CEU1 haplotypes (13%) carry the telomeric replacement polymorphism.

Sequence data obtained from the centromeric partial repeat (GRCh37/hg19 chr8: 6876778–6880877) identified two similar events (Figure 2B). In each case, the sequence in the centromeric partial repeat location has been replaced by sequence from the equivalent region of the full repeats. Exchange 1 haplotypes contain a replacement extending over at least 140 bp. The centromeric boundary of the replacement falls within the region chr8: 6876928–6876969. The telomeric boundary extends beyond chr8: 6876788, although it may be continuous with the block of full repeats. Exchange 2 haplotypes contain a replacement extending over at least 1990 bp. The centromeric boundary of the replacement falls within the region chr8: 6878778–6879044. The telomeric boundary extends beyond chr8: 6876788, although it may be continuous with the block of full repeats.

**DEFA1A3 haplotype classes**

Sequence similarity across the 4.1 kb centromeric flanking region allowed the identification of five different classes of DEFA1A3 haplotype, in which haplotypes within each class shared identical or highly similar flanking sequence, which was distinct from the sequence of haplotypes within the other classes. These five classes are: Reference Sequence (the sequence found in the GRCh37/hg19 human reference assembly), Class 1, Class 2 (each contains multiple unique sequence differences compared to the Reference Sequence), Exchange 1 (contains the Exchange 1 sequence replacement polymorphism) and Exchange 2 (contains the Exchange 2 sequence replacement polymorphism).

Four SNPs were identified, through a combination of sequencing and analysis of phased HapMap data [35] (see Methods), which, due to their pattern of LD, are able to tag the five DEFA1A3 haplotype classes (Figure 1, Table 1 and Additional file 1: Table S1). The diploid DEFA1A3 class genotype was identified for an additional two 1000 Genomes project Asian (ASN) haplotypes. The diploid genotype profile can be used to determine the haplotype classes of the two haplotypes for the genotyped individual.

| Haplotype class | SNP Genotypes |
|-----------------|---------------|
| Reference        | T T A C        |
| Class 1          | C T G C        |
| Class 2          | C C A C        |
| Exchange 1       | T T A A        |
| Exchange 2       | C T A C        |

**Table 1 Linkage disequilibrium of DEFA1A3 flanking SNPs**

| Haplotype class | SNP Genotypes |
|-----------------|---------------|
| Reference        | T T A C        |
| Class 1          | C T G C        |
| Class 2          | C C A C        |
| Exchange 1       | T T A A        |
| Exchange 2       | C T A C        |

The genotype combinations for the four tag SNPs able to identify each of the five DEFA1A3 haplotype classes. Due to the pattern of linkage disequilibrium (Additional file 1: Table S1), a diploid genotype profile can be used to determine the haplotype classes of the two haplotypes for the genotyped individual.

Features of DEFA1A3 haplotype classes

The SNP rs4300027 has previously been identified as a tag of DEFA1A3 copy number in the European population ($p = 1.3 \times 10^{-45}$), with the C allele tagging haplotypes with a low DEFA1A3 copy number (2–3 copies) and the T allele tagging haplotypes with a high DEFA1A3 copy number (4–5 copies) [18]. We investigated whether a combination of four SNP genotypes may further partition this association or identify tags of additional features. Therefore, the diploid DEFA1A3 haplotype class (e.g. Class 1 homozygous/ heterozygous/ negative) was compared to features of the DEFA1A3 copy number variable region using either a Chi Square or Cochran Armitage test, for 599 unrelated European individuals (Table 2 and Additional file 1: Table S2). The diploid DEFA1A3 copy number, as well as the ratio of the number of copies of DEFA1 versus DEFA3, inserted to deleted form of a 5 bp InDel (InDel5) and unduplicated to duplicated form of a 7 bp duplication (7 bp dup) are known for these samples (Figure 1A) [18], allowing a comparison between DEFA1A3 haplotype class and multiple variants at the locus.

This demonstrates that each DEFA1A3 haplotype class has its own unique profile of features. Both the Reference Sequence and Exchange 1 haplotypes are associated with a high DEFA1A3 copy number, whilst Class 1 and Class 2 are associated with a low DEFA1A3 copy number (Table 2). This forms the basis of the previously reported association between the SNP rs4300027 and DEFA1A3 copy number in the European population, with

The genotype combinations for the four tag SNPs able to identify each of the five DEFA1A3 haplotype classes. Due to the pattern of linkage disequilibrium (Additional file 1: Table S1), a diploid genotype profile can be used to determine the haplotype classes of the two haplotypes for the genotyped individual.
Reference Sequence and Exchange 1 haplotypes having the T allele and Class 1 and Class 2 haplotypes the C allele. However, haplotype-specific copy number information, obtained by Khan et al. demonstrates that each DEFA1A3 haplotype class is not associated with a single copy number state [18]. For example, whilst Exchange 1 haplotypes are generally associated with a high copy number, Exchange 1 haplotypes with between 2–7 copies have been observed in the HapMap CEU1 population (data not shown) [18]. In addition to associations with DEFA1A3 copy number, all five classes show an association with the frequency of the DEFA3 gene and the Indel-5 insertion, whilst only two DEFA1A3 haplotype classes are significantly associated with the frequency of the 7 bp duplication (Table 2).

However, these associations are based solely on the European population and it is not clear if the same features can be extended to other worldwide populations. Whilst information for the three internal allelic variants (DEFA1/DEFA3, Indel-5 and 7 bp duplication) was unavailable, the DEFA1A3 copy number has been estimated for 1047 of the 1092 individuals within the 1000 Genomes project. This allowed a comparison of DEFA1A3 haplotype class with DEFA1A3 diploid copy number in non-European populations (Table 3 and Additional file 1: Table S3). The 1000 Genome AMR samples show an association of Class 2 with a low DEFA1A3 copy number and Reference Sequence and Exchange 1 with a high copy number, the same as was observed in the CEU and HRC samples. However, the association of Class 1 and a low copy number, which was observed for the European samples, is not replicated in the AMR samples. Despite this, the SNP rs4300027 is still significantly associated with DEFA1A3 copy number in the AMR samples ($p = 7 \times 10^{-10}$). The ASN samples also show associations of Class 2 with a low DEFA1A3 copy number and of Reference Sequence with a high copy number, as well as of Exchange 2 with a low copy number. Whilst this mirrors the association with rs4300027 observed in the EUR samples, Class 1 is significantly associated with a high DEFA1A3 copy number in the ASN samples, leading to a weaker association between rs4300027 genotype and DEFA1A3 copy number ($p = 5 \times 10^{-4}$). There are no significant associations between haplotype class and DEFA1A3 copy number in the AFR samples, suggesting high within-class variability in copy number and explaining the observation that the SNP rs4300027 does not tag copy number in this population ($p = 0.114$). The 1000 Genome dataset also provides information on additional EUR samples. The associations identified differ from those observed previously; although the association between Class 2 and a low

### Table 2 Comparing DEFA1A3 haplotype class with features of the locus in individuals with European ancestry

| Haplotype class | DEFA1A3 copy number | DEFA3 frequency | Indel5 insertion frequency | 7 bp duplication frequency |
|-----------------|---------------------|-----------------|---------------------------|---------------------------|
| Reference Sequence | $6 \times 10^{-25}$; high | $1 \times 10^{-1}$; high | $2 \times 10^{-10}$; low | $3 \times 10^{-11}$; low |
| Class 1 | $2 \times 10^{-1}$; low | $4 \times 10^{-9}$; low | $3 \times 10^{-3}$; high | NS |
| Class 2 | $1 \times 10^{-12}$; low | $1 \times 10^{-12}$; high | $2 \times 10^{-10}$; low | NS |
| Exchange 1 | $3 \times 10^{-5}$; high | $2 \times 10^{-14}$; low | $1 \times 10^{-39}$; high | $2 \times 10^{-39}$; high |
| Exchange 2 | NS | $3 \times 10^{-5}$; low | $2 \times 10^{-2}$; low | NS |

$p$-values and direction of significant associations obtained for Chi Square or Cochran-Armitage tests (Additional file 1: Table S2) comparing DEFA1A3 haplotype class with features of the DEFA1A3 locus, based on the HapMap CEU and ECACC HRC samples. All $p$-values were adjusted for multiple testing using Bonferroni correction. NS = not significant. High = associated with a high DEFA1A3 copy number/allele frequency; Low = associated with a low DEFA1A3 copy number/allele frequency.
DEFA1A3 copy number and Reference Sequence with a high copy number are observed, Class 1 and Exchange 1 show no association with DEFA1A3 copy number and a novel association of Exchange 2 with a low copy number is observed.

DEFA1A3 haplotype structures

Emulsion haplotype fusion PCR (EHF-PCR) has previously been applied at the DEFA1A3 locus to determine the relative positions of the DEFA1 and DEFA3 genes across a haplotype [33]. By applying this technique to additional variants, a more detailed picture of the haplotype structures of the locus can be determined, allowing the mechanisms of change in copy number at DEFA1A3 to be identified. For 84 independent haplotypes in the HapMap CEU1 population, the DEFA1A3 copy number, as well as the ratio of DEFA1 vs. DEFA3 and inserted to deleted form of Indel5 are known [18]. The positions of these allelic variants were determined for 61 of these 84 haplotypes using EHF-PCR and sequencing (Figure 4). The centromeric-most Indel5 site is located within the sequenced flanking region (GRCh37/hg19 chr8: 6876778–6880877) and as such, the centromeric-most position was captured for all 120 haplotypes. In this location, only the deleted form was observed.

The structures of 17 Reference Sequence haplotypes were determined, 12 of which display one of the structures shown in Figure 4. Whilst there are many different structural haplotypes, with 4 or 5 copies of DEFA1A3, the structures are highly similar to each other, with single copy differences between most structures that could be accounted for by a single deletion or duplication event. Two of the haplotypes with structures that do not fit this pattern have only two copies of DEFA1A3, which is unusual for haplotypes within the Reference Sequence class, whilst the three others have the Indel5 insertion at the second repeat from the telomeric end, suggesting a structural rearrangement.

For Class 1, 9 of the 11 haplotypes observed display a structure shown in Figure 4, in which there are a variable number of copies of DEFA1A3. The remaining 2 haplotypes display a different structure, with fewer copies of DEFA1A3.

| Haplotype class | Africa | America | Asia | Europe |
|-----------------|--------|---------|------|--------|
| Reference Sequence | NS | 3×10^-4; high | 0.009; high | 2×10^-26; high |
| Class 1 | NS | NS | 9×10^-15; high | NS |
| Class 2 | NS | 1×10^-6; low | 1×10^-8; low | 1×10^-18; low |
| Exchange 1 | - | 0.003; high | - | NS |
| Exchange 2 | NS | NS | 0.045; low | 3×10^-6; low |

Table 3 Comparing DEFA1A3 haplotype class with copy number in 1000 Genome individuals

p-values and direction of significant associations obtained for Chi Square or Cochran-Armitage tests (Additional file 1: Table S3) comparing DEFA1A3 haplotype class with DEFA1A3 copy number, based on the 1000 Genomes samples. All p-values were adjusted for multiple testing using Bonferroni correction. NS = not significant, - = no test performed (see Additional file 1: Table S3). High = associated with a high DEFA1A3 copy number; Low = associated with a low DEFA1A3 copy number.

Figure 4 DEFA1A3 haplotype structures. The common DEFA1A3 haplotype structures observed in haplotypes with European ancestry (HapMap CEU1) obtained using EHF-PCR and sequencing. “Del” and “Ins” refer to the deleted and inserted forms of the Indel5 variant. 52 of the 61 haplotypes observed (85%) conformed to one of these structures.
number of copies of a repeat unit containing \textit{DEFA1} and the Indel5 insertion. In this case, not only do the haplotypes share a similar structure, but the repeat units are also highly similar. The two haplotypes which do not fit this pattern both have a copy of the \textit{DEFA3} gene, which is usually absent from Class 1 haplotypes.

For Class 2, 14 of the 15 haplotypes observed display a structure shown in Figure 4. Again, there is a single deletion or duplication event that could account for the differences between the haplotypes. The haplotype which does not fit this pattern lacks \textit{DEFA3}, which is uncommon for Class 2 haplotypes.

For Exchange 1, 14 of the 15 haplotypes display the structure shown in Figure 4; these are similar to Class 1 haplotypes, but vary in copy number from 1–5 copies. Again, it is not just the structures, but the repeat units that are highly similar. The haplotype that does not fit this pattern includes the \textit{DEFA3} gene, which is usually absent from Exchange 1 haplotypes. There are only three examples of Exchange 2 haplotypes analysed, all with the same structure and with all three repeats containing both \textit{DEFA1} and the Indel5 deletion.

Although there is within-class variation, the structures identify common features of each \textit{DEFA1A3} class. For example, if Reference Sequence haplotypes contain a copy of the \textit{DEFA3} gene, it is in the centromeric-most copy, whereas for Class 2 haplotypes, it is in the telomeric-most copy. For the Indel5 variant, both Class 1 and Exchange 1 haplotype have the deletion allele in the centromeric-most copy of the array. For Reference Sequence haplotypes, the Indel5 insertion allele is always in the telomeric-most copy, whereas for Class 2, it is in the middle copy of three-copy haplotypes.

\textbf{Discussion}

In order to fully understand the relationship between multiallelic CNV and expression, it is necessary to not only reliably genotype the copy number, but to understand the positions of these copies across a haplotype. This is especially true at a locus like \textit{DEFA1A3}, in which each repeat unit in the array can be occupied by one of two different genes. Through the use of flanking sequence information, five common \textit{DEFA1A3} haplotype classes have been identified, each of which has specific associations with internal variants in populations of European ancestry. A small-scale analysis on 1000 Genomes sample demonstrates between-population differences within each \textit{DEFA1A3} haplotype class. This is due to the combination of the different copy number distributions associated with each haplotype class and the varying frequencies of the classes between populations.

The use of EHF-PCR to provide spatial information at the \textit{DEFA1A3} locus has been expanded in this study, to look not only at the positions of the \textit{DEFA1} and \textit{DEFA3} genes across a haplotype, but also the positions of an additional allelic variant, Indel5. Although the Indel5 variant does not necessarily change the expression or function of HNP1-3, it provides an additional landmark across a haplotype, allowing a more detailed comparison of haplotype structures. In total, structural haplotypes were obtained for 61 independent haplotypes within the HapMap CEU1 population. The CEU1 population sample appears to be representative of the wider European cohort in terms of the associations between \textit{DEFA1A3} class and features of the locus (data not shown). Therefore, the structures observed should be representative of haplotypes with European ancestry. The structures observed show that haplotypes within each \textit{DEFA1A3} class have highly similar structures, despite having different copy numbers. In addition, some classes have multiple copies of a repeat unit containing the same gene and Indel5 allele, which was expected, given that all five \textit{DEFA1A3} classes are significantly associated with either a high or low frequency of both \textit{DEFA3} and the Indel5 insertion. This repeat unit similarity is likely to promote NAHR, which relies on high sequence identity to facilitate rearrangements. Given that the \textit{DEFA1A3} locus falls within a region of high LD, this information suggests that the major mechanisms for copy number change at the \textit{DEFA1A3} locus involve intra-allelic rearrangements—i.e. NAHR between haplotypes from the same \textit{DEFA1A3} haplotype class. This process would allow changes in \textit{DEFA1A3} copy number, via NAHR resulting in chromosomal crossover, but would preserve the surrounding LD, as rearrangements would occur between haplotypes within the same \textit{DEFA1A3} class. This presumably results from a bias towards NAHR between sister chromatids, rather than between homologous chromosomes, during meiosis. A bias for NAHR between sister chromatids has been observed previously at the tandemly duplicated human alpha satellite DNA [36], as well as a bias towards NAHR between homologue chromosomes at the a1 locus in maize [37].

A clear example of intra-allelic NAHR is observed in the Exchange 1 class, which is expected to be younger than the four other classes, given that it is absent from the Asian population and is very rare in the African population. Exchange 1 haplotypes with between 2 and 7 copies have been observed and the structures identified in samples with European ancestry show there are variable numbers of copies of a repeat unit with the same gene and Indel5 allele.

However, intra-allelic rearrangements will not be the only mechanism operating at the locus. Inter-allelic rearrangements will occur, but given the conservation of LD across the \textit{DEFA1A3} locus, it is likely that inter-allelic NAHR more often results in gene conversion than chromosomal crossover. This study has identified
three gene conversion events occurring in the flanking regions of the DEFA1A3 locus, supporting this idea. In addition, the vast majority of Class 1 haplotypes lack DEFA3, but DEFA3-positive Class 1 haplotypes have been observed and this is likely to have resulted from an intro
duction of DEFA3 to a Class 1 background, via gene conversion. This idea is consistent with previous studies identifying gene conversion events at variable number tandem repeat loci [38-42], to which DEFA1A3 is com
parable, given it contains multiple copies of a repeat unit with high sequence similarity, positioned in tandem. Gene conversion will homogenise repeat units, which in turn will facilitate further rearrangement events.

Many studies at regions of CNV fail to identify a robust association between copy number and disease risk. This is due to a combination of two factors. Firstly, many studies fail to accurately measure multiallelic copy number, leading to an association that cannot be repro
duced [43-47]. Secondly, there is an expectation of a linear relationship in which an increase in copy number results in a proportionate increase in protein expression; however, this is not always the case [12]. As shown for the FCGR3B and NBPF23 loci, knowledge of the allelic structures of the region may be necessary to determine how CNV influences gene expression [13,14,48]. Despite accurate measurement of DEFA1A3 copy number [18], the relationship between gene copy number and expres
sion at this locus remains unclear. Although a positive correlation between DEFA1A3 copy number and HNP1-3 expression has been reported previously [19], this was a small-scale study. Structural information may be required to fully understand the relationship between DEFA1A3 copy number and HNP1-3 expression, as well as explain the association between the DEFA1A3 locus and IgA ne
phropathy risk [26]. In samples with European ancestry, haplotypes within each DEFA1A3 class have highly similar structures. Therefore, the simple genotyping of the four SNPs which tag DEFA1A3 haplotype class, identified in this study, will be sufficient for inferring haplotype structures for haplotypes with European ancestry. This approach could easily be applied to studies comparing DEFA1A3 structure with HNP1-3 expression or associ
ation with a disease phenotype. Therefore, the use of structural information, as derived here for DEFA1A3, should be applied to other copy number variable loci, in order to explain associations between the variation observed and protein expression. This may, in turn, aid the understanding of the features of a copy number variable locus that influence disease risk.

Conclusions
We have defined five common classes of haplotype at the DEFA1A3 locus. Each class is associated with particular features of the DEFA1A3 locus and these associations differ between populations. Structural haplotypes have been obtained across the DEFA1A3 locus for 61 haplo
types with European ancestry, allowing the identification of the common allelic structures at DEFA1A3. The structures suggest that intra-allelic rearrangement is the predominant mechanism resulting in copy number variation at the DEFA1A3 locus.

Methods
DNA samples
180 HapMap phase I (CEU1) and II (CEU2) samples [49] and 480 Human Random Control (HRC; panels 1-5) unrelated UK samples from the European Collection of Cell Cultures (ECACC) [50] were used for the study. The DNA was extracted from lymphoblastoid cell lines.

Measuring DEFA1A3 copy number
Diploid DEFA1A3 copy number was measured for the 180 HapMap and 480 ECACC HRC samples and haplo
type DEFA1A3 copy numbers were defined for 84 haplotypes from the HapMap CEU1 population as de
scribed by Khan et al. [18]. The diploid DEFA1A3 copy number was estimated for 1047 samples within the 1000 Genomes project using read depth analysis of whole genome sequence data [51]. Raw read data was down
loaded from the 1000 Genomes project [51,52]. The reads mapping to the DEFA1A3 locus (GRCh37/hg19 chr8: 6829298–6837591, 6848458–6856701 and 6867561–6875800) and two single-copy flanking regions (GRCh37/ hg19 chr8: 6700000–6830000 and 6900000–7000000) were counted using Samtools [53], with the command samtools view –c. Flanking regions were selected to have a similar GC content to the copy number variable region selected from the DEFA1A3 locus. The ratio of reads per base for the DEFA1A3 locus to the reads per base for the flanking regions was obtained and multi
plied by two to give the diploid DEFA1A3 copy number estimation, which was rounded to the nearest integer value. The 1047 samples consist of individuals with European (EUR) (n = 364), Asian (ASN) (n = 280), African (AFR) (n = 228) and American (AMR) (n = 175) ancestry. EUR = CEU + FIN + GBR + IBS + TSI. ASN = CHB + CHS + JPT. AFR = ASW + LWK + YRI. AMR = CLM + MEX + PUR. A comparison with copy numbers estimated by Khan et al. [18] shows read depth provides an accurate estimation of DEFA1A3 copy number (supplementary methods and Additional file 1: Figure S1).

Identification of DEFA1A3 haplotype classes
A 4.1 kb region immediately centromeric to the DEFA1A3 locus (GRCh37/hg19 chr8: 6876778–6880877) was resequenced across the 30 HapMap CEU1 trios, allowing complete phased haplotype sequences to be obtained using segregation or allele-specific PCR (supplementary
methods and Additional file 1: Table S4). The primers designed to amplify the region were designed to ensure amplification specifically from the DEFA1A3 centromeric partial repeat. PCR products were purified using Ampure-xP (Agencourt), according to the manufacturer's protocol. Approximately 20 ng of purified PCR product was Sanger sequenced using Big Dye (Invitrogen), according to the manufacturer's protocol. Sequenced products were cleaned using CleanSeq (Agencourt) according to the manufacturer's protocol and analysed using an ABI 3730. Twelve primers were used to sequence the region (Additional file 1: Table S5).

Genotyping of DEFA1A3 haplotype classes
Flanking sequence information allowed the identification of five major DEFA1A3 haplotype classes. A combination of four SNPs is able to tag these five DEFA1A3 haplotype classes; three SNPs were identified from sequencing, as described above, whilst the fourth, rs7826487, was identified from phased HapMap genotype data [35,54]. A PCR-RFLP assay was used to genotype each SNP across the HapMap CEU2 and ECACC HRC samples. The SNP rs4300027 was genotyped as described by Khan et al. [18]. The primers and cycling conditions for the other three assays are shown in Additional file 1: Table S6. All assays used 1 µM each primer, 0.5 Units Taq DNA polymerase (NEB), 10 ng genomic DNA and a standard buffer, with final reaction concentrations of 50 mM Tris HCl pH8.8, 1.4 mM Ammonium Sulphate, 7.5 mM 2-mercaptoethanol, 200 µM each dNTP and 125 µg/ml BSA. The same four SNPs were genotyped as part of the 1000 Genomes project [51,55].

Telomeric replacement polymorphism
A three-primer assay was designed to genotype the telomeric replacement polymorphism in the HapMap CEU1 individuals. The forward primer AGACAGCAGATCCGG TATAATC produces a 645 bp product with the reverse primer AGAGCCCAATAAATCTAACGG from non-replacement haplotypes or a 453 bp product with the reverse primer GACTCGTTCTTTCTGGATTAC from haplotypes carrying the replacement. The cycling conditions consisted of an initial denaturation at 95°C for 3 minutes, followed by 36 cycles of 95°C for 30 seconds, 60°C for 30 seconds and 70°C for 30 seconds. Each 10 µl reaction contained 1 µM each primer, 0.5 Units Taq DNA polymerase (NEB), 10 ng genomic DNA and a standard buffer, as described above.

Statistical analysis
A series of Chi Square and Cochran-Armitage tests were used to compare DEFA1A3 haplotype class with features of the DEFA1A3 locus. The copy number and frequency categories were designated such that each category was comparably populated. For Class 1 and Exchange 2, homozygous and heterozygous individuals were grouped, due to their low frequencies. Individuals were counted multiple times; for example, an individual homozygous positive for the Reference Sequence would have also been counted as homozygous negative for Class 1, Class 2, Exchange 1 and Exchange 2. To account for this, p-values were adjusted using Bonferroni correction.

Emulsion haplotype fusion PCR
Four emulsion haplotype fusion PCR (EHF-PCR) systems were used, based on a design described by Tyson and Armour [33]. Two single-copy flanking regions, one centromeric and one telomeric to the DEFA1A3 locus, were fused to the region containing the variant distinguishing DEFA1 from DEFA3 (termed “telomeric gene” and “centromeric gene” respectively) or the region containing the Indel5 variant (termed “telomeric Indel5” and “centromeric Indel5”). The telomeric flanking region contains four SNPs (rs2978951, rs2738046, rs2738045 and rs2702910), the phased genotype data for which was obtained from the HapMap project [35,54]. The centromeric flanking region contains three SNPs (rs4300027, rs4512398 and rs17382102), the phased genotype data for which was obtained from sequencing in this work.

The 25 µl aqueous phase for each EHF-PCR contained 1× Phusion GC buffer (NEB), 0.2 mM each dNTP, 1 µM F1 primer, 25nM F2’R1 primer, 1 µM R2 primer, 2 Units Phusion DNA polymerase (NEB) and 50 ng genomic DNA. The preparation of the emulsion was adapted from a method first described by Turner and Hurles [56]. The aqueous phase was aliquoted to 0.5 ml PCR tubes, to which 50 µl silicone oil (described in [56]) was added. A 3 mm tungsten carbide bead (Qiagen) was added to the lid of the tube and the tube was closed such that it remained in an inverted position; otherwise, the bead remains in a constrained space at the bottom of the tube during vortexing. Inversion of the tube is essential for emulsion formation. The inverted tube was vortexed at speed 5 for 1 minute 30 seconds using a Vortex Genie 2, to give aqueous droplets approximately 5 µM in size (data not shown). The primers and cycling conditions are shown in Additional file 1: Table S7.

25 µl of 1× Phusion GC buffer (NEB) was added to each sample post-PCR to increase the volume of the aqueous phase; this was recovered as described by Tyson and Armour [33]. Allele-specific reamplification of the fusion products was performed to allow haplotype-specific sequencing of the fused region. The primers and cycling conditions used are shown in Additional file 1: Table S8. All reamplifications were performed in a 20 µl reaction containing 1× NH₄ buffer (Bioline), 2 mM MgCl₂,
0.2 mM each dNTP, 0.5 μM each primer, 1 Unit *Taq* DNA polymerase (Bioline) and 1 μl PCR template. The exception was reamplifications using centromeric Indel5 products, for which the 20 μl reaction contained 0.5 μM each primer, 1 Unit *Taq* DNA polymerase, 1 μl PCR template and 1× standard buffer, as described above. The reamplified products were purified and sequenced, as described above.

**Availability of supporting data**
The data sets supporting the results of this article are included in Additional files 2, 3, 4, 5 and 6.

**Additional files**

**Additional file 1**: Supplementary methods, figures and tables. Contains additional information explaining the methods used to phase the *DEFA1A3* centromeric flanking region and how read depth was used to estimate *DEFA1A3* copy number, as well as supplementary Tables 1–8 and supplementary Figure 1.

**Additional file 2**: Data file. Includes the data used for analysis in this publication: *DEFA1A3* copy numbers estimated for 1047 of the 1000 Genomes samples using read depth; SNP genotypes and the *DEFA1A3* haplotype types assigned to the HapMap CEU and ECACC HRC samples (rs4300027 genotypes from Khan et al. [18]); Telomeric Replacement Polymorphism genotypes for the HapMap CEU1 samples; haplotype structures for 61 HapMap CEU1 haplotypes.

**Additional file 3**: *DEFA1A3* read depth copy number estimates. *DEFA1A3* copy numbers estimated for 1047 of the 1000 Genomes samples using read depth.

**Additional file 4**: *DEFA1A3* haplotype classes. SNP genotypes and the *DEFA1A3* haplotype classes assigned to the HapMap CEU and ECACC HRC samples (rs4300027 genotypes from Khan et al. [18]).

**Additional file 5**: Telomeric replacement polymorphism. Telomeric Replacement Polymorphism genotypes for the HapMap CEU1 samples.

**Additional file 6**: *DEFA1A3* haplotype structures. Haplotype structures for 61 HapMap CEU1 haplotypes.

**Competing interests**

The authors declare that they have no competing interests.

**Authors’ contributions**

HAB performed Sanger sequencing to identify the *DEFA1A3* haplotype classes, performed rs7826487, rs7825750 and rs62487514 genotyping of the HapMap CEU and ECACC HRC samples, estimated *DEFA1A3* copy number for the 1000 Genomes samples, designed and performed the EHF-PCR assays, analysed the data and drafted and revised the manuscript. FFX identified the 1000 Genomes samples, designed and performed the EHF-PCR assays, reamplified products were purified and sequenced, as described above. The reamplified products were analysed and revised, as described above.

**Acknowledgements**

We are grateful to Danielle Carpenter for helpful suggestions. HAB is supported by a BBRC Doctoral Training Award (BB/F016999/1). FFX was supported by a scholarship from the Government of Pakistan and University of Karachi (B/Estt(Tb)2007). This work was also made possible by the award of a Leverhulme Trust Research Fellowship to JA.

Received: 29 April 2014 Accepted: 14 July 2014 Published: 21 July 2014

**References**

1. Stankiewicz P, Lupski JR: Structural Variation in the human genome and its role in disease. *Annu Rev Med* 2010, 61:437–455.

2. Feuk L, Carson AR, Scheer SW: Structural variation in the human genome. *Nat Rev Genet* 2006, 7(2):85–97.

3. Fanucci M, Petretto E, Atman TJ: Gene copy number variation and common human disease. *Clin Genet* 2010, 77(3):201–213.

4. Freeman JL, Perry GH, Feuk L, Redon R, McCarthy SA, Aithalher DM, Aburanti H, Jones KW, Tyler-Smith C, Hurles ME, Carter NP, Scherer SW, Lee C: Copy number variation: new insights in genome diversity. *Genome Res* 2006, 16(8):949–961.

5. Holloix EJ, Haffmeier U, Zeeuwen PLJM, Palla R, Lasorz T, Riedijk-Oth Huis D, van de Kerkhof PCGM, Traupe H, de Jongh G, den Heijer M, Reis A, Armour JAL, Schalkwijk J: Priorisation is associated with increased beta-defensin genomic copy number. *Nat Genet* 2008, 40(1):23–25.

6. Stewart GE, Haffmeier U, Nair RP, Palla R, Tejasvi T, Schalkwijk J, Elder JT, Reis A, Armour JAL: Association of beta-defensin copy number and psoriasis in three cohorts of European origin. *J Invest Dermatol* 2012, 132(10):2407–2413.

7. Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BL, Quinones MP, Ramshad MJ, Marrohy KK, Rovin BH, Bradley W, Clark RA, Anderson SA, O’Connell RJ, Agan BK, Ahuja SS, Bologna R, Ten L, Dolan MJ, Ahuja SK: The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* 2005, 307(5714):1434–1440.

8. Willcocks LC, Lyons PA, Claxworthy MR, Robinson JJ, Yang W, Newland SA, Fagnan V, McGavren NN, Condiffe MN, Chivers ER, Audu D, Jolly EC, Watts R, Lau YL, Morgan AW, Nash G, Smith KG: *Copy number of FGR3B*, which is associated with systemic lupus erythematosus, correlates with protein expression and immune complex uptake. *J Exp Med* 2008, 205(7):1573–1582.

9. Fellermann K, Stange DE, Schaeffeler E, Schmalzl H, Wehkamp J, Bevins CL, Reinsch W, Temm A, Schwab M, Lichten P, Radlwinner B, Stange EFG: A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. *Am J Hum Genet* 2006, 79(3):439–448.

10. Bentley RW, Pearson J, Geary RB, Barclay ML, McKinney C, Merriman TR, Roberts RL: Association of higher *DEFB4* genomic copy number with Crohn’s disease. *Am J Gastroenterol* 2010, 105(2):354–359.

11. McKinney C, Merriman ME, Chapman PT, Gow PJ, Hanson AA, Highton J, Jones PBB, McLean L, O’Donnell J, Pokorny V, Spillerberg M, Stamp LJ, Willis J, Steer S, Merriman TR: Evidence for an influence of chemokine ligand 3-like 1 (CCL3L1) gene copy number on susceptibility to rheumatoid arthritis. *Ann Rheum Dis* 2008, 67(3):409–413.

12. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C, Tyler-Smith C, Carter N, Scherer SW, Tavare S, Deloukas P, Hurles ME, Dernitzkis ET: Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 2007, 315(5813):848–853.

13. Mueller M, Barros P, Witherden AS, Roberts AL, Zhang Z, Schaschl H, Yu CY, Hurles ME, Schaffner C, Floto RA, Game L, Steinhöfer KM, Wilson RK, Graves TA, Eichler EE, Cook HT, Vyte JS, Atman TJ, Genomic pathology of SLE-associated copy number variation at the *FGR3C2/FGR3B/FGR3B2* Locus. *Am J Hum Genet* 2013, 92(1):28–40.

14. van der Heijden J, Breunis WB, Kuijpers TW: Phenotypic Variation in IgG Receptors by Nonclassical FCGR2C Alleles. *A 450-kb contig of defensin genes on human chromosome 8p23.* *Genes Immun* 2003, 4(7):257–264.

15. Bird CP, de Grassi A, Lee C, Tyler-Smith C, Carter N, Scherer SW, Tavare S, Deloukas P, Hurles ME, Dernitzkis ET: Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 2007, 315(5813):848–853.

16. Linnemayer MC, Petretto E, Atman TJ: Gene copy number variation and common human disease. *Clin Genet* 2010, 77(3):201–213.

17. Chan YK, Tam TSC, Cho RW, Au CY: Copy number polymorphism and expression level variation of the human alpha-defensin genes *DEFA1* and *DEFA3*. *Hum Mol Genet* 2005, 14(1):2045–2052.

18. Khan FF, Carpenter D, Mitchell L, Mansouri O, Black HA, Tyson J, Armour JA: Accurate measurement of gene copy number for human alpha-defensin *DEFA1A3*. *BMC Genomics* 2013, 14:719.

19. Linneker RM, Ganz T: Human defensin gene copy number polymorphisms: comprehensive analysis of independent variation in alpha- and beta-defensin regions at 8p22-p23. *Genomics* 2005, 86(4):432–450.

20. Cheng FJ, Zhou XJ, Zhao YF, Zhao WH, Zhang H: Alpha-defensin *DEFA1A3* gene copy number variation in Asians and its genetic association study.
in Chinese systemic lupus erythematosus patients. Gene 2013, 517(2):158–163.

21. Beckmann JS, Estvill X, Antonarakis SE: Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. Nat Rev Genet 2007, 8(8):589–596.

22. Gantz T, Selsted ME, Szklarek D, Harwig SS, Daher K, Bainton DF, Lehrer RI: Defensins. Natural peptide antibiotics of human neutrophils. J Clin Invest 1985, 76(4):1427–1435.

23. Lehrer RI, Barton A, Daher KA, Harwig SS, Gantz T, Selsted ME: Interaction of human defensins with Escherichia coli. Mechanism of bactericidal activity. J Clin Invest 1989, 84(2):533–561.

24. Chertov O, Yang D, Howard OMZ, Oppenheim JJ: Leukocyte granule proteins mobilize innate host defenses and adaptive immune responses. Immunol Rev 2000, 177:68–78.

25. Valore EV, Gantz T: Posttranslational processing of defensins in immature human myeloid cells. Blood 1992, 79(6):1538–1544.

26. Yu XQ, Li M, Zhang H, Low HQ, Wei X, Wang JQ, Sun LD, Sim KS, Li Y, Foo JH, Wang W, Li Y, Xin Y, Tang XQ, Fan L, Chen J, Lin RS, Wang XY, Liu ZS, Luo TQ, Zhu L, Huang XJ, Zhang XJ, Liu ZH, Liu JJ: A genome-wide association study in Han Chinese identifies multiple susceptibility loci for IgA nephropathy. Nat Genet 2012, 44(1):178–182.

27. Shaw CJ, Lupski JR: Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. Hum Mol Genet 2004, 13:R57–R64.

28. Stankiewicz P, Lupski JR: Genome architecture, rearrangements and genomic disorders. Pathogenetcs 2008, 11(1):344–357.

29. Hastings PJ, Lupski JR: Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. Hum Mol Genet 2004, 13:R57–R64.

30. Attia El, Otendalski-Hesse L, Bowdrey J, Hattab-Golov I, Naimi E, Yimer G, Amogne W, Mugusi S, Mintz O, Makonnen E, Janabi M, Mugusi F, Adareyaye G, Hardwick R, Fu BY, Viskaduraki M, Yang FT, Hollfess EJ: CCL3L1 copy number, HIV load, and immune reconstitution in sub-Saharan Africans. BMC Infect Dis 2013, 13:536.

31. Diskin SJ, Hou CP, Glesner JT, Attiyeh EF, Laundenslager M, Bosse K, Cole K, Mosse YP, Wood A, Lynch JE, Pecor K, Diamond M, Winter C, Wang K, Kim C, Geier EA, McGrady PW, Blakemore AF, London WB, Shakh TH, Bradfield J, Grant SF, Li HZ, Devoto M, Rappaport ER, Hakinsoon H, Maris JM: Copy number variation at 1q21.1 associated with neuroblastoma. Nature 2009, 459(7249):987–9112.

32. Corex Cell Repositories. http://ccl.coriell.org.

33. The European Collection of Cell Cultures. http://www.hpacultures.org.uk.

34. Altshuler DM, Durbin RM, Abeecas GS, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eicher EE, Field SF, Howson KM, Smith DR, Lehrer RI: Meiotic recombination and flanking sequence variation from 1,092 human genomes. Nature 2010, 464(7297):704–712.

35. Authier D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P, Gibbs RA, Belmont JW, Boudreau A, Leal SM, Hardenbol P, Pasterkam S, Wheeler DA, Willis TJ, Yu FY, Yang HM, Zeng CCQ, Qiao Y, Hu HR, Hu WT, Li CH, Lin W, Liu SQ, Pan H, Tang XL, Wang J, Wang W, Yu J, Zhang B, Zhang QR, et al: A haplotype map of the human genome. Nature 2005, 437(7063):1299–1302.

36. Warburton PE, Willard HF: Interhomolog recombination sequence variation of alpha satellite DNA from human chromosome 17: evidence for concerted evolution along haploidy lineages. J Mol Evol 1995, 41(6):1006–1015.

37. Yandoeu-Nelson MD, Xia Y, Li J, Neuffer MG, Schnabel PS: Unequal sister chromatid and homolog recombination at a tandem duplication of the A1 locus in maize. Genetics 2006, 173(4):2211–2226.

38. Board J, Shone AC, Jeffreys AJ: Meiotic recombination and flanking markers exchange at the highly unstable human minisatellite CEB1 (D2S990). Am J Hum Genet 2000, 67(2):333–344.

39. Board J, Boudet A, Yardley J, Dubrova Y, Jeffreys AJ: Influence of array size and homogeneity on minisatellite mutation. EMBO J 1998, 17(12):3495–3502.

40. Kidd JM, Graves T, Newman TL, Fulton R, Hayden HS, Malig M, Kalikci K, Kaul R, Wilson RK, Eichler EE: A human genome structural variation sequencing resource reveals insights into mutational mechanisms. Cell 2010, 143(3):837–847.

41. Nottle X, Huddleston J, ORoak BJ, Antonacci F, Fichera M, Romano C, Shendure J, Eichler EE: Rapid and accurate large-scale genotyping of duplicated genes and discovery of interlocus gene conversions. Nat methods 2013, 10(9):903–909.

42. Jeffreys AJ, Tamaki K, MacLeod A, Mondton DG, Neil DL, Armour JA: Complex gene conversion events in germline mutation at human minisatellites. Nat Genet 1994, 6(2):136–145.

43. Aldhous MC, Abu Bakar S, Prescott NJ, Falla R, Soo K, Mansfield JC, Mathew CG, Satang J, Armour JAL: 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(24):4930–4938.

44. Craddock N, Hurles ME, Cardin N, Pearson RD, Plagnol V, Robson S, Vulcic D, Barnes C, Conrad DF, Giannoulou E, Holmes C, Marchini JL, Stirrup K, Tobin MD, Wain LV, Yau C, Aerts J, Ahmad T, Andrews TD, Arbuky H, Attwood A, Anton A, Ball SG, Balmforth AJ, Barret JC, Barros I, Barton A, Bennet AJ, Bhaskar S, Blazczyk K, et al: Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. Nature 2010, 464(7289):713–718.

45. 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.

46. Jeffreys AJ, Tamaki K, MacLeod A, Mondton DG, Neil DL, Armour JA: Complex gene conversion events in germline mutation at human minisatellites. Nat Genet 1994, 6(2):136–145.

47. Black et al. BMC Genomics 2014, 15:614 http://www.biomedcentral.com/1471-2164/15/614

Submit your next manuscript to BioMed Central and take full advantage of:

• Convenient online submission
• Thorough peer review
• No space constraints or color figure charges
• Immediate publication on acceptance
• Inclusion in PubMed, CAS, Scopus and Google Scholar
• Research which is freely available for redistribution

Submit your manuscript at www.biomedcentral.com/submit
doi:10.1186/1471-2164-15-614
Cite this article as: Black et al.: Inferming mechanisms of copy number change from haplotype structures at the human DEFA1A3 locus. BMC Genomics 2014 15:614.