

PROCEEDINGS

Open Access



Multipoint association mapping for longitudinal family data: an application to hypertension phenotypes

Yen-Feng Chiu^{1*}, Chun-Yi Lee¹ and Fang-Chi Hsu²

From Genetic Analysis Workshop 19
Vienna, Austria. 24-26 August 2014

Abstract

It is essential to develop adequate statistical methods to fully utilize information from longitudinal family studies. We extend our previous multipoint linkage disequilibrium approach—simultaneously accounting for correlations between markers and repeat measurements within subjects, and the correlations between subjects in families—to detect loci relevant to disease through gene-based analysis. Estimates of disease loci and their genetic effects along with their 95 % confidence intervals (or significance levels) are reported. Four different phenotypes—ever having hypertension at 4 visits, incidence of hypertension, hypertension status at baseline only, and hypertension status at 4 visits—are studied using the proposed approach. The efficiency of estimates of disease locus positions (inverse of standard error) improves when using the phenotypes from 4 visits rather than using baseline only.

Background

Approaches for analyzing longitudinal family data have been categorized into 2 groups [1]: (a) first summarizing repeated measurements into 1 statistic (eg, a mean or slope per subject) and then using the summarized statistic as a standard outcome for genetic analysis; or (b) simultaneous modeling of genetic and longitudinal parameters. In general, joint modeling is appealing because (a) all parameter estimates are mutually adjusted, and (b) within- and between-individual variability at the levels of gene markers, repeat measurements, and family characteristics are correctly accounted for [1].

The semiparametric linkage disequilibrium mapping for the hybrid family design we developed previously [2] uses all markers simultaneously to localize the disease locus without making an assumption about genetic mechanism, except that only 1 disease gene lies in the region under study. The advantages of this approach are (a) it does not require the specification of an underlying genetic model,

so estimating the position of a disease locus and its standard error is robust to a wide variety of genetic mechanisms; (b) it provides estimates of disease locus positions, along with a confidence interval for further fine mapping; and (c) it uses linkage disequilibrium between markers to localize the disease locus, which may not have been typed. We extended this approach to map susceptibility genes using longitudinal nuclear family data with an application to hypertension. Four different outcomes were used based on the proposed method: (I) ever having hypertension (“Ever”), (II) incidence event with status changed from unaffected to affected (“Progression”), (III) first available visit as baseline only (“Baseline”), and (IV) all available time points (“Longitudinal”). We compared the estimates of the disease locus positions, their standard errors, the genetic effect estimate at the disease loci, and their significance for the 4 phenotypes to examine the efficiency gained from using repeated longitudinal phenotypes.

Methods

Genome-wide genotypes and phenotype data

Association mapping was conducted on chromosome 3 of the genome-wide association study (GWAS) data. A

* Correspondence: yfchiu@nhri.org.tw

¹Institute of Population Health Sciences, National Health Research Institutes, Miaoli, 35053, Taiwan, Republic of China

Full list of author information is available at the end of the article

total of 65,519 single-nucleotide polymorphisms (SNPs) included in 1095 genes were genotyped on chromosome 3 for 959 individuals from 20 original pedigrees in Genetic Analysis Workshop 19 (GAW19). Of these individuals, there were 178 (38 %) affected offspring out of 469 offspring for phenotype (I) “Ever”; 130 (31 %) out of 421 offspring for phenotype (II) “Progression”; 64 (11 %) out of 600 offspring for phenotype (III) “Baseline”; and 60 (11 %) out of 565 offspring to approximately 85 (45 %) out of 189 offspring across the 4 visits (or 87 [21.63 %] out of 402 offspring on average) for phenotype (IV) “Longitudinal” (Table 1). To compare phenotypes (I) and (II), only individuals with at least 2 measurements were included in the “ever” phenotype. PedCut [3] was used to split large pedigrees with members more than 20 members into nuclear pedigrees. Consequently, we analyzed a total of 138 pedigrees with 1,495 individuals (the IDs for missing parents were added to form trios). In divided pedigrees, the nuclear families contained between 3 and 25 individuals. Five SNPs were removed because they failed the test of Hardy-Weinberg equilibrium (HWE) (p value $< 10^{-4}$). The HWE test was performed using PLINK 1.07 [4] based on 56 unrelated subjects. (For information on PLINK, see <http://pngu.mgh.harvard.edu/purcell/plink/>.) A total of 22,056 genotypes from various SNPs with genotyping errors (genotyping error rate was around 3.51×10^{-4}) were further excluded by the MERLIN 1.1.2 computing package (see <http://www.sph.umich.edu/csg/abecasis/merlin/tour/linkage.html>). None of the covariates was adjusted for in this approach.

Multipoint linkage disequilibrium mapping

Suppose M markers were genotyped in the region R at locations of $0 \leq t_1 < t_2 < \dots < t_M \leq T$. We assume there are 2 alleles per marker. With $H(t)$ being the target allele at marker position t , and $h(t)$ being the nontarget allele, we define

$$Y_1^{D_{k_{il}}}(t) = \begin{cases} 1 & \text{if the transmitted paternal allele at } t \text{ is } H(t) \\ 0 & \text{if the transmitted paternal allele at } t \text{ is } h(t) \end{cases},$$

$$Y_2^{D_{k_{il}}}(t) = \begin{cases} 1 & \text{if the nontransmitted paternal allele at } t \text{ is } H(t) \\ 0 & \text{if the nontransmitted paternal allele at } t \text{ is } h(t) \end{cases},$$

for the affected offspring $D_{k_{il}}$,

and

$$Y_1^{\bar{D}_{k_{il}}}(t) = \begin{cases} -1 & \text{if the transmitted paternal allele at } t \text{ is } H(t) \\ 0 & \text{if the transmitted paternal allele at } t \text{ is } h(t) \end{cases},$$

$$Y_2^{\bar{D}_{k_{il}}}(t) = \begin{cases} -1 & \text{if the nontransmitted paternal allele at } t \text{ is } H(t) \\ 0 & \text{if the nontransmitted paternal allele at } t \text{ is } h(t) \end{cases},$$

for the unaffected offspring $\bar{D}_{k_{il}}$. Then, we define the preferential transmission statistic $Y_{T_{k_{il}}}(t) = Y_1^{D_{k_{il}}}(t) - Y_2^{D_{k_{il}}}(t)$ for the paternal side and $X_{T_{k_{il}}}(t) = X_1^{D_{k_{il}}}(t) - X_2^{D_{k_{il}}}(t)$ for the maternal side for a trio; similarly, the preferential transmission statistic $Y_{U_{k_{il}}}(t) = Y_1^{\bar{D}_{k_{il}}}(t) - Y_2^{\bar{D}_{k_{il}}}(t)$ and $X_{U_{k_{il}}}(t) = X_1^{\bar{D}_{k_{il}}}(t) - X_2^{\bar{D}_{k_{il}}}(t)$ for an unaffected trio for both parental sides, respectively, where $k_{il} = 1, \dots, N_{1il}$ (for unaffected), $N_{1il}(N_{2il})$ is the number of affected (unaffected) offspring in the family i at the l^{th} time point, $i = 1, \dots, n$, $l = 1, \dots, L$ ($L = 1$ or 4 in this study).

The expectation of the statistic is $\mu_{1k_{ij}}(\delta, \pi) = E[Y_{T_{k_{il}}}(t_j) | \Phi_1] = (1 - 2\theta_{t_j, \tau}) C (1 - \theta_{t_j, \tau})^N \pi_j$ for case-parent trios and $\mu_{2k_{ij}}(\delta, \pi) = E[Y_{U_{k_{il}}}(t_j) | \Phi_2] = (1 - 2\theta_{t_j, \tau}) C^* (1 - \theta_{t_j, \tau})^N \pi_j$ for control-parent trios, where $\theta_{t_j, \tau}$ is the recombination fraction between marker position t_j and disease locus position τ , the recombination fraction Θ is a parametric function of the parameter of primary interest (τ , the physical position of the functional variant), N is the number of generations since the initiation of the disease variant, Φ_1 denotes the event that the offspring is affected, Φ_2 represents the event that the offspring is unaffected, $C = E[Y_{T_{k_{il}}}(t) | \Phi_1] = E[X_{T_{k_{il}}}(t) | \Phi_1]$, $C^* = E[Y_{U_{k_{il}}}(t) | \Phi_2] = E[X_{U_{k_{il}}}(t) | \Phi_2]$, $\delta = (\tau, N, C, C^*)$ is the vector of parameters, and $\pi_j = \Pr[h(t_j) | h(\tau)]$. $\mu_{1k_{ij}}$ is the probability for an affected offspring to receive a target allele, and $-\mu_{2k_{ij}}$ is the probability for an unaffected offspring to receive a target allele. The statistic $Z_{1k_{ij}} = X_{T_{k_{il}}} + Y_{T_{k_{il}}}$ and $Z_{2k_{ij}} = X_{U_{k_{il}}} + Y_{U_{k_{il}}}$ were used to estimate the parameters. The estimating equations used to solve for parameters δ are:

Table 1 Number of offspring for different phenotypes

	Ever	Progression	Baseline	Visit 1	Visit 2	Visit 3	Visit 4
Affected offspring	178	130	64	60	78	125	85
All offspring	469	421	600	565	426	429	189
Percentage	0.38	0.31	0.11	0.11	0.18	0.29	0.45
Number of nuclear families	174	149	213	203	168	165	79

$$S_1(\delta) = \sum_{i=1}^n \sum_{l=1}^L \sum_{k_{il}=1}^{N_{1il}} \sum_{j=1}^M \frac{\partial \mu_{1k_{il}}(\delta, \hat{\pi}_j)}{\partial \delta} \quad (1)$$

$$Cov^{-1}(Z_{1k_{il}}) \left\{ Z_{1k_{il}} - 2\mu_{1k_{il}}(\delta, \hat{\pi}_j) \right\},$$

$$S_2(\delta) = \sum_{i=1}^n \sum_{l=1}^L \sum_{k_{il}=1}^{N_{2il}} \sum_{j=1}^M \frac{\partial \mu_{2k_{il}}(\delta, \hat{\pi}_j)}{\partial \delta} \quad (2)$$

$$Cov^{-1}(Z_{2k_{il}}) \left\{ Z_{2k_{il}} - 2\mu_{2k_{il}}(\delta, \hat{\pi}_j) \right\},$$

where $\hat{\pi}_j$ is the average of nontransmitted parental alleles in the sample.

The estimating equations were solved iteratively for parameters τ , N , C , and C^* , where τ and C are the 2 parameters of interest. The variance of the disease locus position estimate was estimated to make inferences about the disease locus position (τ) and its genetic effect (C) [2]. Theoretically, the genetic effect of τ , characterized by C , is the transmission probability that the affected offspring will carry the disease allele, H , at τ . Detailed derivations for case-parent trios in a cross-sectional design can be found in Chiu et al. [2, 5]. We

will present the details of this proposed methodology elsewhere.

Gene-based association mapping was conducted for all SNPs on chromosome 3. This approach accounts for correlations between markers and repeated phenotypes within subjects, and correlations between subjects per family. The consistent estimates of hypertension locus position using “Ever” and “Progression” are shown in Table 2 and Fig. 1, while the consistent estimates of hypertension locus position using baseline and longitudinal data (at all 4 visits) are listed in Table 3 and Fig. 2.

Results and discussion

A total of 119 (11 %), 79 (7 %), 49 (4 %), and 42 (4 %) of 1095 genes had a significant genetic effect ($P < 4.57 \times 10^{-5}$ with Bonferroni correction) based on hypertension status at “Ever,” “Progression,” baseline (“Baseline”), and 4 visits (“Longitudinal”), respectively. There are only 3 significantly associated genes ($P \leq 0.05$) for baseline and longitudinal phenotypes duplicated with the significantly associated genes for “Ever” and “Progression” outcomes: *FETUB*, *ILIRAP*, and *C3orf21*. Several hits identified here have been reported from linkage or GWAS studies. Table 2

Table 2 Significant and consistent estimates of disease locus positions and their genetic effects using “Ever” and “Progression” phenotypes

Gene*	Ever			Progression			Previous hits
	$\hat{\tau} \pm SE$	\hat{C}	p Value	$\hat{\tau} \pm SE$	\hat{C}	p Value	
<i>FBLN2</i>	13.6464 ± 0.00026	0.80	8.85 × 10 ⁻⁷	13.6462 ± 0.00030	0.88	2.47 × 10 ⁻⁶	L
<i>C3orf19</i>	14.6810 ± 0.00099	0.34	1.61 × 10 ⁻¹²	14.6802 ± 0.0010	0.34	6.87 × 10 ⁻¹¹	L
<i>C3orf20</i>	14.7245 ± 0.00077	0.51	1.31 × 10 ⁻⁶	14.7244 ± 0.00091	0.45	2.81 × 10 ⁻⁷	L
<i>OSBPL10</i>	31.6853 ± 0.00051	0.41	7.69 × 10 ⁻⁸	31.6856 ± 0.00024	0.62	3.95 × 10 ⁻⁶	LG
<i>CMTM8</i>	32.3186 ± 0.00080	0.56	6.01 × 10 ⁻⁶	32.3183 ± 0.00052	0.70	1.20 × 10 ⁻⁵	
<i>BSN</i>	49.6596 ± 0.00062	0.83	3.42 × 10 ⁻⁸	49.6594 ± 0.00077	0.77	3.58 × 10 ⁻⁶	
<i>RFT1</i>	53.1117 ± 0.0012	0.37	4.07 × 10 ⁻⁶	53.1111 ± 0.0012	0.36	2.54 × 10 ⁻⁵	
<i>ADAMTS9</i>	64.5214 ± 0.00028	0.53	1.05 × 10 ⁻¹¹	64.5216 ± 0.00030	0.54	2.54 × 10 ⁻¹¹	L
<i>EPHA3</i>	89.6014 ± 0.00042	0.80	1.41 × 10 ⁻⁶	89.6018 ± 0.00040	0.89	1.19 × 10 ⁻⁵	
<i>EPHA6</i>	98.2999 ± 0.00047	0.41	3.57 × 10 ⁻⁸	98.2997 ± 0.00052	0.48	7.3 × 10 ⁻⁷	L
<i>C3orf52</i>	113.3097 ± 0.0026	0.62	6.55 × 10 ⁻⁹	113.3088 ± 0.0030	0.58	7.29 × 10 ⁻⁶	L
<i>SIDT1</i>	114.7743 ± 0.00039	0.78	8.45 × 10 ⁻⁷	114.7741 ± 0.00071	0.67	8.46 × 10 ⁻⁶	L
<i>IFT122</i>	130.7107 ± 0.0012	0.57	1.10 × 10 ⁻⁵	130.7118 ± 0.00060	0.71	4.90 × 10 ⁻⁷	
<i>RBP1</i>	140.7325 ± 0.00019	0.65	6.40 × 10 ⁻⁷	140.7345 ± 0.00033	0.42	3.89 × 10 ⁻¹¹	L
<i>PLOD2</i>	147.3469 ± 0.00098	0.34	3.56 × 10 ⁻⁶	147.3469 ± 0.0016	0.34	3.02 × 10 ⁻⁵	L
<i>LEKR1</i>	158.2181 ± 0.00036	0.72	1.66 × 10 ⁻¹⁰	158.2183 ± 0.00043	0.77	4.35 × 10 ⁻¹⁰	L
<i>RSRC1</i>	159.4005 ± 0.00059	0.51	4.35 × 10 ⁻⁶	159.4003 ± 0.00064	0.51	1.64 × 10 ⁻⁵	L
<i>ECT2</i>	174.0021 ± 0.00064	0.88	1.91 × 10 ⁻⁶	174.0022 ± 0.00063	1.00	2.92 × 10 ⁻⁷	L
<i>PEXSL</i>	181.0080 ± 0.0078	0.29	2.99 × 10 ⁻⁵	181.0145 ± 0.013	0.23	5.23 × 10 ⁻⁷	LG
<i>LPP</i>	189.5573 ± 0.00035	0.50	1.71 × 10 ⁻⁶	189.5574 ± 0.00022	0.53	7.05 × 10 ⁻⁶	
<i>OSTN</i>	192.4272 ± 0.0018	0.73	5.33 × 10 ⁻¹⁴	192.4301 ± 0.0012	0.80	4.11 × 10 ⁻⁹	G

\hat{C} , the genetic effect estimate; G, previous GWAS hits; L, previous linkage hits; $\hat{\tau}$, the disease locus position estimate in cM

*Because of space limitations, we list only the 2 phenotypes with consistent estimates for the disease locus positions (the difference between the 2 $\hat{\tau}$ for both phenotypes is less than 10⁻² cM) and significant estimates for the genetic effects (both with $P < 4.57 \times 10^{-5}$, Bonferroni)

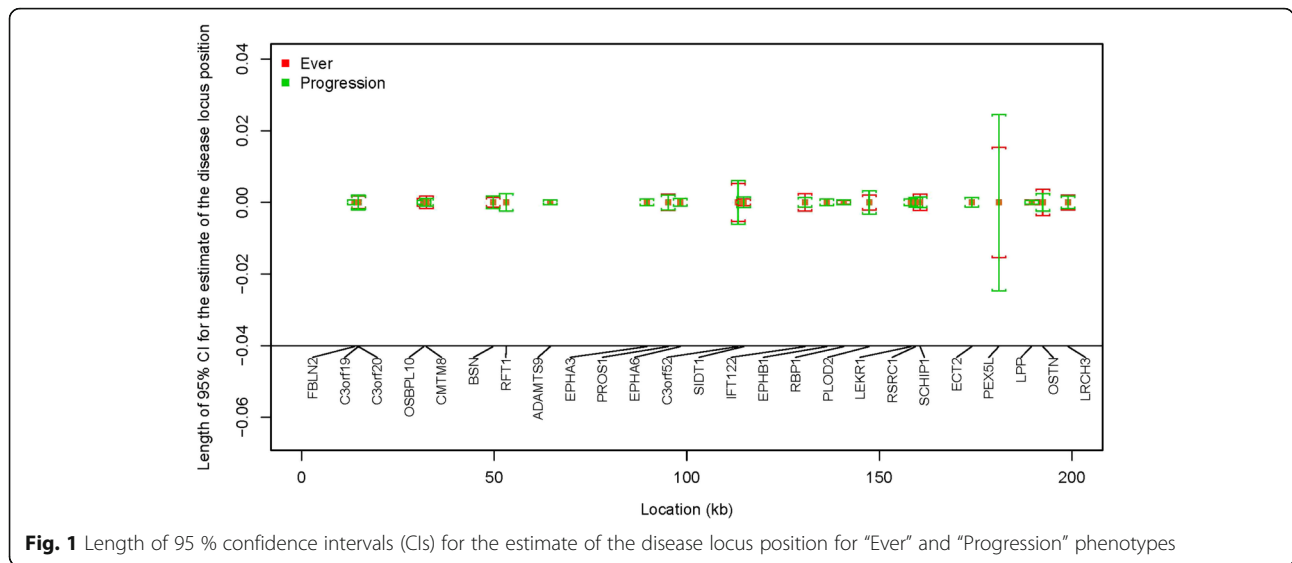


Fig. 1 Length of 95 % confidence intervals (CIs) for the estimate of the disease locus position for “Ever” and “Progression” phenotypes

Table 3 Significant and consistent estimates of disease locus positions and their genetic effects using “Baseline” and “Longitudinal” phenotypes

Gene*	Baseline			Longitudinal			Previous hits
	$\hat{\tau} \pm SE$	\hat{C}	p Value	$\hat{\tau} \pm SE$	\hat{C}	p Value	
GRM7 [†]	7.4917 ± 0.00048	0.44	2.87 × 10 ⁻⁵	7.4871 ± 0.0015	0.75	6.04 × 10 ⁻¹⁴	LG
SLC4A7	27.4521 ± 0.000045	0.30	0.014	27.4520 ± 0.000067	0.30	0.0024	LG
SCN10A	38.7559 ± 0.0089	0.088	0.019	38.7611 ± 0.0018	0.73	0.0022	
AC092058.3	39.5105 ± 0.0020	0.076	0.036	39.5102 ± 0.00024	0.21	0.00022	
LTF	46.4731 ± 0.00059	0.17	0.046	46.4733 ± 0.00045	0.31	0.0099	
NEK4	52.7326 ± 0.00071	0.83	0.00010	52.7277 ± 0.0024	0.86	0.00024	
FAM116A	57.6101 ± 0.00023	0.69	2.58 × 10 ⁻⁶	57.6107 ± 0.00032	0.61	0.011	
LRIG1	66.5968 ± 0.0026	0.28	0.018	66.5961 ± 0.00064	0.60	0.0022	L
TBC1D23	101.5084 ± 0.0011	0.46	0.026	101.5148 ± 0.0010	0.73	0.0011	L
ALCAM	106.7625 ± 0.00069	0.83	0.028	106.7598 ± 0.00041	0.62	0.00013	L
PLCXD2	112.9440 ± 0.00087	0.50	0.0016	112.9422 ± 0.0062	0.48	0.00020	L
LSAMP	117.0676 ± 0.00060	0.43	0.00022	117.0671 ± 0.00025	0.86	0.00012	L
ILDR1	123.2009 ± 0.0011	0.70	0.013	123.2008 ± 0.00098	0.91	0.023	
PDIA5	124.3194 ± 0.0076	0.065	0.0028	124.3225 ± 0.0020	0.68	0.0086	
HPS3	150.3484 ± 0.0016	0.14	1.65 × 10 ⁻⁵	150.3521 ± 0.00063	0.77	0.0080	L
CASRL1	157.2304 ± 0.0037	0.19	0.012	157.2295 ± 0.00094	0.28	0.031	L
C3orf55	158.7595 ± 0.00074	0.90	0.0051	158.7634 ± 0.0012	0.91	3.90 × 10 ⁻⁶	L
IGF2BP2	186.9725 ± 0.00018	0.74	0.041	186.9719 ± 0.00031	1.00	0.031	
FETUB [‡]	187.8470 ± 0.00031	0.38	0.0012	187.8503 ± 0.017	0.042	0.0021	
IL1RAP [‡]	191.8193 ± 0.012	0.074	0.0060	191.8203 ± 0.0012	0.79	4.75 × 10 ⁻⁶	
C3orf21 [‡]	196.2815 ± 0.0036	0.62	<10 ⁻¹⁸	196.2821 ± 0.0011	0.97	0.00057	
KIAA0226	198.9161 ± 0.0038	0.071	0.024	198.9168 ± 0.00076	0.15	0.022	

\hat{C} , the genetic effect estimate; G, previous GWAS hits; L, previous linkage hits; $\hat{\tau}$, the disease locus position estimate in cM

*Displayed are all genes where $p \leq 0.05$

[†]The gene is significant with the Bonferroni correction ($P < 4.57 \times 10^{-5}$) and its P values are 2.31×10^{-6} and 0.00044 for “Ever” and “Progression,” respectively

[‡] The same genes for the “Ever” and “Progression” phenotypes had P values < 0.05 but $> 4.57 \times 10^{-5}$ for the genetic effect estimate

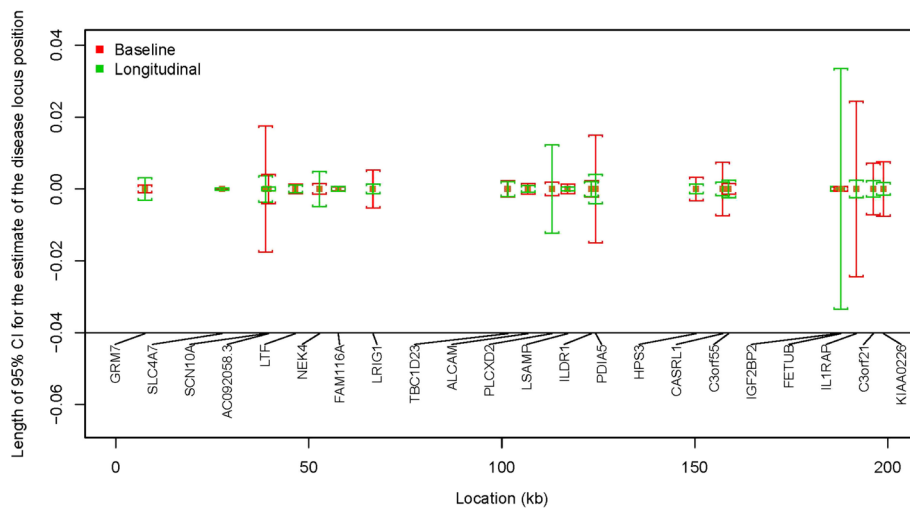


Fig. 2 Length of 95 % confidence intervals (CIs) for the estimate of the disease locus position for “Baseline” and “Longitudinal” phenotypes

shows genes with a significant genetic effect ($P < 4.57 \times 10^{-5}$). Table 3 presents the genes that are significant at a significance level of 0.05. Only 1 gene, *GRM7*, is significant at the level of $P < 4.57 \times 10^{-5}$.

Figures 1 and 2 display the 95 % confidence intervals for the estimate of the hypertension locus position for the 4 phenotypes centered at the estimated disease locus position. The comparison is shown for the genes listed in Tables 2 and 3. The standard errors of the estimates for the disease locus position are smaller in 64 % of the genes based on longitudinal data (Table 3) compared to those based on baseline data. This is because those incidence cases included in “Progression” were also included in the analysis of “Ever.” Only prevalent cases, a relatively small proportion, are additionally included in the analysis of “Ever.” Thus, the results from “Progression” and “Ever” are similar.

Conclusions

Methods of genetic analysis rely heavily on correlations among family members’ outcomes to infer genetic effects, whereas longitudinal studies allow investigators to study factors’ effects on outcomes and changes over time [1]. To retrieve full information from longitudinal family data, appropriate statistical approaches are crucial. We proposed a multipoint linkage disequilibrium approach accounting for multilevel correlations between markers per subject, within-subject longitudinal observations, and subjects within families, aiming to correctly localize the disease locus and assess its genetic effects. This approach has several advantages: it allows us to estimate the disease locus position, the disease locus’s genetic effect, and the 95 % confidence intervals without specifying a disease genetic mode and yet making full use of the markers and repeated measurements. In addition,

this approach treats the genotype data as random conditional on the phenotype, eliminating the problem of ascertainment bias. We applied this approach to the baseline and longitudinal prevalence/incidence of hypertension events. The efficiency of parameter estimates was similar for the “Ever” and “Progression” categories, but was improved with repeated longitudinal outcomes compared to the use of “Baseline” only. This difference between analyses might largely result from the different total sample sizes and proportions of hypertensive subjects for different phenotypes. Several identified genes on chromosome 3 for hypertension were consistent with findings from previous linkage and association studies. Despite its advantages, this proposed approach also has limitations; for example, covariate adjustment is not available.

Acknowledgements

We deeply appreciate the reviewers’ thorough reviews and constructive suggestions, which greatly improve the quality of this manuscript. This project was supported by a grant from Ministry of Science and Technology, Taiwan (MOST102-2118-M-400-005) and a grant from the National Health Research Institutes, Taiwan (PH-103-pp-04). We thank Ms. Karen Klein (Biomedical Research Services and Administration, Wake Forest School of Medicine) for her editorial contributions to this manuscript.

Declarations

This article has been published as part of *BMC Proceedings* Volume 10 Supplement 7, 2016: Genetic Analysis Workshop 19: Sequence, Blood Pressure and Expression Data. Summary articles. The full contents of the supplement are available online at <http://bmcpoc.biomedcentral.com/articles/supplements/volume-10-supplement-7>. Publication of the proceedings of Genetic Analysis Workshop 19 was supported by National Institutes of Health grant R01 GM031575.

Authors’ contributions

YFC designed the overall study, CYL conducted statistical analyses, YFC and FCH drafted the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Institute of Population Health Sciences, National Health Research Institutes, Miaoli, 35053, Taiwan, Republic of China. ²Department of Biostatistical Sciences, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem 27157, USA.

Published: 18 October 2016

References

1. Gauderman WJ, Macgregor S, Briollais L, Scurrah K, Tobin M, Park T, et al. Longitudinal data analysis in pedigree studies. *Genet Epidemiol.* 2003;25 Suppl 1:S18–28.
2. Chiu YF, Lee CY, Kao HY, Pan WH, Hsu FC. Analysis of family- and population-based samples using multiple linkage disequilibrium mapping. *Ann Hum Genet.* 2013;77(3):251–67.
3. Liu F, Kirichenko A, Axenovich TI, van Duijn CM, Aulchenko YS. An approach for cutting large and complex pedigrees for linkage analysis. *Eur J Hum Genet.* 2008;16(7):854–60.
4. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet.* 2007;81(3):559–75.
5. Chiu YF, Liang KY, Pan WH. Incorporating covariates into multipoint association mapping in the case-parent design. *Hum Hered.* 2010;69(4):229–41.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

