

# Design choices in Haplin

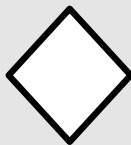
## DESIGN: CASE-CONTROL

**CASE**



**CT**

**CONTROL**



**CC**

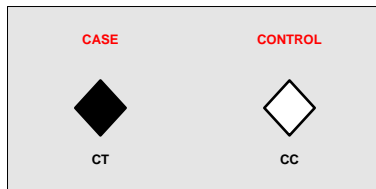
**Genotype case children and control children**

# EXAMPLE: CASE-CONTROL DESIGN WITH A SINGLE CT SNP

## Simplest possible approach:

- Compare frequencies among cases and controls
- For instance, T may have a higher frequency among cases than controls
- Use, for instance, logistic regression, three levels: CC, CT, TT

| Cases | Controls |
|-------|----------|
| CT    | CC       |
| TT    | CC       |
| TT    | TT       |
| CC    | CC       |
| CT    | CT       |
| TT    | CC       |
| ⋮     | ⋮        |

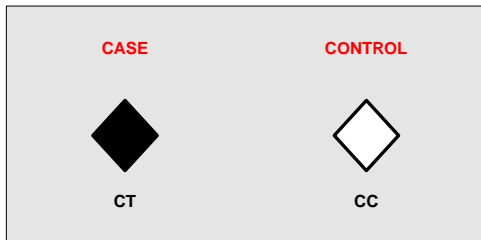


# EXAMPLE: CASE-CONTROL DESIGN WITH A SINGLE CT SNP

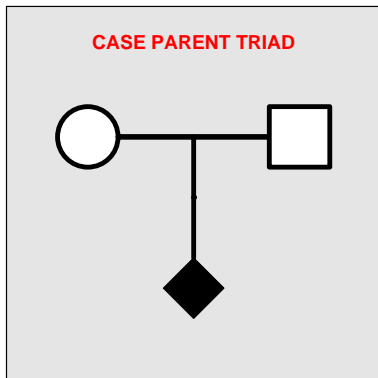
Possible problem:

- **Population stratification**
- That is, cases and controls may come **from different subpopulations**
- Which may cause confounding between SNP and subpopulation

| Cases | Controls |
|-------|----------|
| CT    | CC       |
| TT    | CC       |
| TT    | TT       |
| CC    | CC       |
| CT    | CT       |
| TT    | CC       |
| ⋮     | ⋮        |

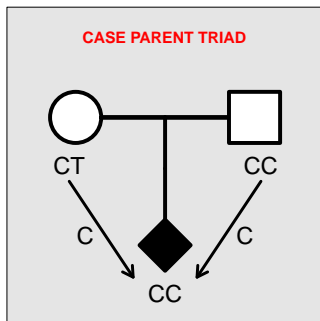


# CASE-PARENT TRIAD DESIGN



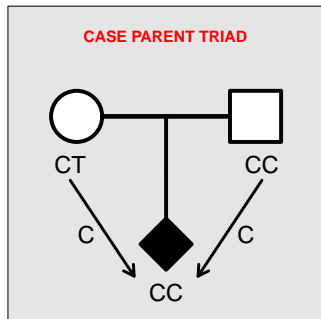
- **Objective:**  
Detect association between candidate gene and disease
- Sample case child with parents

# CASE-PARENT TRIAD DESIGN: TRANSMISSION-DISEQUILIBRIUM TEST (TDT)



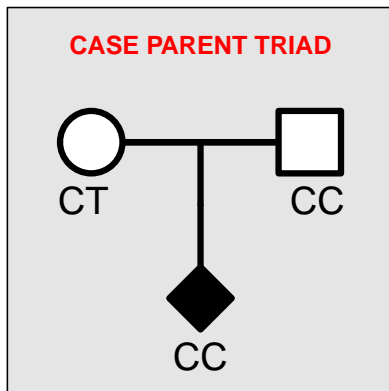
- **Objective:**  
Detect association between candidate gene and disease
- Sample case child with parents
- **TDT:**  
In retrospect: is  $A_1$  transmitted more often than  $A_2$ ?  
(Conditional on affected child)
- **TDT:** Count transmitted alleles, use standard  $\chi^2$ -test

# CASE-PARENT TRIAD DESIGN: HAPLIN MODEL



- **Objective:**  
Detect association between candidate gene and disease
- Sample case child with parents
- **Haplin model:**  
Estimate relative risk (RR) with confidence intervals
- **Haplin model:**  
Compute p-values for each RR and for total effect (LRT)

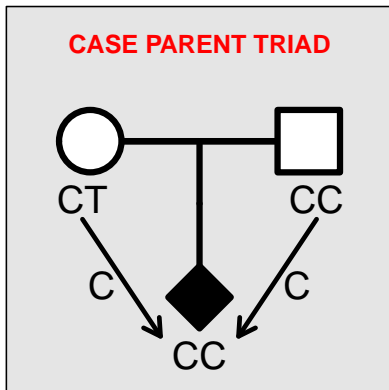
# TRIADS: IDENTIFY TRANSMISSION



??



# TRIADS: IDENTIFY TRANSMISSION

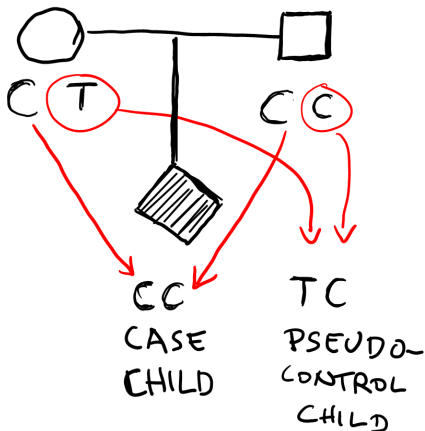


“C” from the mother  
“C” from the father

# PSEUDO-CONTROLS

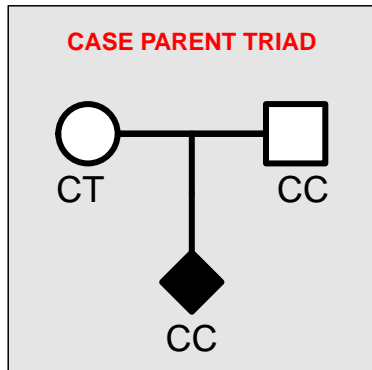
We can, to some extent, think of triads in terms of pseudo-controls (non-transmitted alleles)

## CASE TRIAD



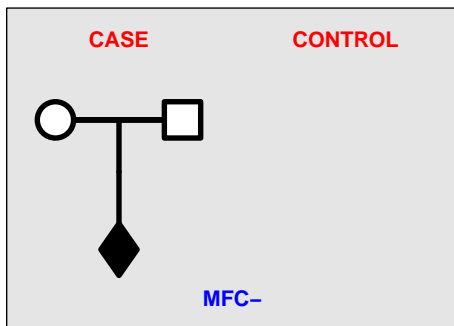
## DESIGN: TRIAD (TRIO) WITH TIME-TO-EVENT DATA

- Cohort data
- Genotype child and its parents
- Outcome is time-to-event

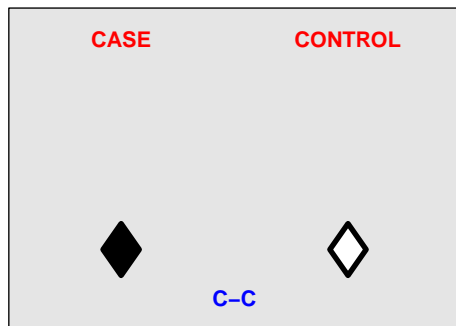


# THE STANDARD COMPARISON: TRIAD VERSUS CASE-CONTROL

Case-triad



Case-control



## WHAT EFFECTS SHOULD WE/CAN WE LOOK FOR?

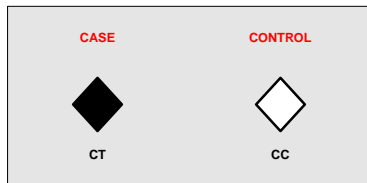
|                              | Case-triad | Case-control | Time-to-event |
|------------------------------|------------|--------------|---------------|
| Fetal genes                  | Yes        | Yes          | Yes           |
| Maternal genes               | Yes        | –            | Yes           |
| Parent-of-origin             | Yes        | –            | Yes           |
| Fetal-maternal interaction   | Yes        | –            | Yes           |
| Main effect of environment   | –          | Yes          | Yes           |
| Gene-environment interaction | Yes        | Yes          | Yes           |

**NOTE:** Gene-environment interaction is possible in all, but only for the genetic effects available in that design.

# CASE-CONTROL DESIGN

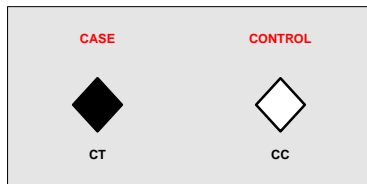
## Advantages:

- Statistical **Power** is (usually) high
- Estimate **Odds Ratio** (OR) associated with **fetal** SNPs and haplotypes
- Estimate **Odds Ratio** (OR) associated with **maternal** SNPs and haplotypes
- Estimate **environment main effect**
- Estimate **gene-environment interactions**
- Simplest analyses can be done with standard logistic regression software



## Disadvantages:

- **Population stratification**, must be controlled for
- Possible **confounding** between fetal and maternal genes
- Loss of power when **reconstructing haplotypes**
- Vulnerable to genotyping errors



## CASE PARENT TRIAD (TRIO) DESIGN

### Advantages:

- Parents are often easier to recruit than independent controls
- Easy to reconstruct **haplotypes**
- Easier to detect data problems (**Mendelian inconsistencies**)
- Estimate **Relative Risk** (RR) associated with **fetal** SNPs and haplotypes
- Estimate **Relative Risk** (RR) associated with **maternal** SNPs and haplotypes
- Fetal and maternal genes “adjusted” for one another
- Avoid **population stratification**
- Estimate **parent-of-origin** effects
- Estimate **gene-environment interactions**



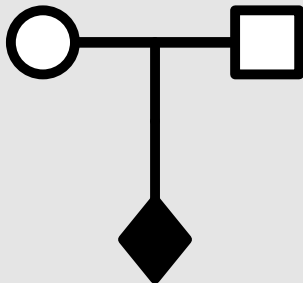
# CASE PARENT TRIAD (TRIO) DESIGN

## Disadvantages:

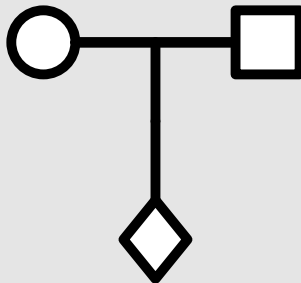
- Power sometimes lower than the case-control design
- No estimate for **environment main effect**
- Parents may not be available for late onset disease

# HYBRID: CASE-PARENT TRIADS + CONTROL-PARENT TRIADS

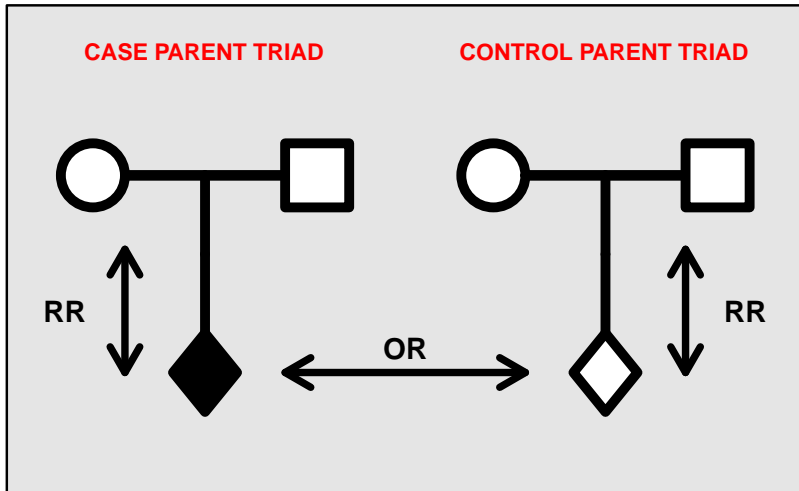
CASE PARENT TRIAD



CONTROL PARENT TRIAD



# HYBRID: CASE-PARENT TRIADS + CONTROL-PARENT TRIADS



## THREE FUNDAMENTAL HAPLIN DESIGNS

Case-parent triads: `design = "triad"`

Case-control: `design = "cc"`

Hybrid: `design = "cc.triad"`

### NOTE:

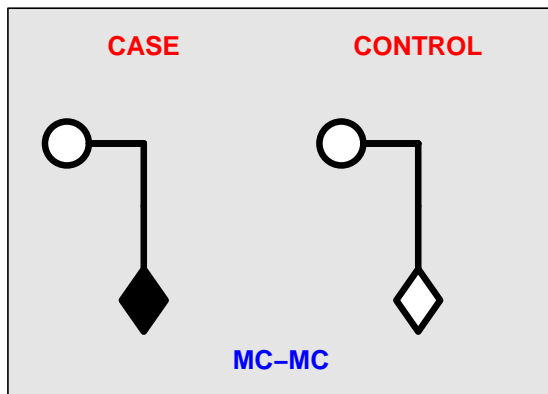
Other designs obtained by setting family members missing (NA) in data file:

CASE-MOTHER DYADS: `design = "triad"`  
with case fathers missing

CASE-PARENT TRIADS + CONTROL CHILDREN: `design = "cc.triad"`  
with control parents missing

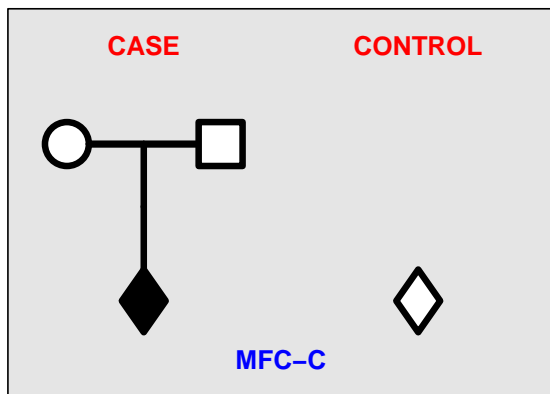
CASE-MOTHER DYADS + CONTROL-MOTHER DYADS: `design = "cc.triad"`  
with case and control fathers missing

## HYBRID: CASE-MOTHER DYADS + CONTROL-MOTHER DYADS



- Mothers often used as case-controls (without children) to test for maternal effects.
- This may result in confounding.
- **In Haplin:** Use `design = "cc.triad"`, set fathers as missing in file.

## HYBRID: CASE-PARENT TRIADS + CONTROL CHILDREN



- Case-parent triad “fortified” with independent controls.
- **In Haplin:** Use `design = "cc.triad"`, set control parents as missing in file.

# WHICH ALLELE ESTIMATES WHICH EFFECT?

## CASE TRIAD

