



**Supplementary Fig. 5.** Principal component analysis (PCA) in (A) the Hyperglycemia and Adverse Pregnancy Outcome-Hong Kong (HAPO-HK) Study, (B) the Tianjin Study, and (C) the Treated GDM Cases vs. Non-diabetes Controls (TGDM-NDM) Study. The PCA plots show the first two principal components (PCs), based on genotype data of 26 different populations from the 1000 Genomes Project, as well as each discovery cohort of Chinese women (A: 961 women from the HAPO-HK Study; B: 455 women from the Tianjin Study; and C: 266 women from the TGDM-NDM Study). The 26 populations from the 1000 Genomes Project have been divided into five super populations: African (AFR) includes Yoruba in Ibadan, Nigeria, Luhya in Webuye, Kenya, Gambian in Western Divisions in the Gambia, Mende in Sierra Leone, Esan in Nigeria, Americans of African Ancestry in SW USA, and African Caribbeans in Barbados; Ad Mixed American (AMR) includes Mexican Ancestry from Los Angeles USA, Puerto Ricans from Puerto Rico, Colombians from Medellin, and Colombia, Peruvians from Lima, Peru; South Asian (SAS) includes Gujarati Indian from Houston, Texas, Punjabi from Lahore, Pakistan, Bengali from Bangladesh, Sri Lankan Tamil from the UK, and Indian Telugu from the UK; European (EUR) includes Utah Residents (CEPH) with Northern and Western European Ancestry, Toscani in Italia, Finnish in Finland, British in England and Scotland, and Iberian Population in Spain; and East Asian (EAS) includes Han Chinese in Beijing, China, Japanese in Tokyo, Japan, Southern Han Chinese, Chinese Dai in Xishuangbanna, China, and Kinh in Ho Chi Minh City, Vietnam.