

Born in Bradford Data Dictionary

Sample IDs and counts to link with with bioinformatics data files

Contents

Background	1
Born in Bradford	1
Study identifiers	2
Child cord blood DNA methylation linkage	3
Mother pregnancy bloods DNA methylation linkage	4
Exome seq linkage: Child	5
Exome seq linkage: Mother	6
Exome seq linkage: Mother-Pregnancy	7
Exome seq linkage: Mother-Pregnancy replicates	8
GWAS Linkage: Child replicates	9
GWAS Linkage: Child	10
GWAS Linkage: Mother	12
GWAS Linkage: Mother-Pregnancy	13
GWAS Linkage: Mother-Pregnancy replicates	14

Background

This document is a data dictionary for Sample IDs and counts to link with with bioinformatics data files. It describes 28 variables from 11 sources. This document was built from Born in Bradford database version BUILD-JAN2018.

Born in Bradford

Born in Bradford is a longitudinal multi-ethnic birth cohort study aiming to examine the impact of environmental, psychological and genetic factors on maternal and child health and wellbeing. Bradford is a city in the North of England with high levels of socio-economic deprivation and ethnic diversity. Women were recruited at the Bradford Royal Infirmary at 26-28 weeks gestation. For those consenting, a baseline questionnaire was completed. The full BiB cohort recruited 12,453 women and 3353 of their partners across 13,776 pregnancies and 13,858 children between 2007 and 2010. The cohort is broadly characteristic of the city's maternal population. Mean age of the mothers at study recruitment was 27 years old. Researchers are looking at the links between the circumstances of a child's birth, the context in which they grow up, their health and well-being and their educational progress. Ethical approval for the data collection was granted by Bradford Research Ethics Committee (Ref 07/H1302/112).

Study identifiers

Study identifiers are standardised across Born in Bradford data sources to enable linking of data from different sources.

Variable	Variable Label	Details
ChildID	BiB Child ID	Unique ID assigned to each child at birth. Where birth outcome is unknown for a given pregnancy, ChildID will be blank and there is no child recruited to the study from that pregnancy. Use MotherID with ChildID to link siblings together. Note that twins have separate ChildIDs but the same PregnancyID.
FatherID	BiB Father ID	Unique ID assigned to partners post-recruitment. Use FatherID with PregnancyID to link fathers across pregnancies. Where FatherID matches across two PregnancyIDs, but those PregnancyIDs are associated with different MotherIDs, this is a father with two separate pregnancies in the cohort with different mothers. Likewise, where MotherID matches across two PregnancyIDs, but those PregnancyIDs are associated with different FatherIDs, this is a mother with two separate pregnancies in the cohort with different fathers.
MotherID	BiB Mother ID	Unique ID assigned to each mother post-recruitment. MotherID should be used when looking for pregnancies or children associated with the same mother. Data collected at pregnancy level will duplicate for MotherIDs that are in the study for more than one pregnancy.
PregnancyID	BiB Pregnancy ID	Unique ID assigned to each mother at recruitment. It is named PregnancyID because a mother can enrol for more than one pregnancy. If a mother returns to enrol for a second or third pregnancy, she is assigned a new PregnancyID. Children and partners from that pregnancy can be linked to the mother by the PregnancyID

Child cord blood DNA methylation linkage

Database ID for source: `dnamch`

This source is measured at the **child** level. It contains data from 951 children with one observation per child. There are 4 variables with a total of 951 observations.

Description

DNA methylation on Illumina MethylationEPIC (850k) chip. Linkage information for child samples, DNA from cord bloods. Selected 1000 paired mother-child samples where: Mother completed GTT; 50% Pakistani, 50% white British; Singleton birth; Existing good quality genotype data for both in the pair.

Variable	Variable Label	Details
dnambarcode	DNA methylation BiB lab sample ID (child at birth)	Administrative: Integer value <hr/> DNA methylation BiB lab sample ID (child at birth) <hr/> Range 98272086 to 117248923 Mean 107451998.78 951 non-missing values
dnamclabid	DNA methylation assaying lab sample ID (child at birth)	Administrative: Text value <hr/> DNA methylation assaying lab sample ID (child at birth) <hr/> 951 unique values 951 non-missing values
dnamreplicate	DNA methylation sample replicate submitted (child at birth)	Administrative: Categorical value <hr/> DNA methylation sample replicate submitted (child at birth) <hr/> 7 non-missing values <hr/> Coding [<code>dnamchdnamreplicate_label</code>]: 1 = replicate
dnamcsentrixid	DNA methylation assay data ID (child at birth)	Administrative: Text value <hr/> DNA methylation assay data ID (child at birth) <hr/> 951 unique values 951 non-missing values

Mother pregnancy bloods DNA methylation linkage

Database ID for source: `dnampp`

This source is measured at the **pregnancy** level. It contains data from 934 pregnancies with one observation per pregnancy. There are 4 variables with a total of 934 observations.

Description

DNA methylation on Illumina MethylationEPIC (850k) chip. Linkage information for mother samples, DNA from pregnancy bloods. Selected 1000 paired mother-child samples where: Mother completed GTT; 50% Pakistani, 50% white British; Singleton birth; Existing good quality genotype data for both in the pair.

Variable	Variable Label	Details
dnammbarcodes	DNA methylation BiB lab sample ID (mother in pregnancy)	Administrative: Integer value DNA methylation BiB lab sample ID (mother in pregnancy) Range 98272075 to 117249006 Mean 106224954.54 934 non-missing values
dnammcreplicates	DNA methylation sample replicate submitted (mother in pregnancy)	Administrative: Categorical value DNA methylation sample replicate submitted (mother in pregnancy) 3 non-missing values Coding [<code>dnamppdnamreplicate_label</code>]: 1 = replicate
dnammlabids	DNA methylation assaying lab sample ID (mother in pregnancy)	Administrative: Text value DNA methylation assaying lab sample ID (mother in pregnancy) 934 unique values 934 non-missing values
dnammsentrixids	DNA methylation assay data ID (mother in pregnancy)	Administrative: Text value DNA methylation assay data ID (mother in pregnancy) 934 unique values 934 non-missing values

Exome seq linkage: Child

Database ID for source: exseqc

This source is measured at the **child** level. It contains data from 262 children with one observation per child. There are 2 variables with a total of 262 observations.

Description

Exome sequence linkage information for child samples.

Variable	Variable Label	Details
exomecsangerid	Exome Seq Sanger Sample ID (Child)	Administrative: Text value Exome Seq Sanger Sample ID (Child) 262 unique values 262 non-missing values
has_exomec	Child has exome sequence data	Derived: Categorical value Subcohort indicator: Child has exome sequence data 262 non-missing values Coding [has]: 1 = has

Exome seq linkage: Mother

Database ID for source: exseqm

This source is measured at the **mother** level. It contains data from 2333 mothers with one observation per mother. There are 2 variables with a total of 2333 observations.

Description

Exome sequence linkage information for mother samples.

Variable	Variable Label	Details
exomempregs	Number of mother's pregnancies with exome sequenced samples	Derived: Integer value Number of mother's pregnancies with exome sequenced samples Range 1 to 3 Mean 1.07 2333 non-missing values
has_exomem	Mother has exome sequence data	Derived: Categorical value Subcohort indicator: Mother has exome sequence data 2333 non-missing values Coding [has]: 1 = has

Exome seq linkage: Mother-Pregnancy

Database ID for source: exseqp

This source is measured at the **pregnancy** level. It contains data from 2489 pregnancies with one observation per pregnancy. There are 3 variables with a total of 2489 observations.

Description

Exome sequence linkage information for mother samples from different pregnancies, i.e. mother-within-pregnancy samples.

Variable	Variable Label	Details
exomepsangerid	Exome Seq Sanger Sample ID (Mother-Pregnancy)	Administrative: Text value <hr/> Exome Seq Sanger Sample ID (Mother-Pregnancy) <hr/> 2488 unique values 2488 non-missing values
exomerepreg	Mother has exome sequence replicate at this pregnancy	Derived: Categorical value <hr/> Mother has exome sequence replicate at this pregnancy <hr/> 2489 non-missing values <hr/> Coding [gwasreplicates]: 0 = No replicate 1 = Has replicate
has_exomempreg	Mother has exome sequence data at this pregnancy	Derived: Categorical value <hr/> Subcohort indicator: Mother has exome sequence data at this pregnancy <hr/> 2489 non-missing values <hr/> Coding [has]: 1 = has

Exome seq linkage: Mother-Pregnancy replicates

Database ID for source: `exsqpr`

This source is measured at the **pregnancy** level. It contains data from 1 pregnancies with more than one observation per pregnancy. There is 1 variable with a total of 2 observations.

Description

Exome sequence linkage information for mother-within-pregnancy samples with replicates. This sample was accidentally aliquoted into two wells. This gives us an unexpected replicate.

Variable	Variable Label	Details
exomepsangeridr	Exome Seq Replicate Sanger Sample ID (Mother-Pregnancy)	Administrative: Text value <hr/> Exome Seq Replicate Sanger Sample ID (Mother-Pregnancy) <hr/> 2 unique values 2 non-missing values 1 pregnancies with between 2 and 2 observations each

GWAS Linkage: Child replicates

Database ID for source: `gwascr`

This source is measured at the **child** level. It contains data from 4 children with more than one observation per child. There is 1 variable with a total of 8 observations.

Description

GWAS sample linkage information for child samples with replicates.

Variable	Variable Label	Details
gwassentrixcr	Sentrix ID (child replicates)	Administrative: Text value <hr/> Sentrix array matrix and position for child sample replicates <hr/> 8 unique values 8 non-missing values 4 children with between 2 and 2 observations each

GWAS Linkage: Child

Database ID for source: gwas1c

This source is measured at the **child** level. It contains data from 7157 children with one observation per child. There are 5 variables with a total of 7157 observations.

Description

GWAS sample linkage information for child samples.

Variable	Variable Label	Details
gwasrepchild	Child sample has GWAS replicate	Derived: Categorical value Child sample has GWAS replicate 7157 non-missing values Coding [gwasreplicates]: 0 = No replicate 1 = Has replicate
gwassentrixcs	Sentrix ID (child)	Administrative: Text value Sentrix array matrix and position for child samples 7153 unique values 7153 non-missing values
has_gwas	Has microarray genotype data	Derived: Categorical value Subcohort indicator: has Illumina microarray genotype data 7157 non-missing values Coding [has]: 1 = has
has_gwascpairm	Has genotype data paired with mother	Derived: Categorical value Subcohort indicator: has Illumina microarray genotype data paired with mother, not necessarily from same pregnancy 6256 non-missing values Coding [has]: 1 = has

Variable	Variable Label	Details
has_gwascpairmpreg	Has genotype data paired with mother at same pregnancy	<p>Derived: Categorical value</p> <hr/> <p>Subcohort indicator: has Illumina microarray genotype data paired with mother at same pregnancy</p> <hr/> <p>6131 non-missing values</p> <hr/> <p>Coding [has]: 1 = has</p>

GWAS Linkage: Mother

Database ID for source: `gwas1m`

This source is measured at the **mother** level. It contains data from 8610 mothers with one observation per mother. There are 2 variables with a total of 8610 observations.

Description

GWAS sample linkage information for mother samples.

Variable	Variable Label	Details
gwasmpregs	Number of mother's pregnancies with genotyped samples	Derived: Integer value Number of mother's pregnancies with genotyped samples Range 1 to 3 Mean 1.06 8610 non-missing values
has_gwasm	Mother has genotype data	Derived: Categorical value Subcohort indicator: mother has Illumina microarray genotype data at any pregnancy 8610 non-missing values Coding [has]: 1 = has

GWAS Linkage: Mother-Pregnancy

Database ID for source: gwas1p

This source is measured at the **pregnancy** level. It contains data from 9099 pregnancies with one observation per pregnancy. There are 3 variables with a total of 9099 observations.

Description

GWAS sample linkage information for mother samples from different pregnancies, i.e. mother-within-pregnancy samples.

Variable	Variable Label	Details
gwasrempreg	Mother-pregnancy sample has GWAS replicate	Derived: Categorical value Mother-pregnancy sample has GWAS replicate 9099 non-missing values Coding [gwasreplicates]: 0 = No replicate 1 = Has replicate
gwassentrixps	Sentrix ID (mother-pregnancy)	Administrative: Text value Sentrix array matrix and position for mother-pregnancy samples 9094 unique values 9094 non-missing values
has_gwasmpreg	Mother has genotype data at this pregnancy	Derived: Categorical value Subcohort indicator: mother has Illumina microarray genotype data at this pregnancy 9099 non-missing values Coding [has]: 1 = has

GWAS Linkage: Mother-Pregnancy replicates

Database ID for source: `gwaspr`

This source is measured at the **pregnancy** level. It contains data from 5 pregnancies with more than one observation per pregnancy. There is 1 variable with a total of 10 observations.

Description

GWAS sample linkage information for mother-within-pregnancy samples with replicates.

Variable	Variable Label	Details
gwassentrixpr	Sentrix ID (mother-pregnancy replicates)	Administrative: Text value <hr/> Sentrix array matrix and position for mother-pregnancy sample replicates <hr/> 10 unique values 10 non-missing values 5 pregnancies with between 2 and 2 observations each