## (a) Reference data generation ..... ... ......... .. 1000G (chr. 20) • • • Split into subset panels Removing monomorphic sites Split panels Union Intersection and singletons (b) Accuracy analysis *Imputation* Meta-imputation Split panels Quality control filtering Removing genotypes contained in scaffold Genotype GoT2D Intersection Η scaffold (chr. 20) Matching to variants contained in GoT2D Union G Н • Calculating $r^2$ per variant Extraction of variants typed using Illumina Omni2.5 Chip (c) Association analysis Simulation Defining 1 causal variant per simulation, randomly selected within MAF interval and simulated at relative risk (RR) GoT2D $(RR_{het} = 1.2)$ - Low risk (20, 50][0.5, 1](5, 10] $(RR_{het} = 1.6)$ - Medium risk (chr. 20) (10, 20]- High risk $(RR_{het} = 2.0)$ 100 replicate simulations per combination of MAF interval and RR category *Imputation* Meta-imputation Split panels Quality control filtering Association analysis Simulated Simulated (frequentist score test) Intersection case-control Н genotype Signal detection within data scaffold 1Mb around simulated causal site Union

Extraction of variants corresponding to Illumina Omni2.5 Chip