# POPULATION GENETICS

Kenneth Alonso, MD, FACP

# Hardy-Weinberg equilibrium

- $p^2 + 2pq + q^2 = 1$ (assumes binary distribution of dominant and recessive traits, a two locus model)
- Used to determine allele frequency and heterozygote carrier frequency.
- Non-random mating, non-constant mutation rates, changes in population structure due to migration affect the equilibrium as do traits that do not provide a selection advantage.

# Hardy-Weinberg equilibrium

- M and N blood groups are co-dominant markers.
- Population frequencies are 0.36 for MM and 0.16 for NN genotypes (and, 0.48 for MN).
- Therefore, the frequency of M gene is 0.60 ([0.36 + 0.36 + 0.48] / 2.0).
- The next generation will continue to reproduce these frequencies.
- Neither gene disappears over time as none provides survival advantage or disadvantage.

# Gene frequency

- Genetic drift is a random process of mutation.
- If a gene causes a change that leads to the death of the carrier before it can be transmitted, the frequency in the population will decrease.
- If a gene leads to longer life and the greater likelihood of transmission, the frequency in the population will increase.

# Gene frequency

- Founder effect is the high frequency of a mutant gene in a rapidly expanding population founded by a small ancestral group when one of the founders was a carrier of the mutated gene (e.g., Old Order Amish, Afrikaner, Finns)
- Ashkenazi are a population bottleneck.
- Inheritance, however, may be multifactorial (polygenic).
- A Gaussian (bell-shaped) distribution of frequency suggests a multifactorial inheritance.
- Twin data may provide clues.

# Concordance rate

- A single gene fully expressed (penetrance) will be fully concordant in identical twins.
- A somatic line mutation may affect only one of the twins.
- A non-genetic trait will have the same frequency in identical and fraternal twins.
- In multifactorial traits, concordance rates will be higher in identical as opposed to fraternal twins.
- Estimation of heritability
- Probability that population variance in a trait can be ascribed to genetic factors
- Calculated as 2x(concordance in identical twins – concordance in fraternal twins).

# Autosomal dominant inheritance

- At least one parent carries the gene inherited.
- No sex predilection.
- Transmitted by either sex.
- No one without the gene can transmit the gene.
- Father to son transmission possible.
- May not be fully expressed in child.

# Autosomal recessive inheritance

- Both parents must have the gene.
- However, neither parent expresses the gene product.
- ather, they are carriers of the gene.
- Increased likelihood of parental consanguinity.
- Affects either sex.
- Father to son transmission possible.

# X-linked recessive inheritance

- Affected males.
- Mother usually asymptomatic carrier and may have affected male relatives.
- Affected males cannot transmit gene to sons.
- All daughters are carriers.
- Unaffected males cannot transmit gene.
- Females can be affected if both parents have the gene.
- Occasionally, females may be affected if the X-chromosome inactivated (Lyonized) is the one that does not contain the gene.

# X-linked dominant inheritance

- Affects either sex
- Female predilection (2:1)
- The child of an affected female has a 50% probability of being affected.
- An affected male transmits the gene to all daughters but not to sons.

# Mitochondrial inheritance

- Inheritance only through females.
- An affected female transmits the trait to all children.
- An affected male never transmits the trait.
- Males and females are equally affected.

# Y linked inheritance

- Genes related to the male specific region of the Y chromosome relate to spermatogenesis.
- Mutations here generally result in infertile males.
- No true Y-linked inheritance.
- A few genes with X related homologues have been mapped to the Y chromosome.
- No disorders have been identified resulting from mutations of those genes on the Y chromosome.
- However, if there were, they would only affect males.
- Affected males would arise from affected sons.
- All sons of an affected father would be affected.

# Molecular basis of genetic dominance

- Gain of function
- Increased gene dosage (duplication, amplification)
- Ectopic or temporally altered mRNA expression
- Increased or constitutive protein activity
- Protective effects
- Altered protein structure
- Toxic protein
- Non-protein functions (exxon shuffling; altered substrate specificity)
- Genomic imprinting

# Molecular basis of genetic dominance

- Loss of function
- Haplo-insufficiency
- Subunit imbalance
- Metabolic rate determining step
- Development regulator gene at threshold
- Regulatory genes at threshold include pax3, pax6, gli3, and c-kit

# Threshold model

- Assumes there is an underlying liability distribution in the underlying population and that the threshold on this distribution must be surpassed before the allele is expressed.
- Those at the lower end of the distribution have few alleles that could be expressed.
- Asymptomatic.
- Those at the higher end of the distribution, above the threshold, likely to express the alleles.

# Expression

- Incomplete penetrance (expression) may result from:
- Location of mutation in gene
- Triple repeat expansion of nucleotides
- Mis-sense mutation
- A single mutation may have pleiotropic effects.

# Expression

- Genetic heterogeneity may result from different mutations at the same allele.
- With varying expression
- Cystic fibrosis, hemoglobinopathy, and muscular dystrophy as examples.
- Genetic heterogeneity may result from mutations at different loci (that rely on a common pathway for expression).
- Retinitis pigmentosa, myotonia congenita as examples.

# Risk of transmission

- The recurrence risk is the probability that an affected offspring will be produced in families in which one or more affected offspring have been manifest.
- 50% for autosomal dominant
- 25% for autosomal recessive
- Multifactorial risk more difficult to calculate.
- Based on observations of large populations.
- Called empirical risks.

# Risk of transmission

- The empirical risk is more elevated if:
- More than one family member is more affected
- If the expression of the allele is more severe in the proband
- If the proband is of the less commonly affected sex (gene dosage).
- The empirical risk is lower in more remote relations.
- May have significant environmental contribution.
- Studies of identical twins separated at birth and raised separately used to minimize environmental contribution to gene expression.

# Examples of shared loci

- Age related macular degeneration related to CFH gene abnormality
- CDKN2A and B involved in coronary disease, type 2 DM, invasive melanoma
- CDKAL1 cell cycle regulator involved in type 2 diabetes and in prostate cancer
- GIP found in 80% Europeans, 5% Africans; worse prognosis in type 2 DM in blacks

# Examples of shared loci

- ATG16LI autophagy abnormality in Crohn’s
- ORMDL3 in childhood asthma and Crohn’s(unknown function)
- LRRK2 in Crohn’s and Parkinson’s
- JAFZ1 height, type 2 DM, prostate cancer
- C10-ORF67 in sarcoidosis and celiac disease
- KITLG in testicular carcinoma, blond or brown hair