Cheatography

Genetics Cheat Sheet

by dolly via cheatography.com/183950/cs/38348/

Historical views of heredity and inheritance

Bricks and mortar theory by Hippocrates

- Elements are originated from all parts of the body and became concentrated in male semen. This will be developed and formed into human in the womb
- Inheritance is an acquired characteristics

Blueprint theory by Aristotle

- Transmission of information from parents to offspring
- Heredity is partly assymetric
- Transmission is particulate (definitely one trait or another)

Lamarckian Inheritance

- To explain while some features persisted while others disappeared
- Traits acquired/ lost when depends on need

Modern genetics introduced by Darwin and Mendel

Darwin's blending inheritance

- Offspring inherit the parents' average characteristic
- All parts of the parents can contribute to the evolution and development of the offspring (Pangenesis)

Modern genetics introduced by Darwin and Mendel (cont)

Mendel's particulate inheritance

1. Law of segregation
2. Law of independent assortment
3. Law of dominance

Exception to independent assortment - Linked genes

Genes can be linked together if it is located close together on the same chromosome

Rediscovery of Mendel's work

Allele

One of two or more versions of DNA sequence at a given genomic location

Conflict between Mendelian and Biometrician

Debates between Mendelian and Biometrician

- Do the hereditary and evolutionary properties for a trait like height were the same as those for Mendel's peas?
- Whether inheritance of complex trait was by 'blending' of parental phenotypes (Darwin) which was seen as different to the inheritance of discrete characters as in Mendel's peas

Conflict between Mendelian and Biometrician (cont)

Biuometrician's claim

Traits are continuous (Blending inheritance) and heritable

Mendelian's claim

Mendelian genetics work in inheritance

Achondroplasia

<90cm height

Marfan Syndrome

>200cm height

Emergence of Biometrical genetics

Whar are Galton's claims in trait hereditary?

1. Traits like height, weight, arm length are normally distributed, not binary
2. Traits are resemble between parents and offspring

Galton's claim(1): Traits are "-normally distributed" means

1. A trait has a mean value given a population
2. A trait can be subject to mathematical transformation

What Galton use to study continuous variation in organism?

- Regression
- Correlation

Emergence of Biometrical genetics (cont)

Galton's claim(2): Traits are resemble between parents and offspring

- When measuring the height of parents and offspring, the mid-parental height has almost no deviation to their offspring height
- Therefore, traits are resemble between parents and offspring

Brownlee's Multi-Gene model to explain Mendelian inheritance in the blending inheritance model

1. Parents and Child: 0.5 correlation because parents transmit 50% of their genome to their child
2. Parents and Grandchild: 0.25 correlation
3. Parents and Great-grandc-hild: 0.125 correlation
- These correlations are based on Mendelian's segregation law

Polygenic Model

Fisher`s Infinitesimal model

- Polygenic inheritance :A quantitative trait could be explained by Mendelian inheritance if several genes affect the trait
- Include additive and dominant factors
- Resemblance between relatives occur due to their genetic covariance

Polygenic Model (cont)

Fisher's single locus model

Assume:
1. Dominant allele is not known
2. A locus either follows dominant or additive
- Used to determine whether the locus follow dominant or additive

Fisher's single locus model, assume A2A2= -a and A1A1= a

1. If d>0 : A1 is dominant to A2
2. If d<0 A2 is dominant to A1
3. If d=a/-a: Complete dominance (Heterozygote)
4. If d=a/-a: Over-dominance
5. If d=0: Locus is additive

Continuous distribution of quantitative traits

Alleles in our genome is limited but environmental factors are not. Therefore, traits are also influenced by environmental factors

Fisher's partitioning variance

Genetic and non-genetic factors

Three genetic factors (G)

1. Additive (A)
2. Dominance
3. Epistasis: Interaction between additive factors/ additive - dominant factors

Non-genetic factor

Environment (E)

Polygenic Model (cont)

Phenotypic variance (P)

Interaction between genetic and environmental factors (GxE)
2. P= G+E+GxE

Heritability

How much of the variation in a trait is due to variation in genetic factors (G)

Genetic Architecture

- Composition of various genetic factors upon a phenotype
- Include additive, dominance and epiptasis

Genetics

To identify genetic factor associated with traits/disease but also study the contribution of a genetic factors

Trait is not dichotomy (contrast between two things)

The features of an organisms are due to the individual's genotype and environment

Allele

Allele and Allele frequencies

- Proportion of chromosomes in population carrying the allele of traits/disease
- Different combination of alleles determine traits or diseases
- Allele frequencies indicate the proportion of observed genotypes in a given population

Allele (cont)

Allele transmission to next generation

- Same with Mendel's first law
- Totally independent and not influenced by environmental factors

Hardy-Weinberg Equilibrium

Do segregation in Mendelian inheritance law affected by the segregant (allele)?

No. This is called "stable"

Hardy-Weinberg principle

- Assumed that allele frequencies will not change from generation to generation
- p2+2pq+q2=1
- p+q=1

Assumptions of Hardy-Weinberg equilibrium

1. Random mating
2. No natural selection
3. Equal genotype frequencies in two sexes
4. No mutation/migration
5. No differential viability
6. Infinite population size

However, all of these are not realistic!

Hardy-Weinberg Equilibrium (cont)

Mendelian segregation

Preserved in any organism with sexual reproduction regardless of allele frequency in the population

Chi-Square Test

Chi square

Use statistics to determine whether a locus of interest is under HWE or not

Null hypothesis

There is no difference between observed value and the expected value

Degree of freedom (DF)

- Number of phenotypic possibilities in the cross
- Example DF: 3(AA,Aa,aa)-1=2
- If the level of significance read from the table is greater than 0.05/5%, the null hypothesis is not rejected

When the null hypothesis is supported by analysis

Assumptions
1. Mating is random
2. Normal gene segregation
3. Independent assortment

Chi-Square Test (cont)

When the null hypothesis is not supported by analysis

Assumptions
1. Non-random occur
2. Genes are not randomly segregating because they are linked on the same chromosome/inherited together.

Introduction of Heritability

Model to describe heritability

1. Fisher's model
2. Falconer's model

Falconer's Model

Mathematical formula used in twin studies to estimate the relative contribution of genetics vs environment to variation in a particular trait
- Heritability of the trait based on the difference between twin correlations
- Heritability=2(rMZ-rMD)
- Where r=concordance of the phenotype, MZ=Monozygotic Twins, DZ= Dizygotic twins

Heritability

Phenotypic similarity in family depends on

1. Genetic relationship
2. Traits

Heritability (cont)

Total phenotypic variance for a character?

- VP=VG+VE
- Combined effects of genotypic and environmental variance

Genetic variance (VG)

- The variance among the mean phenotypes of different genotypes
-Additive genetic varian-ce(VA): Variation due to the additive effects of alleles
- Dominance genetic variation (VD): Variation due to dominance relationships among alleles
- Epistatic genetic variation (VI): Variation due to intera-ctions among loci

Environmental variance (VE)

The variance among phenotypes expressed by replicate members of the same genotype
- Differences between monozygotic twins are due to environmental factors

Dominance genetic variance (VD)

Due to dominance deviations which describe the extent to which heterozygotes are not exactly intermediate between the homozygotes

Heritability (cont)

Additive genetic variance (VA)

- Responsible for the resemblance between parents and offspring
- The basis for the response to selection

Degree of relatedness and the components of phenotypic covariance

1. Identical twins: VA+VD+VE
2. Parent-offspring:1/2VA
3. Full siblings:1/2VA+1/4VD+VE
4. Grandparent-Grandchil-d:1/4VA

Heritability of a trait

- A measure of the degree resemblance between relatives
- Estimates the degree of variation in a phenotypic trait in a population that is due to genetic variation between individuals in that population that is due to genetic variation between individuals in that population

Narrow sense heritability (h2)

The proportion of trait variance that is due to additive genetic factors

Broad sense heritability (H2)

The proportion of trait variance that is due to all genetic factors including VD, VA, VI

Normal Distribution

Two quantities that describe a normal distribution

1. Mean
2. Variance

Deviation

- Distribution of a trait in a population= Proportion of individuals that have each of the possible phenotypes
- In normal distribution, half points are above and half points are below mean
- One standard deviation are located in the mean
- The distribution of a trait in a population implies nothing about its inheritance

Covariance

A measure of the joint variab-ility of two random variables (trait)
- Example: Measure the height deviation of father and son in a population

Resemblance between family members

- When there is genetic variation for a character, there will be a resemblance between relatives
- Relatives will have more similar trait values to each other compared to unrelated individuals

Normal Distribution (cont)

Resemblance between relatives

-Depends on the degree of relationship
- Use slope, not correlation coefficients to compute resemblance of family members
- Identical twin=100%, Full siblings=50%, Parent-offspring=50%, Half-sibling=25%

Morgan Experiment

Morgan's experiment

Proved that chromosomes are the location of Mendel's heritable factors from his fly experiment

Centimorgan

The frequency of crossing over

Linkage Map/ Genetic Map

If the frequency of how often genes crossover is known, the percentage can be used to estimate how far apart the genes are from one another on a chromosome

Genes

Definition of gene

- A core unit of the heredity that control the development of a trait
- Mendel's "discrete particle" in particulate inheritance actually indicated the concept of "gene"
- Those consist of DNA sequences and produces functional elements

Genes (cont)

How many genes can make protein in human?

20,000 genes make proteins and most of them involve in determining traits

Genotype

The part of the genetic makeup of a cell which determine one of its characteristics

Phenotype

The set of observable characteristics of an individual resulting from the interaction of its genotype with the environment

Components in a gene

Gene contain exons, introns, UTRs and promoter in its transcript
- Gene can have various transcripts due to alternative splicing

Exon

A region of a trascribed gene present in the final functional RNA molecule

Intron

Any nucleotide sequence within a gene that is removed by RNA splicing during maturation of the final RNA product

Genes (cont)

UTR

Either of two sections, one on each side of a coding sequence on a strand of mRNA

Promoter

The section of DNA that controls the initiation of RNA transcription as a product of a gene

Cells and Chromosomes

Where does genetic recombination occur in meiosis?

In meiosis I and it occur between Prophase I and Metaphase I

Pros of asexual reproduction

- Produce more offspring as it takes less time
- Require less energy

Cons of asexual reproduction

- No variation in offspring
- Less variation in population
- Mutation can slightly increase variations
- Fragile to environmental change

Pros of sexual reproduction

- Increase variation in offspring
- More resistant to many environmental forces because of genetic variation

Cells and Chromosomes (cont)

Cons of sexual reproduction

- Require two organisms for mating
- Requires more cellular energy
-More time required for offspring development

Elements in Chromosomes

Ploidy

Number of homologous sets of chromosomes in a cell

Locus

A fixed position on a chromosome that may be occupied by one or more gene

The nuclear genome

Consist of 6 billion nucleotides in 46 chromosomes

Chromosomes

Hereditary factors

Genes and allele that are located on chromosomes

Autosomal chromosomes/ autosomes

Pairs number 1 to 22

Sex chromosomes/ somatic cells

Pair number 23

Mitochondrial chromosomes in mitochondria

- Haploid
- Maternal transmission

Chromosomes (cont)

Karyotype

- Visualize chromosome shapes, structures and behaviors of chromosomes during cell division
- Autosomes in metaphase are arranged from the longest to shortest and from number 1 to 22
- Chromosomes number 23 are either XX/XY
- p arm=short, q arm= not p

Mutation

Saltationists

Claim that evolution take place suddenly (saltating)(so that change instantaneous transition into a new species

Gradualists

Believe gradual process of evolution given large-scale variability in a population

Gene can be defined in terms of their behavior as fundamental units based on:

1. Hereditary Transmission
2. Genetic recombination 3. Mutation
4. Gene function

Mutation (cont)

Darwinian view on mutation

- Most mutations have an impact on certain traits
- Natural selection is the primary force of evolution

Post-Mendelian geneticists' view on mutation

Natural selection plays little or modest role but occurrence of mutation would be a major evolution force

"Hopeful Monster" hypothesis by Richard Goldschmidt

- Macroevolution through macromutations
- Called "Hopeful Monsters" because they were the embodiment of large phenotypic changes that had the potential to succeed as new species (saltation)
- Change early development and thus cause large effects in the adult phenotype

Developmental macromutations

Mutations in developmentally important genes could produce large phenotypic effects

Mutation (cont)

Neo-Darwinism

- Natural selection is assumed to play much more important role than mutation
- Creating new characters in the presence of genetic recombination

Kimura's view: Neutral mutation

- The rate of substitution is so high that if each mutation improved fitness, the gap between the most fit and typical genotype would be large
- This rapid rate of mutation means that the majority of the mutations were neutral
- Mutations had little/ no effect on the fitness of the organism
- Not all mutations affect on/ completely determine our trait, including diseases

Mutation is an old term

Describe the situation for permanent change in evolut-ionary process

Mutation (cont)

Variant

- The change in the nucleotide sequences
- Since a change in nucleotide sequence may not be permanent, variants are often called: genetic variant, variation or genetic variation

Polymorphism

Describe a variant with a frequency above 1% but broadly variants that we know the frequency in certain population

Mutation

Saltationists

Claim that evolution take place suddenly (saltating)(so that change instantaneous transition into a new species

Gradualists

Believe gradual process of evolution given large-scale variability in a population

Mutation and Population

Out of Africa Theory

- Explains the origin of modern human beings
- A small subset of this population migrated out in the past 100,000 years and rapidly expanded throughout a broad geographical region

Mutation and Population (cont)

Non-Afracan populations have different variant frequency due to

1. Bottleneck
2. Long migration history

Coalescent Theory

- Two sample lineages find common ancestor
- A model how an allele sampled from a population may have originated from a common ancestor

Stochastic

When coalescence occurs is a stochastic (random probability(process

Genomic study of population structure

Implications of HapMap project and 1000 Genome Project

- Variant frequency is uniquely represented in each population so can identify the population structure
- Genomic data are useful and fundamental resource to identify genes associated with disease and genetic variant in patients

Genomic study of population structure

Implications of HapMap project and 1000 Genome Project

- Variant frequency is uniquely represented in each population so can identify the population structure
- Genomic data are useful and fundamental resource to identify genes associated with disease and genetic variant in patients

Genetic variant by size

SNV - Single Nucleotide Variant

- Substitution of one/another base pair at a particular location in the genome
- Also called SNP if the allele frequency in a population is known
- A point mutation because it only affects a single nucleotide of nucleic acid
- There are ~3,500,000 SNVs per individual (more in African)
- Everyone have different compositions of SNVs so there us variability in traits
- The ratio of heterozygous and homozygous SNVs is ~2:1

Genetic variant by size (cont)

Indel - Insertion/Deletion

- 1-1000bp changes in our genome
- There are ~300,000 to 600,000, indels per individual (more in African)
- Less than SNVs as indels have a large phenotypic effect than SNVs so more selective pressure

Indels can be divided to

1. Microsatellite polymorphism
2. Mobile element insertion polymorphism

Microsatellite polymorphism

2-4 nucleotide unit repeated in tandem 5-24 times

Mobile element insertion polymorphism

Cause human genetic diversity through retrotransposition
- Involves transcription into RNA
- Reverse transcription into DNA sequence
- Insertion into another site in genome

SV - Structural variant

- A genomic change >1000bp

Genetic variant by size (cont)

SV can be divided to

1. Copy Number Variant (CNV) - Deletion/Duplication
2. Copy Number Neutral Variants (CNNV) - Inversion/Insertion/ Translocation

Small variants

SNVs and indels

Large variants

SVs

SV in the gnomAD project

- Represent population structure as small variants
- More singleton SVs are observed in larger SVs
- Singleton: The variant only seen in an individual (rare) Rare: It's under strong natural selection so only seen in few individuals
- Size of SVs are correlated with the effect size of SVs

Genetic variant by size

SNV - Single Nucleotide Variant

- Substitution of one/another base pair at a particular location in the genome
- Also called SNP if the allele frequency in a population is known
- A point mutation because it only affects a single nucleotide of nucleic acid
- There are ~3,500,000 SNVs per individual (more in African)
- Everyone have different compositions of SNVs so there us variability in traits
- The ratio of heterozygous and homozygous SNVs is ~2:1

Genetic variant by size (cont)

Indel - Insertion/Deletion

- 1-1000bp changes in our genome
- There are ~300,000 to 600,000, indels per individual (more in African)
- Less than SNVs as indels have a large phenotypic effect than SNVs so more selective pressure

Genetic variant by Frequency

Selection and Frequency	- Natural selection work on trait so the frequency of variants that contribute to trait can be changed - Level of natural selection is varied by traits and diseases - Some traits are favored by selection, therefore, the frequency variants increase - According to Polygenic model, a single variant is lilkely contributing partially/highly partially to a trait. Therefore, there is a wide range of the frequency of variants

Genetic variant by Frequency (cont)

Selection and Allele Frequency	- Allele frequencies can be changed by selection - Increase beneficial alleles and removes deleterious one - Traits not favored over mating are likely under natural selection (high selective pressure) - Natural selection tends to make allele with higher fitness more common over time, resulting in Darwinian evolution

Genetic variant by Frequency (cont)

Fecundity	- Based on fertility ratio (FR) - Lower fecundity: Higher selective pressure on the trait - If a trait is not suited to mating/.reproduction, allele for this trait disappeared in a population - Similar to reproductive fitness

Genetic variant by Frequency (cont)

FR	Calculated based on the number of children individuals in that group had compared with the general population - If a disease have 0.5 FR, they have average half as many children as the general population
Penetrance	The proportion of individuals carrying a particular variant of a gene that also express an associated trait

Genetic variant by Frequency (cont)

Fitness
- Determine the allele frequency in population
- If fitness is not affected by variant, it will be remained in a population, ultimately increasing its frequency

Genetic variant by Transmission

Type of genetic variants by transmission mode
1. Inherited variants
2. De novo variants
3. Somatic variants

De novo variants
- new variants arise during cell division
- different nucleotide changes compared to DNA template
- Errors are not present in genome thus called de novo=new
- Errors in somatic cell: de novo somatic variants
- Errors in germ cells: de novo germline variants

Genetic variant by Transmission (cont)

Mutability/Mutation rates
How much errors are occured during replication

Mutation signatures
The pattern of somatic mutations in disease

Human germline mutation rate
1.0~1.5x 10-8 bp per generation

How many total of de novo variant from mother and father ?
- ~70 de novo variants
- 80% of de novo variants are from father's sperm

Main contributor to de novo variants
- Advanced parental age
- Father is higher than mother- Because spermatog-onial cells continue to divide throughout life which allow the progressive accumulation of mutations due to errors during DNA replication/failure to repair non-replicative DNA damage between cell divisions

Rarest variants
Have greatest potential to carry for disorders

Genetic variant by Transmission (cont)

Variant frequency and its penetrance for disease
- Inverse relationship
- Allele frequency is low but penetrance is high

Genetic variant by consequence

Missense variants
Single base pairs substitution produce different amino acid

Trucating variants
A genetic variant which results in a shorter version of the protein being produced

Nonsense mediated decay
Destroys the mRNA leading to no protein

Noncoding variants
- Variants located outside the coding regions
- Located in promoters, transcription factor binding sites, enchancers

Protein isoform
Protein that are similar to each other and perform similar roles within the cells

Genetic variant by consequence (cont)

Variant annotation
- The process of assigning functional information to DNA variants
- Can be varied by transcript
- A gene can have more than one transcript

Two schemes for variant annotation
1. Per gene annotation: Choose the most critical consequence by the variant per gene
2. Per-transcript annotation: All consequence for every transcript

Linkage Disequilibrium and Haplotype

Linkage disequilibrium (LD)
- Non-random association of alleles at two or more loci in a given population
- LD between two alleles is related to time of the mutation events, genetic distance and population history
- LD around an ancestral mutation on founder chromosome

Haplotype
A group of alleles in an organism that are inherited together from a single parent