

SUPPLEMENTARY DATA

A. Study population:

The study included individuals recruited in the Indian Migration Study (IMS) which was nested within a cardiovascular risk factor surveillance system that monitored risk factors in industry populations across several large cities in India [1]. Briefly each migrant worker and his/her co-resident spouse were asked to invite one non-migrant full sibling of the same sex and closest to them in age still residing in their rural place of origin. This strategy resulted in rural dwelling sibs being drawn from 20 of the 28 states in India, reflecting the migration patterns of the factory workforce and their spouses. Full details of the study recruitment have been described previously [2]. Fieldwork commenced in March 2005 and ended by December 2007, and phenotypic information was available on 7068 participants, out of which 6774 individuals were true sibling pairs [i.e. 137 pairs (274 individuals) were cousins or unrelated pairs (cousins and friends of an index person) and 20 were singletons]. Biochemical data was available for 6942 individuals recruited from four cities in India (Lucknow, Nagpur, Hyderabad and Bangalore). The majority of sibling pairs (n=1832 pairs) consisted of a rural to urban migrant factory worker and a sibling who had remained in a rural area. There were also 1012 urban-urban sibling pairs and 245 rural- rural sibling pairs. Ethical approval was obtained from the ethics committee of the All India Institute of Medical Sciences, New Delhi, India (AIIMS; reference number A-60/4/8/2004).

B. Stratification Analysis

The within-family analysis in the Fulker model is potentially conservative as it treats each family as a separate stratum [3]. Therefore in order to explore whether we could adjust for a reduced number of strata, we applied the STRUCTURE software [4] to the genotype data, using one sib per family. The best fit to the data used 10 sub-populations within our study sample, but the subjects did not form discrete clusters (Figure 1). We tried a total association model adjusting for a smaller number of clusters (n=10) using the companion STRAT program, but this yielded results similar to those of the Fulker model. We have therefore only reported the results of the Fulker model.

References

1. Lyngdoh T, Kinra S, Shlomo YB et al. (2006) Sib-recruitment for studying migration and its impact on obesity and diabetes. *Emerg Themes Epidemiol* 3:2
2. Ebrahim S, Kinra S, Bowen L et al. (2010) The effect of rural-to-urban migration on obesity and diabetes in India: a cross-sectional study. *PLoS Med* 7(4):e1000268
3. Fulker DW, Cherny SS, Sham PC, Hewitt JK (1999) Combined linkage and association sib-pair analysis for quantitative traits. *Am J Hum Genet* 64(1):259-267
4. Pritchard JK, Stephens M and Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155: 945–959