TwinsUK - The Most Genotyped and Phenotyped Twin Cohort in the World

Jordana Bell, Mario Falchi, Chris Hammond, Juliette Harris, Deborah Hart, Pirro Hysi, Massimo Mangino, Cristina Menni, Kerrin Small, Claire Steves, Frances Williams and Tim Spector

Department of Twin Research, Faculty of Life Sciences & Medicine, King's College London   www.twinsuk.ac.uk/
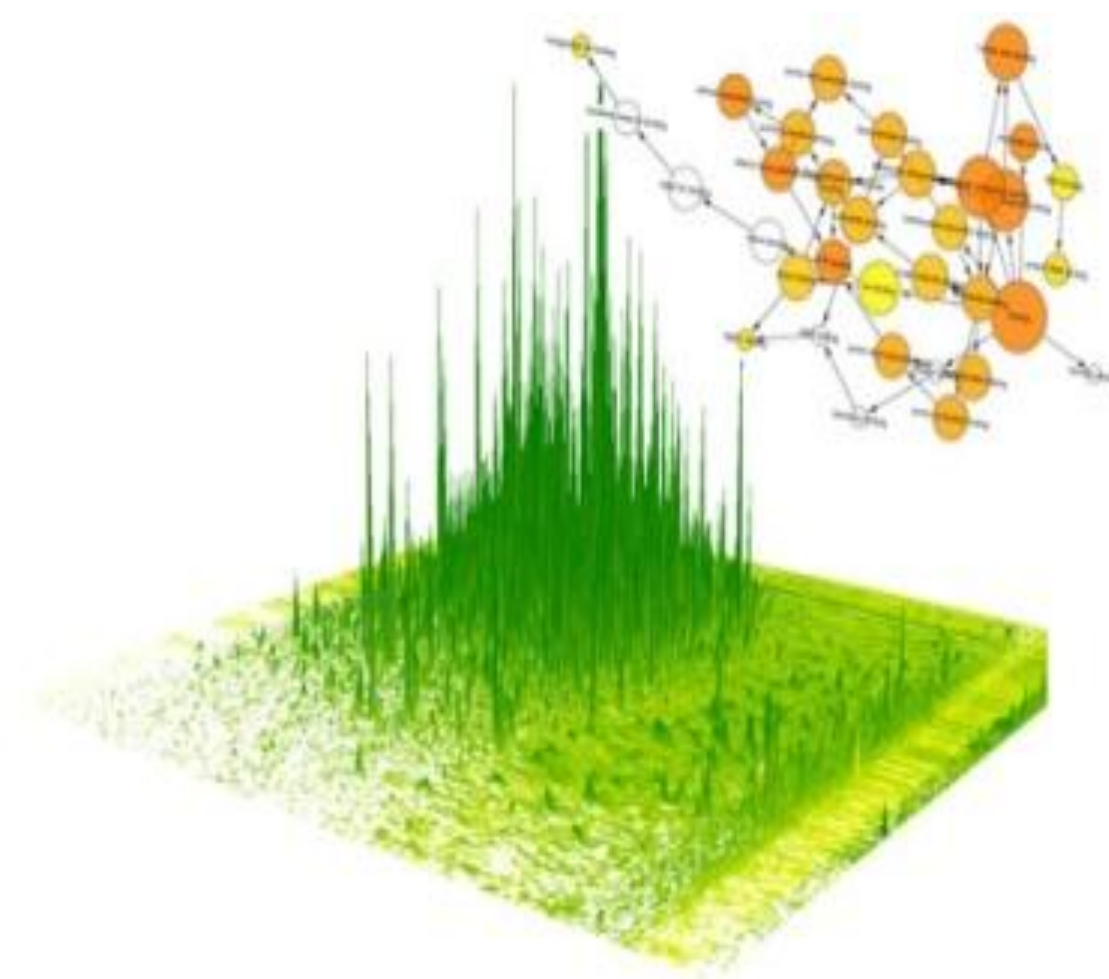
Since 1992, the Department of Twin Research (DTR) has collected data on over 13,000 members of the TwinsUK registry in order to understand how genetic variations relate to health, disease and aging. The data has enabled multiple collaborations with over 800 groups worldwide, resulting in over 600 research papers, including publications on the social and ethical implications of genetic technological advances. Genome-wide association studies using TwinsUK data have identified more than 400 novel gene loci in 30 disease areas including osteoporosis, osteoarthritis and melanoma. In addition, every variant in the genome has been analysed in 2,000 twins (UK10K project) with a further 7000 having been genotyped (Illumina platform). The dataset has rapidly expanded in the recent years to examine gene expression in multiple tissues (MuTHER study), Epigenetics (EpiTwin project) and to also investigate metabolomic (of over 500 metabolites), transcriptomics, immunomics and glycomics data. Data has also been collected on the gut microbiome of over 3000 twins, using sequencing of 16s rDNA and entire metagenomes, to understand the role of the microbiome in disease and ageing. Such an extensive dataset has resulted in the TwinsUK cohort being one of the most genotyped and phenotyped in the world. The scope of its 'omic' and genetic data provides a unique opportunity to investigate the relationship between our health, genes, environment and microbiome in a way, and to an extent that is unprecedented.
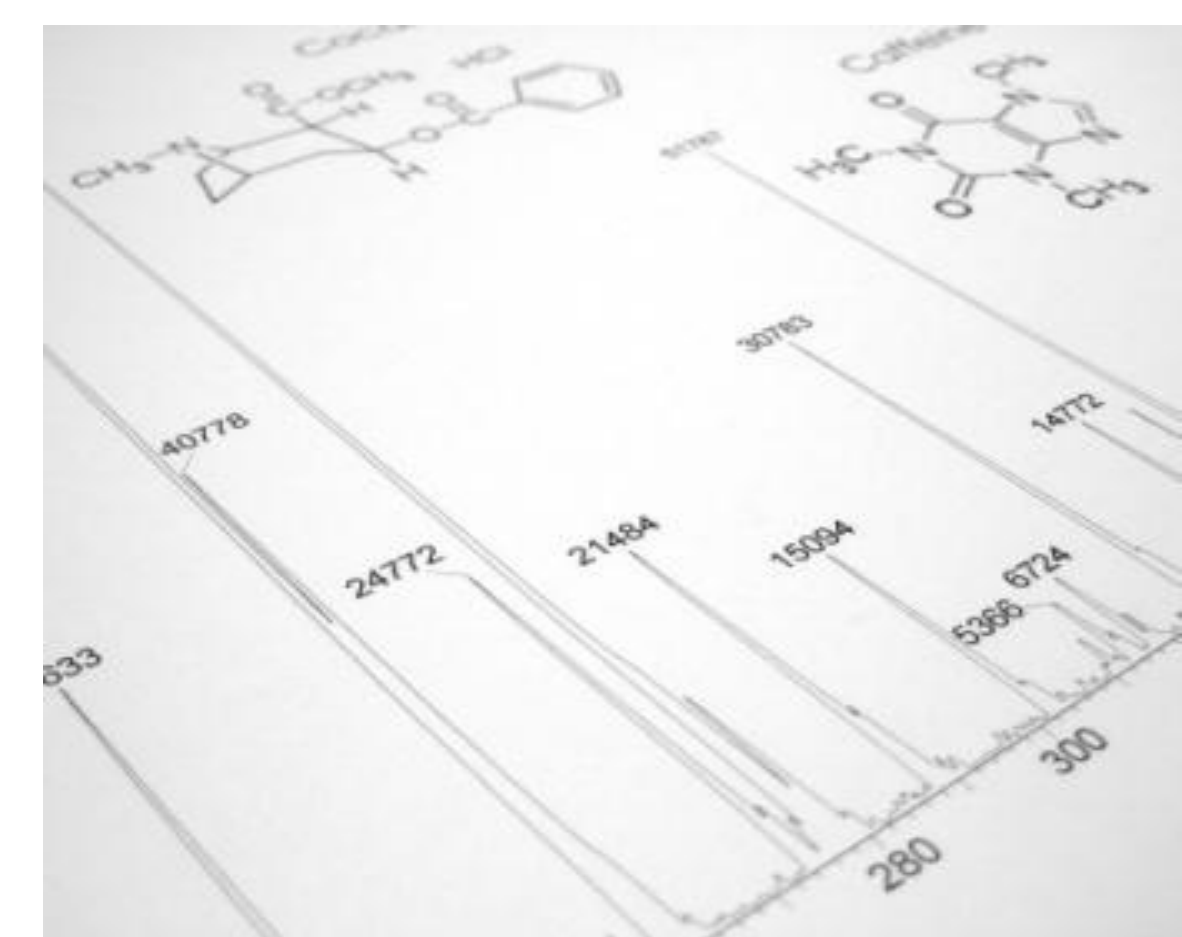
Genomics

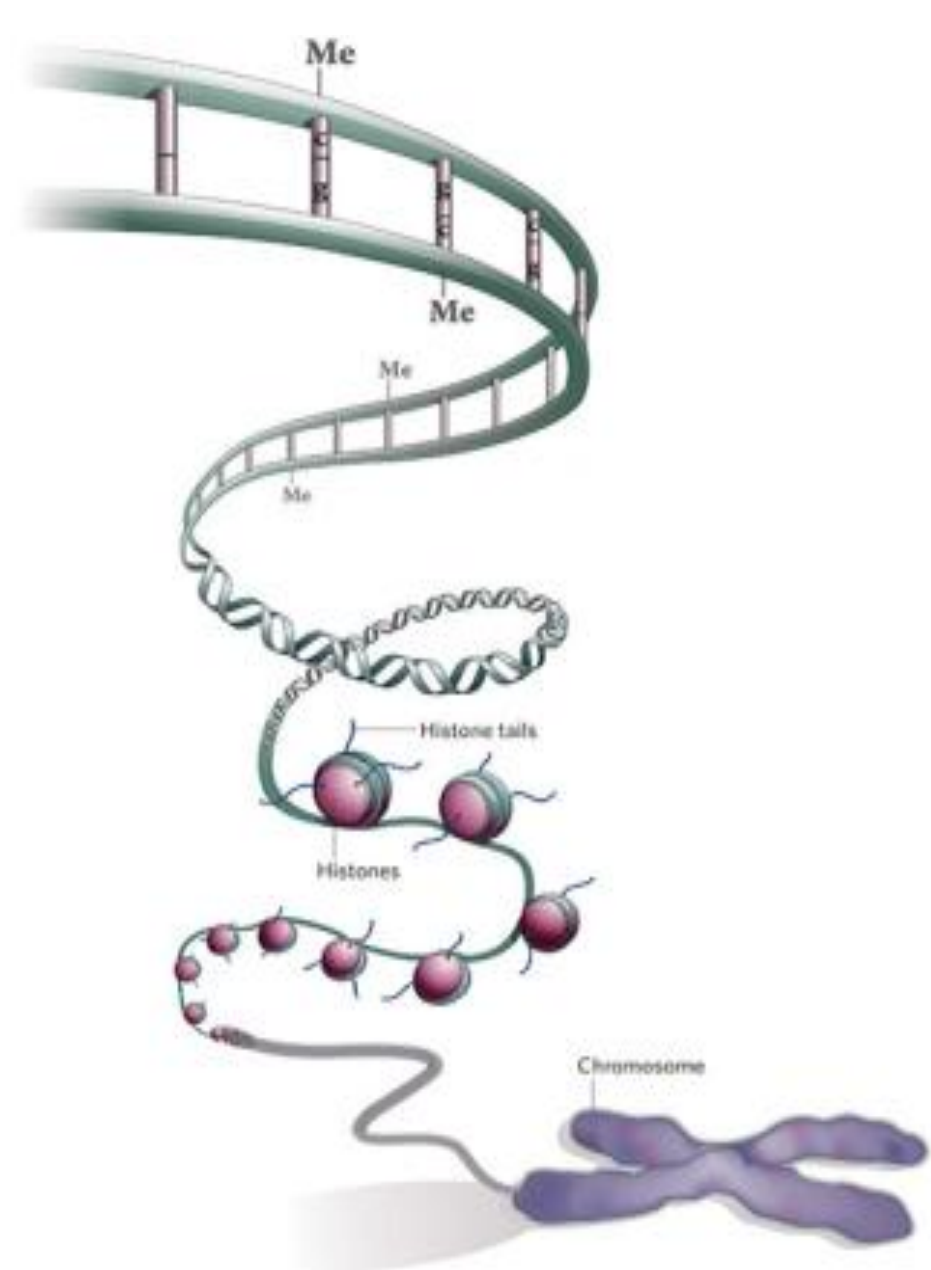Genotypes (6,922 Illumina CHIP; 3,300 Illumina Exome CHIP); DNA Sequence (3,500 NGS)

Proteomics

Somalogic panel of over 1000 proteins on 220 individuals with longitudinal phenotypic data

Metabolomics

Plasma/serum: 1,000 Biocrates; 8000 Metabolon, subset of 2000 with longitudinal profiles, 2000 Biocrates at three time points; 800 faeces
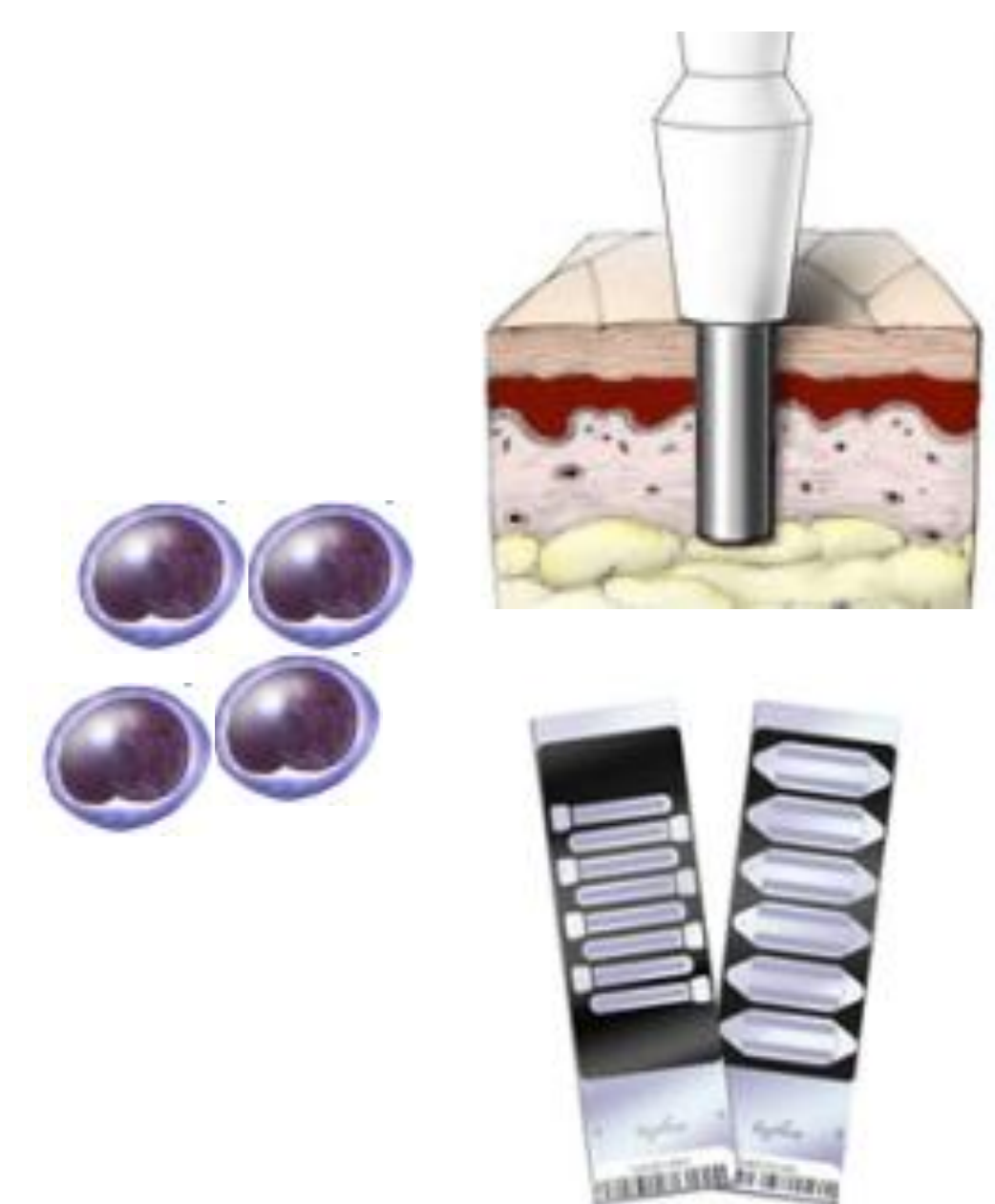
Epigenomics

Whole blood: 5,000 MEDIP-seq and 1000 Illumina 450K; adipose tissue: 600 Illumina 450K and subset in skin tissue
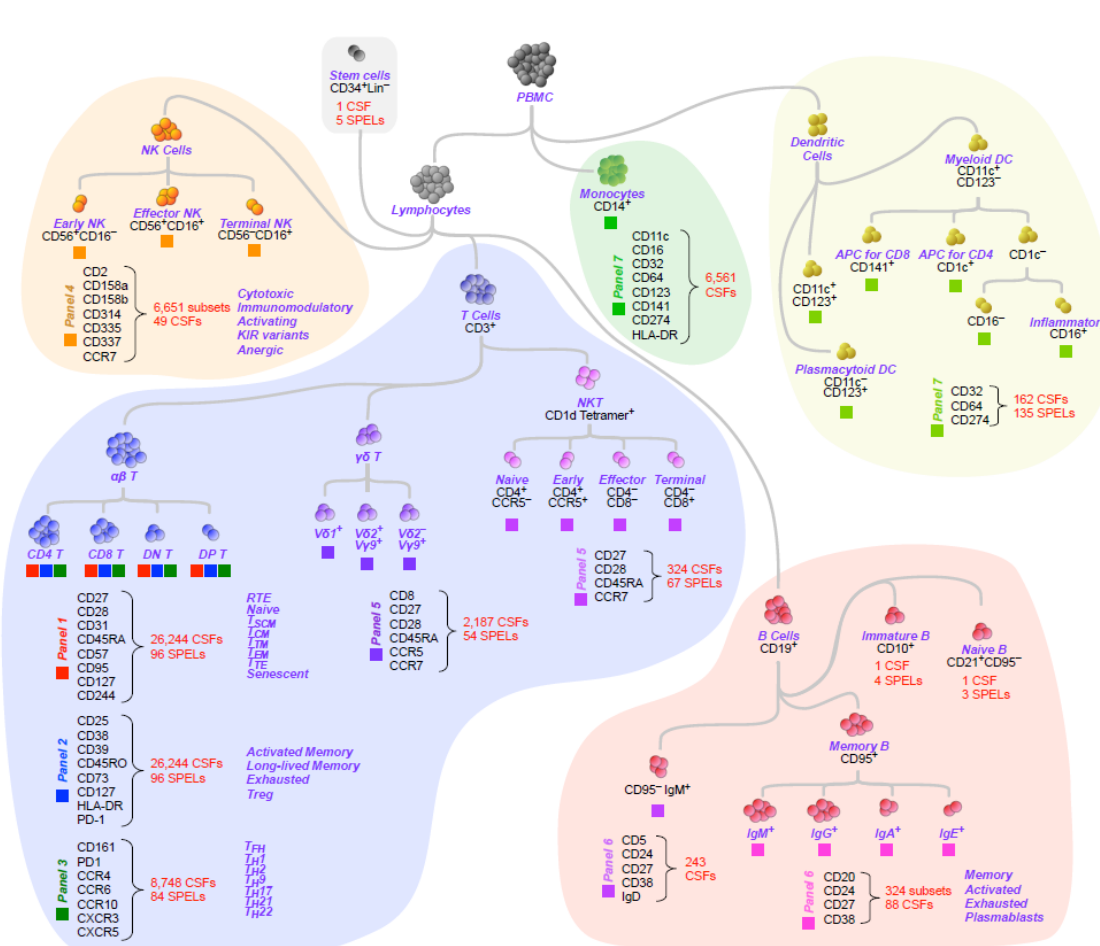
TwinsUK 'OMICS' BIG DATA

from 13,000 twins, longitudinally over 24 years

Transcriptomics

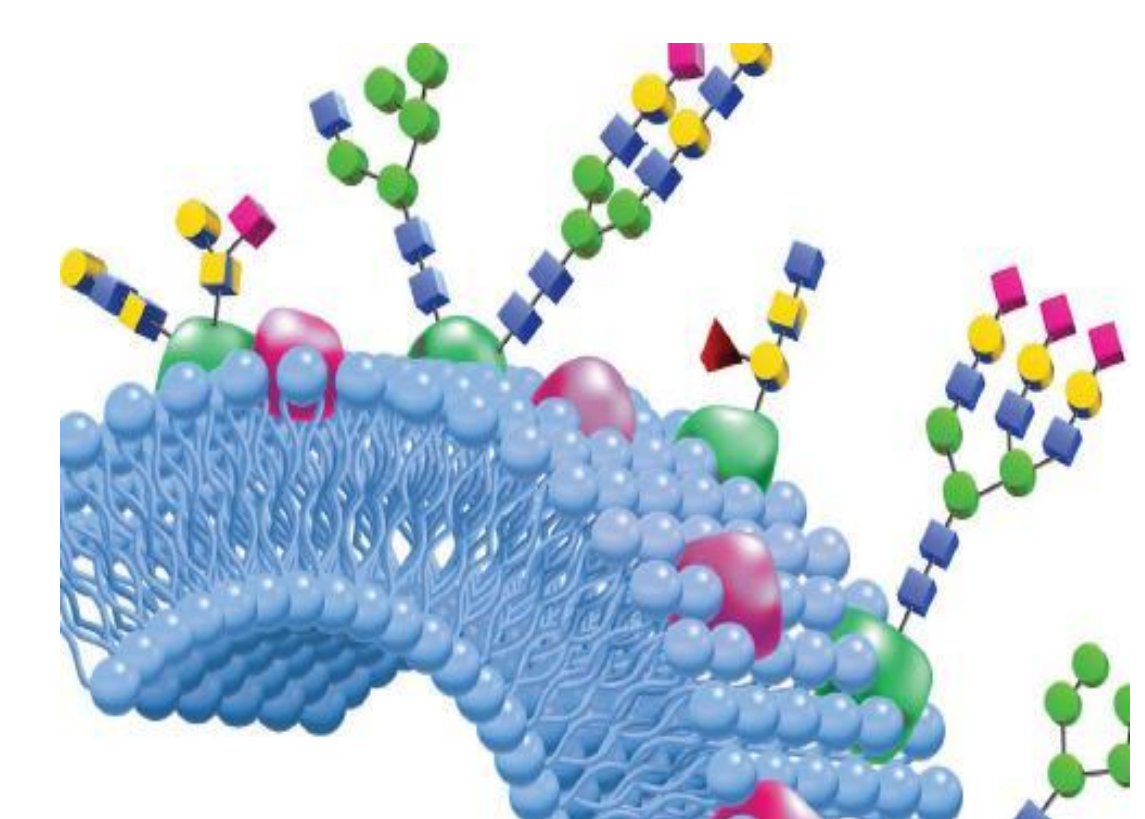4 tissue RNAseq profiles in 850 individuals, subset with longitudinal profiles

Immunomics

89,051 data points over 500 individuals

Microbiome

Quality controlled 16S data on over 2700 individuals, 700 of which have >1 datapoint. Shotgun Metagenomes data on over 1000 samples. Faecal metabolites on over 1000 samples

Glycomics

IgG and total plasma N-glycome for ~4600 individuals