**Pipeline → Linkage Disequilibrium**

[**Information data:**](#_wahhul854hi4) **1**

[**SmartCleaning:**](#_qzjrmw4on4g5) **2**

[**QC specific**](#_nrskn2z275l3) **3**

[Sex Check](#_esly18xf21mo) 3

[Heterozigozity for samples](#_75jlqpmoaiez) 3

[Relationship + Imbreeding Inference](#_jzzv34xuuq38) 3

[Merge with reference](#_k2h6818iwouh) 3

[Remove monomorphic markers](#_r1kpbsglouwt) 4

[Remove Markers Within Regions That are Known for Long LD](#_6kskxji5r8d4) 4

[Prun-in](#_ycn60d9erqzr) 5

[Centimorgan Positions](#_6s9apwff1ym4) 8

[**Linkage Analysis**](#_wfuyruaikn9b) **9**

[Make Inputs Merlin](#_exgazcbqh9uy) 9

[.dat](#_qxr9brld4wpc) 10

[.map](#_g6npsop2qx99) 11

[.ped](#_y6zs49zeiuma) 11

[pedstat](#_ex3chdl99sz4) 11

[merlin](#_af4stcillo5t) 12

[PCA](#_4hzptvuag82p) 13

[**Linkage Disequilibrium Analysis**](#_n4awever1uys) **13**

[**IBD Sharing Analysis**](#_pconhiy5zky3) **13**

[Phasing process](#_jke5r43uu4hs) 13

# Information data:

vcf:

/hpf/largeprojects/tcagstor/tools/data/MSSNG/ILMN/variants/SNVs+indels/joint

samples:

IDs in MSSNG are:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **SUBMITTED\_ID** | **DNA\_NUMBER** | **DNA\_SOURCE** | **FATHER\_ID** | **MOTHER\_ID** | **AFFECTION** | **RELATION** | **SEX** |
| MSSNG00421-001 | 07-44A | Blood | 0 | 0 | 1 | Mother | F |
| MSSNG00421-002 | 07-45A | Blood | 0 | 0 | 1 | Father | M |
| MSSNG00421-003 | 07-43A | Blood | MSSNG00421-002 | MSSNG00421-001 | 2 | Proband | F |
| MSSNG00421-004 | 07-183A | Blood | MSSNG00421-002 | MSSNG00421-001 | 2 | AffectedSibling | M |
| MSSNG00421-005 | 13-82A | Blood | MSSNG00421-002 | MSSNG00421-001 | 2 | AffectedSibling | M |
| MSSNG00421-006 | 13-83A | Blood | MSSNG00421-002 | MSSNG00421-001 | 2 | AffectedSibling | M |
| MSSNG00421-007 | 13-84A | Blood | MSSNG00421-002 | MSSNG00421-001 | 2 | AffectedSibling | M |
| MSSNG00421-008 | 18-55A | Blood | MSSNG00421-002 | MSSNG00421-001 | 2 | AffectedSibling | F |
| MSSNG00421-009 | 18-56A | Blood | MSSNG00421-002 | MSSNG00421-001 | 2 | AffectedSibling | M |
| **SUBMITTED\_ID** | **DNA\_NUMBER** | **DNA\_SOURCE** | **FATHER\_ID** | **MOTHER\_ID** | **AFFECTION** | **RELATION** | **SEX** |
| 1-0627-001 | 238162 | Blood | 2-1689-002 | 2-1689-001 | 1 | Mother | F |
| 1-0627-002 | 238067 | Blood | 0 | 0 | 1 | Father | M |
| 1-0627-003 | 202774 | Blood | 1-0627-002 | 1-0627-001 | 2 | Proband | M |
| 1-0627-004 | 194812 | Saliva | 1-0627-002 | 1-0627-001 | 2 | AffectedSibling | M |
| 1-0627-005 | 203639 | Blood | 1-0627-002 | 1-0627-001 | 2 | AffectedSibling | M |
| 1-0627-006 | 171543 | Blood | 1-0627-002 | 1-0627-001 | 2 | AffectedSibling | M |
| 1-0627-007 | 202211 | Blood | 1-0627-002 | 1-0627-001 | 2 | AffectedSibling | M |

# SmartCleaning:

<https://docs.google.com/document/d/1S_ke6YCPBsYNTN5tgf0wRRUYMBuTMwtKquYyUa_8BOg/edit?usp=sharing>

1. Remove chr 0
2. Remove duplicate data
3. Remove missing data
4. Infer individual sex
5. Remove A|T and C|G variants
6. Remove 100% heterozigotes variants
7. Select Reference common SNPs
8. Add Pop Name
9. SNP-> CHR:POS

# QC specific

*Filter Family Individuals*

for fam in 0627 0042 ; do plink --bfile /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/MosaicQC\_2022/MSSNG\_Autossomic\_SmartQC\_ReferenceSNPs --keep /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_${fam}/Family${fam}.txt --make-bed --out /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_${fam}/Family${fam} ; done

cat /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042.txt /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627.txt > /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/FamiliesList.txt

## Sex Check

Done on SmartQCstep 4.

## Heterozigozity for samples

Standart QC TCAG?

## Relationship + Imbreeding Inference

module load plink/1.9.beta3a

for fam in 0627 0042 ; do plink --bfile /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_${fam}/Family${fam} --genome --het --out /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_${fam}/Family${fam} ; done

## Remove monomorphic markers

freqNonMonomorphic = 1/((9\*2)\*10) #10x to have sure that just remove monomorphic

freqNonMonomorphic =1/18 = 0.05

plink --bfile /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042 --maf 0.05 --make-bed --out /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042\_NoMonomorphic

631,006 variants and 9 people pass filters and QC.

### Family 0627:

freqNonMonomorphic = 1/((7\*2)\*10) #10x to have sure that just remove monomorphic

freqNonMonomorphic =1/14 = 0.07

plink --bfile /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627 --maf 0.07 --make-bed --out /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627\_NoMonomorphic

630546 variants and 7 people pass filters and QC.

## Remove Markers Within Regions That are Known for Long LD

<https://genome.sph.umich.edu/wiki/Regions_of_high_linkage_disequilibrium_(LD)>

FilterRegions.ipynb

plink --bfile /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042\_NoMonomorphic --exclude /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/SNPs2Remove.txt --make-bed --out /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042\_NoMonomorphic\_NoLDRegions

615,746 variants and 9 people pass filters and QC.

plink --bfile /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627\_NoMonomorphic --exclude /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/SNPs2Remove.txt --make-bed --out /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627\_NoMonomorphic\_NoLDRegions

616,143 variants and 7 people pass filters and QC

## Prun-in

plink --bfile /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042\_NoMonomorphic\_NoLDRegions --indep-pairwise 100 10 0.1 --out /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042\_NoMonomorphic\_NoLDRegions\_LD0.1

plink --bfile /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627\_NoMonomorphic\_NoLDRegions --indep-pairwise 100 10 0.1 --out /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627\_NoMonomorphic\_NoLDRegions\_LD0.1

plink --bfile /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042\_NoMonomorphic\_NoLDRegions --extract /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042\_NoMonomorphic\_NoLDRegions\_LD0.1.prune.in --geno 0.01 --mind 0.01 --make-bed --out /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind

5162 variants and 9 people pass filters and QC..

plink --bfile /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627\_NoMonomorphic\_NoLDRegions --extract /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627\_NoMonomorphic\_NoLDRegions\_LD0.1.prune.in --geno 0.01 --mind 0.01 --make-bed --out /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind

5,128 variants and 7 people pass filters and QC.

## Centimorgan Positions

for chr in $(seq 1 22) ; do awk -v chr="$chr" '{if($1=="chr"chr){print $2"\t"$3"\t"$4}}' /hpf/largeprojects/tcagstor/users/marlam/MapaGenetico/recomb-hg38/genetic\_map\_GRCh38\_merged.tab > /hpf/largeprojects/tcagstor/users/marlam/MapaGenetico/genetic\_map\_GRCh38\_chr${chr}\_combined.txt ; done

for chr in $(seq 1 22) ; do cat /hpf/largeprojects/tcagstor/users/marlam/MapaGenetico/Header.txt /hpf/largeprojects/tcagstor/users/marlam/MapaGenetico/genetic\_map\_GRCh38\_chr${chr}\_combined.txt > /hpf/largeprojects/tcagstor/users/marlam/MapaGenetico/genetic\_map\_GRCh38\_chr${chr}\_combined\_Header.txt ; done

for chr in $(seq 1 22) ; do plink --bfile /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind --chr ${chr} --make-bed --out /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr} ; done

for chr in $(seq 1 22) ; do plink --bfile /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind --chr ${chr} --make-bed --out /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr} ; done

for chr in $(seq 1 22) ; do plink --bfile /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr} --cm-map /hpf/largeprojects/tcagstor/users/marlam/MapaGenetico/genetic\_map\_GRCh38\_chr${chr}\_combined\_Header.txt ${chr} --make-bed --out /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map ; done

for chr in $(seq 1 22) ; do plink --bfile /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr} --cm-map /hpf/largeprojects/tcagstor/users/marlam/MapaGenetico/genetic\_map\_GRCh38\_chr${chr}\_combined\_Header.txt ${chr} --make-bed --out /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map ; done

# 

# Linkage Analysis

## Make Inputs Merlin

for chr in {1..22} ; do plink --bfile /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map --recode --out /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map ; done

for chr in {1..22} ; do plink --bfile /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map --recode --out /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map ; done

### .dat

for chr in {1..22} ; do awk '{print "M""\t"$2}' /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map.map > /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map.dat ; done

for chr in {1..22} ; do awk '{print "M""\t"$2}' /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map.map > /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map.dat ; done

for chr in {1..22} ; do cat /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Header.dat /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map.dat > /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map\_H.dat ; done

for chr in {1..22} ; do cat /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Header.dat /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map.dat > /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map\_H.dat ; done

### .map

for chr in {1..22} ; do awk '{print $1"\t"$2"\t"$3}' /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map.map > /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map\_merlin.map ; done

for chr in {1..22} ; do awk '{print $1"\t"$2"\t"$3}' /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map.map > /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map\_merlin.map ; done

### .ped

/hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map.ped

/hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map.ped

## pedstat

/hpf/largeprojects/tcagstor/users/marlam/Programs/merlin-1.1.2/executables/pedstats -d /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map\_H.dat -p /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map.ped

/hpf/largeprojects/tcagstor/users/marlam/Programs/merlin-1.1.2/executables/pedstats -d /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr22\_map\_H.dat -p /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr22\_map.ped

## merlin

“There are two scoring functions of nonparametric linkage in GENEHUNTER, NPLpairs and NPLall. NPLpairs simply calculates the number of pairs of alleles from distinct affected pedigree members that are IBD. NPLall, introduced by Whittemore and Halpern (1994a), puts extra weight on three or more affected pedigree members who are IBD (Zhao, 1999 - doi: 10.1086/302607).”

for chr in {1..22} ; do /hpf/largeprojects/tcagstor/users/marlam/Programs/merlin-1.1.2/executables/merlin -d /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map\_H.dat -p /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map.ped -m /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0042/Family0042\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map\_merlin.map --pairs --ibd --npl --pdf --markerNames --prefix Family0042\_chr${chr}\_NoMaf ; done

for chr in {1..22} ; do /hpf/largeprojects/tcagstor/users/marlam/Programs/merlin-1.1.2/executables/merlin -d /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map\_H.dat -p /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map.ped -m /hpf/largeprojects/tcagstor/users/marlam/ASD/MSSNG/ASD\_Families/Family\_0627/Family0627\_NoMonomorphic\_NoLDRegions\_LD0.1\_GenoMind\_chr${chr}\_map\_merlin.map --pairs --ibd --npl --pdf --prefix Family0627\_chr${chr} ; done