A method to assess linkage disequilibrium between CNVs and SNPs inside copy number variable regions

Nathan E. Wineinger1*, Nicholas M. Pajewski2 and Hemant K. Tiwari1* 

1 Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL, USA 
2 Department of Biostatistical Sciences, Wake Forest University Health Sciences, Winston-Salem, NC, USA 

INTRODUCTION

Examination of linkage disequilibrium (LD) between single nucleotide polymorphisms (SNPs) has played a key role in our understanding of worldwide patterns of genetic variation, including determining the extent of haplotype diversity (Conrad et al., 2006), detecting regions of positive selection (Sabeti et al., 2007), and guiding the design of most current genotyping arrays through the selection of appropriate haplotype tagging SNPs. Traditional pairwise metrics of LD, including \( r^2 \), \( D^\prime \), and \( D^\prime \prime \), have been designed to quantify the degree of non-independence between neighboring genetic polymorphisms (Lewontin and Kojima, 1960; Lewontin, 1964; Hill and Robertson, 1968). With the current understanding that copy number variation (CNV) also significantly contributes to genetic polymorphisms; (2) differ based upon the arbitrary choice of SNP allele. Based on this result, we modify traditional techniques to appropriately quantify the covariance in the case of SNPs residing within CNVRs.

MATERIALS AND METHODS

We begin with a brief review of current statistical metrics for the quantification of LD, discuss their performance in the presence of CNV, and conclude with our proposed statistics based on CNV–SNP covariance.

REVIEW OF CURRENT LD METRICS FOR SNPs AND CNVs

In accordance with current LD metrics, let \( X \) denote the integer copy number state for a CNVR on a single maternal/paternal chromosome or haploid, where we assume for simplicity that \( X \) can take three values representing a deletion (0), normal copy number (1), and duplication (2). Similarly define \( Y \) as the count of reference alleles at a SNP on the same chromosome, where we arbitrarily label the SNP alleles as A (reference) and B. The marginal probability distributions for \( X \) and \( Y \) can then be defined as:

\[
X = \begin{cases} 
0 & \text{with } P(X = 0) = f_0, \\
1 & \text{with } P(X = 1) = f_1, \\
2 & \text{with } P(X = 2) = f_2 \end{cases} \quad \text{and} \\
Y = \begin{cases} 
1 & \text{with } P(Y = 1) = f_A, \\
0 & \text{with } P(Y = 0) = f_B. 
\end{cases}
\]  

(1)
Assuming that the joint frequencies \((f_{x,y})\) are known, the covariance between \(X\) and \(Y\) can be written as:

\[
\text{Cov}(X,Y) = \sum_{x,y} xy \cdot f_{x,y} - \left(\sum_x x \cdot f_x\right) \left(\sum_y y \cdot f_y\right).
\]

We consider this covariance between CNVs and SNPs in the following four scenarios.

**Scenario 1a:** The SNP is outside a CNVR (exterior SNP) that contains a normal (one copy) variant and deletion (zero copies). Then:

\[
\text{Cov}(X,Y) = f_{1,A} - f_x f_y = -(f_{1,B} - f_y f_b).
\]

**Scenario 1b:** The SNP is outside a CNVR (exterior SNP) that contains a normal (one copy) variant and duplication (two copies). Then:

\[
\text{Cov}(X,Y) = f_{2,A} - f_x f_y = -(f_{2,B} - f_y f_b).
\]

In both of the above scenarios, the covariance between the CNV and SNP will appropriately be zero when \(X\) and \(Y\) are independent (i.e., the joint frequency is equivalent to the product of the marginal frequencies). Also, any inference concerning the relationship between the CNV and SNP does not depend on as the choice of reference allele, since only the direction of the covariance differs. Given these features, traditional measurements of LD between CNVs and SNPs are sufficient for exterior SNPs.

**Scenario 2a:** The SNP is inside a CNVR (interior) that contains a normal (one copy) variant and deletion (zero copies). Table 1 provides definitions of CNV–SNP allele frequencies based on haploid, three copy number state model (zero to two copies per haploid). In situations where the SNP lies within the CNVR, SNP allele counts are dependent on copy number state. For example, whenever a deletion is present, both \(X\) and \(Y\) must be equal to zero. Thus,

\[
\text{Cov}(X,Y) = f_{1,A} (1 - f_1) = -f_{1,B} (1 - f_1).
\]

**Scenario 2b:** The SNP is inside a CNVR (interior) that contains a normal (one copy) variant and duplication (two copies). This final scenario represents the most complex case. The sample space of \(Y\) needs to change to reflect the possibility of zero to two copies of the \(A\) allele. Namely:

\[
Y = \begin{cases} 
0 & \text{with } P(X = 0) = f_0 + f_{1,B} + f_{2,BB}, \\
1 & \text{with } P(X = 1) = f_{1,A} + f_{2,AB}, \\
2 & \text{with } P(X = 2) = f_{2,AA}.
\end{cases}
\]

The covariance then becomes,

\[
\text{Cov}(X,Y) = f_1 (f_{2,AB} + 2f_{2,AA}) + (1-f_1)f_{1,A} \\
\neq -f_1 (f_{1,AB} + 2f_{1,BA}) - (1-f_1)f_{1,B}.
\]

Based on the covariances calculated in scenarios 2a and 2b, we find two undesirable features of current metrics when used to assess LD between interior SNP and CNVs: (1) polymorphic SNPs inside CNVRs will never be uncorrelated with the CNV; and (2) the correlation between variants will differ based upon which SNP allele is considered as the reference. In these scenarios the use of traditional LD measurements could impact association results. Consider a population where a monomorphic SNP lies within a CNVR that includes a moderately frequent deletion (for instance: \(f_0 = 0.1\) and \(f_{1,A} = 0.9\)). Traditional metrics would conclude that the SNP and CNV are in perfect LD; and that any inference based upon the SNP would apply to the CNV. However, in the absence of copy number data, an association analysis based upon the SNP would be completely uninformative – leading to, perhaps, the incorrect conclusion that CNV is also not associated with the trait. In general, we show that high values of \(\rho^2\) between an interior SNP and deletion are obtained whenever the SNP minor allele frequency is low (Figure 1). However, in the absence of CNV data, the same incorrect conclusion would again be applied to the CNV. In these situations we would hope LD measurements would conclude independence. However that is not the case. We also note the result that the correlation between the SNP and CNV depends on the SNP allele considered as the reference. We have provided an example in the results section signifying this property. Together, these features demonstrate that traditional LD metrics are inappropriate when applied to interior SNPs and CNVs.
To address these deficiencies, we now propose a new metric to quantify LD between CNVs and SNPs that functions equivalently to traditional measures for exterior SNPs, and solves these issues for interior SNPs.

DERIVATION OF NEW PROPOSED STATISTIC

We consider a bi-allelic SNP present within a CNVR with three potential haploid copy number states: zero, one, or two copies — although methods here can be expanded to higher copy number, or multiple SNP alleles (Kalinowski and Hedrick, 2001). Define the CNV–SNP allele at this locus to be a combination of the haploid copy number state and nucleotide frequency with two differing, generically labeled SNPs A and B. Then this model can be treated similar to a multiallelic locus with alleles: 0, A, B, AA, AB, and BB; where 0 represents a deletion (Table 1). Combined in pairs, these alleles form a CNV–SNP genotype which provides information on the total number of copies of each nucleotide (Table 2). This model is consistent with those in the majority of copy number calling algorithms for array-based CNV detection (Wang et al., 2007; Korn et al., 2008; Coin et al., 2010). Note, however, that while CNV–SNP genotypes can be inferred from common genotyping platforms (Korn et al., 2008; Coin et al., 2010), the phase, particularly in duplicated regions, may be ambiguous. For example, an AAB genotype may have either of the phased haploid configurations AA/B or AB/A.

We note that in the case of interior SNPs, a deletion should not provide any information on the relationship between the copy number state and SNP allele(s) present. Therefore, let X be the integer haploid copy number state and Y represent the presence of a particular SNP allele, conditional on haploid copy number state not equal to zero, so that:

\[
X = \begin{cases} 
1 \text{ with } P(X=1) = \frac{f_1}{1-f_0}, \\
2 \text{ with } P(X=2) = \frac{f_2}{1-f_0}; \text{ and }
\end{cases}
\]

\[
Y = \begin{cases} 
1 \text{ with } P(Y=1) = \frac{f_1}{1-f_0}, \\
0 \text{ with } P(Y=0) = \frac{f_0}{1-f_0}; \text{ where }
\end{cases}
\]

\[
f_1 = f_{1,AA} + f_{1,AB} + f_{1,BA} = f_{1,AA} + f_{1,AB} + 1/2 f_{1,AB}; \text{ and }
\]

\[
f_0 = f_{0,AA} + f_{0,AB} + f_{0,BA} \quad \text{according to the CNV–SNP allele frequencies listed in Table 1.}
\]

The covariance between X and Y then becomes,

\[
\text{Cov}(X,Y) = (1-f_0)^{-1} \left( f_{1,AA} - f_{1,AB} f_{1,BA} f_{1,AB} / f_0 \right) = (1-f_0)^{-1} \left( f_{1,AB} - f_{1,AA} f_{1,BA} / f_0 \right),
\]

which does not depend on the particular choice of the reference allele. We denote the inner factor in formula [9] as $D_c$, noting its equivalence to Lewontin’s D (Lewontin and Kojima, 1960) in situations for exterior SNPs. Specifically, let

\[
D_c = f_{1,AA} - f_{1,AB} f_{1,BA} / f_0,
\]

where $f_0^*$ = 0 for exterior SNPs, $f_0$ for interior SNPs.

Similar to $D$, the range of values for $D_c$ is difficult to interpret without proper scaling. Therefore, we propose a method nearly identical to the construction of $D$ (Lewontin, 1964). Define the maximum value that $D_c$ can take based upon allele frequencies as $D_c^\text{max}$. Then:

\[
D_c^\text{max} = \begin{cases} 
\min \left( f_{1,AB} f_{1,AA} / (1-f_0), f_{1,AB} f_{1,BA} / (1-f_0) \right) \quad \text{when } D_c \geq 0,
\max \left( -f_{1,AB} f_{1,AA} / (1-f_0) - f_{1,AB} f_{1,BA} / (1-f_0) \right) \quad \text{when } D_c < 0.
\end{cases}
\]

Finally, let

\[
D_c' = \frac{D_c}{D_c^\text{max}}.
\]

Meanwhile, we can also calculate the correlation between X and Y to be:

\[
\rho_{X,Y} = (1-f_0^*) \frac{D_c}{\sqrt{f_1 f_2 f_3 f_4}},
\]

where

| CNV–SNP genotype | Haploid configuration | Frequency |
|------------------|----------------------|-----------|
| 0/0              | 0/0                  | $f_0^2$   |
| A/0              | 2$f_1 f_2$           |           |
| B/0              | 2$f_1 f_2$           |           |
| AA/0             | $f_1^2$              |           |
| AB/0             | 2$f_1 f_2$           |           |
| BB/0             | $f_2^2$              |           |
| AAA              | 2$f_1 f_2$           |           |
| AAB              | 2$f_1 f_2$           |           |
| ABB              | $f_2^2$              |           |
| BBB              | 2$f_1 f_2$           |           |
| AAAA             | $f_1^2$              |           |
| AAAB             | 2$f_1 f_2$           |           |
| AABB             | $f_1^2 f_2$          |           |
| ABAB             | 2$f_1 f_2$           |           |
| ABBB             | $f_2^2$              |           |
| BABB             | 2$f_1 f_2$           |           |

Frequency estimates are based upon haploid configurations falling into their appropriate Hardy–Weinberg equilibrium proportions.
or, alternatively:
\[
\rho_{k,Y}^2 = \left(1 - f_0^k\right)^2 \frac{D_C^2}{f_{S,T}f_{A,B}f_{B}} = r_k^2. \tag{14}
\]

We again note that \(D_C\) and \(r_k^2\) are identical to the traditional LD measurements \(D'\) and \(r^2\), respectively, for exterior SNPs; and both are an appropriate measurement for interior SNPs.

**ESTIMATION OF CNV–SNP ALLELE FREQUENCIES**

Calculation of \(D_C\) and \(r_k^2\) is straightforward when the CNV–SNP haplotype frequencies are known. However, current methods for array-based genotype/CNV calling do not directly infer the haplotype configuration (phase), though methods for estimating this configuration have been recently proposed (Kato et al., 2008; Su et al., 2010). Here we present a novel method to estimate CNV–SNP allele frequencies based on unphased CNV–SNP genotypes. The method is a direct result of an EM algorithm and nearly identical in construction to the gene-counting, allele frequency estimation method is a direct result of an EM algorithm and nearly identical to the gene-counting, allele frequency estimation procedure in Ceppellini et al. (1955) and Smith (1957). Consider a CNVR with CNV–SNP haplotype configurations \(S/T\) such that \(S, T \in \{0, A, B, AA, AB, BB\}\). In the E-step, haplotype configuration counts are estimated based on the expected counts from estimated CNV–SNP allele frequencies. That is, for each CNV–SNP haplotype configuration \(S/T\):

\[
N_{S,T} = N_{ST} \left( \frac{f_{S,T}}{f_{ST,A}} \right) \tag{15}
\]

where \(N_{ST}\) is number of CNV–SNP genotypes that could possibly result in an \(S/T\) haplotype configuration, \(f_{S,T,A}\) is the estimated frequency of the \(S/T\) haplotype configuration, and \(f_{ST,A}\) is the estimated frequency of CNV–SNP genotypes that could result in an \(S/T\) haplotype configuration for the \(k\)th iteration. In the M-step, CNV–SNP allele frequencies estimates are updated:

\[
f_{S,T,k+1} = \frac{2N_{S,T} + \sum N_{S,T} \sum N_{S,T}}{2N} \tag{16}
\]

as well as new CNV–SNP haplotype configuration frequencies estimates:

\[
f_{S,T,k+1} = \begin{cases} f_{S,T,k+1} & \text{if } S=T, \\ 2f_{S,T,k+1}f_{T,A,k+1} & \text{otherwise.} \end{cases} \tag{17}
\]

The algorithm is based upon haplotype configurations falling into their appropriate Hardy–Weinberg equilibrium proportions. As a result, this approach may perform poorly in de novo mutation hot-spots and CNVs found only in somatic cells.

**RESULTS**

We provide calculations of \(r_k^2\) for various CNV–SNP allele frequencies and compare them to the traditional measurements for SNPs inside CNVRs (Table 3). We define, \(r_k^2\) and \(r_k^2\) are the traditional metrics of LD using SNP allele A or B as the reference allele, respectively. Note how vastly different results can be obtained depending on which allele is used as the reference. The value of \(r_k^2\) is the same irrespective of SNP allele considered as the reference allele.

We theoretically demonstrated how current metrics of LD are inappropriate in certain cases and proposed a new method that solves these issues. Note that the CNV–SNP allele frequencies are critical in calculating \(r_k^2\). We evaluated our method for estimating CNV–SNP allele frequencies via an EM algorithm, as described above in the methods section, using a simulation procedure. These results are provided in Table 4. In summary, our metric accurately and precisely measures SNP/CNV covariance, regardless of the location of the SNP and type of CNV. In particular, high values of \(r_k^2\) will always lead to a proper conclusion about role of CNVs from the study of SNPs. In CNVRs that only include a deletion, our proposed method will always correctly assign independence between interior SNPs and the CNV. Meanwhile, in duplicated regions our metric will provide a value that appropriately quantifies the correlation between SNP allele(s) and the number of copies present.
Table 4 | Copy number variant–single nucleotide polymorphism (CNV–SNP) allele frequency estimation procedure diagnostics based upon 1,000 simulations of a sample size of 1,000 (2,000 haploids) and various CNV–SNP allele frequencies.

| Type                | Simulated frequency | Mean difference |
|---------------------|---------------------|-----------------|
| No CNVs             |                     |                 |
|                     | $f_{AA} = 0.5$      | $0^*$           |
|                     | $f_{BA} = 0.5$      |                 |
| Deletion only (1)   |                     |                 |
| $f_{A} = 0.1$       | $0^*$               |                 |
| $f_{AA} = 0.5$      |                     |                 |
| $f_{BA} = 0.4$      |                     |                 |
| Deletion only (2)   |                     |                 |
| $f_{A} = 0.5$       | $0^*$               |                 |
| $f_{AA} = 0.3$      |                     |                 |
| $f_{BA} = 0.2$      |                     |                 |
| Duplication only (1)|                     |                 |
| $f_{A} = 0.5$       | $0.0019$            |                 |
| $f_{AB} = 0.3$      |                     |                 |
| $f_{BB} = 0.2$      |                     |                 |
| Duplication only (2)|                     |                 |
| $f_{A} = 0.3$       | $0.0030$            |                 |
| $f_{AA} = 0.2$      |                     |                 |
| $f_{AB} = 0.1$      |                     |                 |
| $f_{BB} = 0.1$      |                     |                 |
| Duplication only (3)|                     |                 |
| $f_{A} = 0.3$       | $0^*$               |                 |
| $f_{AB} = 0.3$      |                     |                 |
| $f_{BB} = 0.2$      |                     |                 |
| Multiallelic (1)    |                     |                 |
| $f_{A} = 0.1$       | $0.0040$            |                 |
| $f_{AA} = 0.3$      |                     |                 |
| $f_{AB} = 0.1$      |                     |                 |
| $f_{BB} = 0.2$      |                     |                 |
| Multiallelic (2)    |                     |                 |
| $f_{A} = 0.3$       | $0.0028$            |                 |
| $f_{AB} = 0.2$      |                     |                 |
| $f_{BB} = 0.1$      |                     |                 |

Mean difference represents the mean difference between the true and estimated CNV–SNP allele frequencies. $0^*$: Less than $1 \times 10^{-5}$ for each allele. Haploid configurations can nearly be unambiguously assigned based upon the given three-state haploid model.

DISCUSSION
We have provided rationale for why current metrics used to assess LD between CNVs and interior SNPs are inappropriate. Given that difficulties arise only for these SNPs, one potential solution, as previous studies have done, would be to rely upon exterior SNPs for tagging CNVs. Though this approach has been successful for deletions, duplications tend to be in very low LD with exterior SNPs (Kato et al., 2010). It is possible that duplicate copies are not simply positioned in tandem next to a neighboring SNP in relation to the reference genome. The more extreme case arises when a duplicate copy has been translocated onto a different chromosome. In this situation an exterior SNP will be completely unlinked to the translocated duplicate. However, interior SNPs will segregate within the duplicate – particularly if this copy is not suitably matched for recombination. A similar argument can be made for duplicate segments inserted downstream of its reference location. A larger physical distance between the duplicated copies and an exterior SNP allows for a greater probability of recombination to eliminate LD. However, this distance will be irrelevant in regards to the allelic content within the duplicated genomic segment.

We have included a new method to quantify LD between CNVs and SNPs which provides accurate estimates for interior SNPs and defaults to the traditional measurements otherwise. As our methods require knowledge of CNV–SNP allele frequencies, we have provided an estimation procedure that performs well under a wide range of scenarios. We hope CNV researchers, particularly those hoping to draw conclusions about CNVs from SNPs, will use this method to identify tagging SNPs which may or may not exist within the CNV boundary.

WEB RESOURCES
R code to measure $r^2_{SNP}$ and CNV–SNP allele frequencies from CNV–SNP genotypes is available from the corresponding author upon request.

ACKNOWLEDGMENTS
The work is supported in part by the University of Alabama at Birmingham’s Alumni Associations’ Marie and Emmett Carmichael Fund for Graduate Students in Biosciences, and NIH grants T32 HL-079888 and T32 HL-072757. The opinions expressed herein are those of the authors and not necessarily those of the NIH or any organization with which the authors are affiliated.

REFERENCES
Aitman, T. J., Dong, R., Vyse, T. J., Norsworthy, P. L., Johnson, M. D., Smith, J., Mangion, J., Robertson, Lowe, C., Marshall, A. J., Petretto, E., Hodges, M. D., Bhangal, G., Patel, S. G., Sheehan-Rooney, K., Duda, M., Cook, P. R., Evans, D. J., Domín, J., Flint, J., Boyle, J. J., Pasey, C. D., and Cook, H. T. (2006). Copy number polymorphism in FcgR3 predisposes to glomerulonephritis in rats and humans. Nature 439, 851–855.

Ceppellini, R., Siniscalco, M., and Smith, C. A. B. (1955). The estimation of gene frequencies in a random-mating population. Ann. Hum. Genet. 20, 97–115.

Coi, L. J., Ascher, J. E., Walters, R. G., Moustafa, J. S., de Smith, A. J., Sladek, R., Balding, D. J., Frooguel, P., and Blakemore, A. I. (2010). cnvHap: an integrative population and haplotype-based multipplatform model of SNPs and CNVs. Nat. Methods 7, 541–546.

Conrad, D. F., Jacobsson, M., Coop, G., Wen, X., Wall, J. D., Rosenberg, N. A., and Pritchard, J. K. (2006). A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. Nat. Genet. 38, 1251–1260.

Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gojoburne, O., Zhang, Y., Aerts, J., Andrews, T. D., Barnes, C., Campbell, P., Fitzgerald, T., Hu, M., Ihm, C. H., Kristiansson, K., Macarthur, D. G., Macdonald, J. R., O’Doherty, I., Pang, A. W., Robson, S., Stirrups, K., Valsecchi, A., Walter, K., Wei, J., Wellcome Trust Case Control Consortium, Tyler-Smith, C., Carter, N. P., Lee, C., Scherle, S. W., and Hurles, M. E. (2010). Origins and functional impact of copy number variation in the human genome. Nature 464, 704–712.

Eichler, E. F., Flint, J., Gibson, G., Kong, A., Leal, S. M., Moore, J. H., and Nadeau, J. H. (2010). Missing heritability and strategies for finding the underlying causes of complex diseases. Nat. Rev. Genet. 11, 446–450.

Gonzalez, E., Kulkarni, H., Bolivar, H., Mangano, A., Sanchez, R., Catano, G., Nibbs, R. J., Freedman, B. I., Quinones, M. P., Banham, M. J., Murthy, K. K., Rovin, B. H., Bradley, W., Clark, R. A., Anderson, S. A., O’Connell, R. J., Agar, B. K., Ahuja, S. S., Bologna, R., Sen, L., Dolan, M. J., and Ahuja, S. K. (2005). The influence of CCL3L1
gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science 307, 1434–1440.

Hill, W. G., and Robertson, A. (1968). The effects of inbreeding at loci with heterozygote advantage. Genetics 60, 615–628.

Kalinowski, S. T., and Hedrick, P. W. (2001). Estimation of linkage disequilibrium for loci with multiple alleles: basic approach and an application using data from bighorn sheep. Heredity 87, 698–708.

Kato, M., Awagushi, T., shikawa, S., Umeda, T., Nakamichi, R., Shapero, M. H., Jones, K. W., Nakamura, Y., Aburatani, H., and Tsunoda, T. (2010). Population-genetic nature of copy number variants in the human genome. Hum. Mol. Genet. 19, 761–773.

Kato, M., Nakamura, Y., and Tsunoda, T. (2008). An algorithm for inferring complex haplotypes in a region of copy-number variation. Ann. Hum. Genet. 83, 157–169.

Korn, J. M., Kuruvilla, F. G., McCarthy, S. A., Wysoker, A., Nensley, J., Cawley, S., Hubbell, E., Veitch, J., Collins, P. J., Darvisi, K., Lee, C., Nizami, M. M., Gabrieli, S. B., Purcell, S., Daly, M. J., and Altschuler, D. (2008). Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nat. Genet. 40, 1253–1266.

Leavitt, R. C. (1964). The interaction of selection and linkage. I. General considerations: heterotic models. Genetics 49, 49–67.

Leavitt, R. C., and Kojima, K. (1960). The evolutionary dynamics of complex polymorphisms. Evolution 14, 453–470.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hinderer, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., McCarthy, M. I., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E. H., McCarthy, S. A., Gaudet, R., Schaffner, S. F., Lande, R. E., International HapMap Consortium, Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., Belmont, J. W., Boudreau, A., Hardenbol, P., Leal, S. M., Pastersnak, S., Wheeler, D. A., Willis, T. D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Wu, L., Li, C., Liu, S., Pan, H., Tang, X., Wang, K., Brooks, L. D., He, X., Zhang, Q., Zhao, H., Zhao, H., Zhou, L., Gabriel, S. B., Ray, B., Blumenstein, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R. C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaura, L., Altshuler, D., Shen, Y., Yao, Z., Zhao, H., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Y., Wu, W., Wang, W., Wang, Y., Wang, X., Xiong, X., Xu, L., Weye, M. M., Tsai, S. K., Xue, H., Wong, T. J., Gaber, L., Fan, J.B., Gudmundson, K., Murray, S. S., Oliphant, A. R., Chee, M., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J.F., Philips, M. S., Rouy, S., Salié, C., Verner, A., Hudson, T. J., Kwok, P. Y., Cai, D., Koholdt, D. C., Miller, R. D., Pavlikova, L., Taillon-Miller, P., Xiao, M., Tsai, L. C., Mak, W., Song, Y. Q., Tam, P. K., Nakamura, Y., Kanahashi, R., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C. P., Delgado, M., Derrmitzakis, E. T., Gwilliam, R., Hunter, S., Morrison, P., Powell, D., Stranger, B. E., Whittaker, E. T., Gwilliam, R., Hunt, S., Morrison, P., Andrews, T. D., Arbury, H., Attwood, D., Bain, L. V., Yau, C., Aerts, J., Ahmad, T., Andrews, T. D., Attwood, A., Auton, A., Ball, S. G., Balmforth, A. J., Barrett, I. C., Barnes, L., Barton, A., Bennett, A. L., Bhaskar, S., Blazyczki, K., Bowes, J., Brand, O. J., Braund, P. S., Bredin, E., Breen, G., Brown, M. J., Bruce, I. N., Bull, J., Burren, O. S., Burton, J., Byrones, J., Caesar, S., Clee, C. M., Coffey, A. J., Connell, J. M., Cooper, J. D., Dominiczak, A. F., Downes, K., Drummond, H. E., Dudakia, D., Dunham, A., Ebbs, E., Eccles, D., Edkins, S., Edwards, C., Elliot, A., Emery, P., Evans, D. M., Evans, G., Eyre, S., Farmer, A., Ferrier, I. N., Feuk, L., Fitzgerald, T., Flynn, E., Forbes, A., Forty, L., Franklin, J. A., Freathy, R. M., Gibbs, P., Gilbert, P., Guo, J., Gordon-Smith, K., Gray, E., Green, E., Groves, C. J., Grozeva, D., Gwilliam, R., Hall, A., Hammond, N., Hardy, M., Harris, P., Hassanali, N., Hemberla, H. S., Hinchliffe, J., Hinson, A., Hocking, L., Howard, E., Howard, P., Howson, J. M., Hughes, D., Hunt, S., Isaacs, J. D., Jain, M., Jewell, D. P., Johnson, T., Jolley, J. D., Jones, I. R., Jones, L. A., Kirov, G., Langford, C. F., Lungo-Aland, H., Lathrop, G. M., Lee, J., Lee, K. L., Lees, C., Lewis, K., Lindgren, C. M., Maisuria-Armer, M., Maller, J., Mansfield, J., Martin, P., Massey, D. C., McArdle, W. L., McGriffin, P., McElroy, K. E., Mentzer, A., Mimmacak, M., Morgan, A. E., Morris, A. P., Mowat, C., Myers, S., Newman, W., Ninmoo, E. R., O’Donovan, M. C., Onipila, A., Onyiah, I., Ovington, N. R., Owen, M. J., Pali, K., Parnell, K., Pernet, D., Perry, J. R., Phillips, A., Pinto, D., Precott, N. I., Prokopenko, I., Quail, M. A., Rafels, S., Rayner, N. W., Redon, R. B., Reid, D. M., Renwick, S. M., Robertson, S. C., Robertson, D. S., Sambrook, J., Sanderson, J. D., Schilenberg, H., Scott, C. E., Scott, R., Seal, S., Shaw-Hawkins, S., Shields, B. M., Simmons, M. J., Smyth, D. J., Somaskantharajah, E., Spanova, K., Steer, S., Stephens, J., Stevens, H. E., Stone, M. A., Su, Z., Symmons, D. P., Thompson, J. R., Thomson, W., Travers, M. E., Turnbull, C., Valiesa, A., Walker, M., Walker, N. M., Wallace, C., Warren-Perry, M., Watkins, N. A., Webster, J., Weeden, M. N., Wilson, A. G., Woodburn, M., Wordsworth, B. P., Young, A. H., Zeggini, E., Carter, N. P., Frayling, T. M., Lee, C., McVeag, G., Munroe, P. B., Palotie, A., Sawyer, S. J., Scherer, S. W., Strachan, D. P., Tyler-Smith, C., Brown, M. A., Burton, P. R., Cauliff, M. J., Compston, A., Farrall, M., Gough, S. C., Hall, A. S., Hattersley, A. T., Hill, A. V., Mathew, C. G., Pembrey, M., Satsangi, J., Stratton, M. R., Worthington, J., Deloukas,
P., Duncanson, A., Kwiatkowski, D. P., McCarthy, M. I., Ouwehand, W., Parkes, M., Rahman, N., Todd, J. A., Samani, N. I., and Donnelly, P. (2010). Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. Nature 464, 713–720.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 20 January 2011; accepted: 31 March 2011; published online: 25 April 2011.

**Citation:** Wineinger NE, Pajewski NM and Tiwari HK (2011) A method to assess linkage disequilibrium between CNVs and SNPs inside copy number variable regions. Front. Gene. 2:17. doi: 10.3389/fgene.2011.00017

This article was submitted to Frontiers in Statistical Genetics and Methodology, a specialty of Frontiers in Genetics.

Copyright © 2011 Wineinger, Pajewski and Tiwari. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.