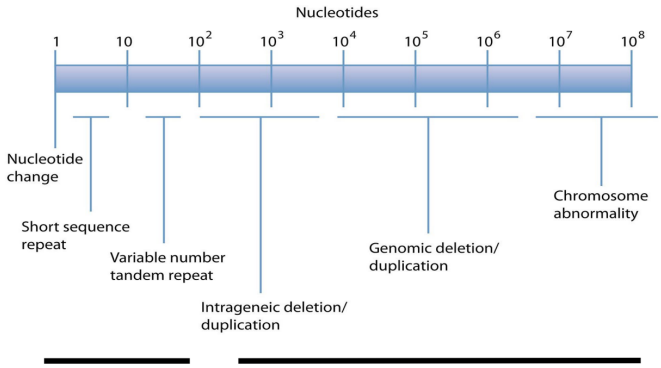
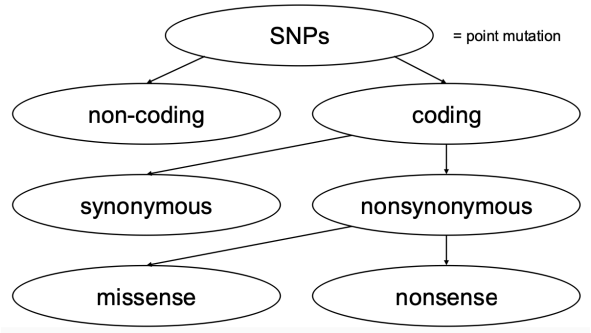
Concept Name	*Page*	*Definition/Notes*
Variation	2	Phenotypic differences; The range of differences that are between individual organisms; Differences between organism; Between different species or within same species but different breeds. Most variation is caused by both gene and environment.
Genetic Variation	2, 37, 54	***3 very similar definition:*** 1. Differences caused by genes within the same species in individual or in population. 2. Genetic variation is the difference in DNA among individuals or the differences between populations among the same species. 3. The presence of differences in DNA among individuals and between populations. ***Examples:*** Blood type is a result of Genetic Variation. IS NOT Gene Mutation (后果/发生的数量/发生的区域). ***Includes:*** 1. Point mutation 2. Copy number variants ***Sources:*** Mutation and genetic recombination. Mutations are the ultimate sources of genetic variation, but other mechanisms, such as genetic drift contribute to it, as well. ***Significance:*** 1. SURVIVE OR EXTINCTION: Genetic variation allows a population to adapt to environmental changes. 2. DIVERGENCE OR CONVERGENCE: Genetic variation drives evolution and leads to biodiversity. ***Features:*** 1. Most variation is EVOLUTIONARY neutral. 2. Traits under negative selection will be largely due to rare variants. 3. Traits not under negative selection will be at least partly explained by common variants.
Environmental Variation	2	Differences caused by environment.

DNA Sequence Variants	3	
DNA Sequence Mutation	5	= DNA damage DNA sequence mutation causes genetic variants Mutation type: benign/pathogenic/uncertain. Basically not include **Chromosome Abnormality**
Chromosome Abnormality	44	(in here) = Chromosome Structure Abnormality = Chromosome Rearrangements • Deletion • Duplication • Inversion • Dicentric/acentric chromosome • Isochromosome • Ring chromosome • Translocation
All nDNA bp (per cell)	3, 46, 51	**Genome sizes** Species / No. of base pairs Arabidopsis 125 million Drosophila 180 million Rice 400 million Maize 2,500 million Human 3,000 million Barley 4,900 million Wheat 16,000 million 3 billion, 10^9, 30 亿, = 2 meters 20,000~25,000 genes About the same number of genes as mice, most of the common genes share the same intron and exon arrangement 100,000 proteins (predicted 100,000 genes) gene average size: 3000bp gene size varies greatly: longest is 2.4 million bp 50% genes have known function Chromosome 1 has the most genes (2968), and the Y chromosome has the fewest (231). 24 linear molecules

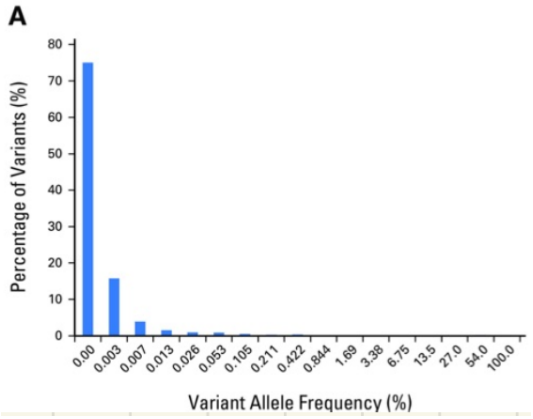
All mtDNA bp (per cell)	/	16Kb
AA-changing Variants (per/individual)	5	10,000-11,000 by Genome Project Consortium 2010
non-AA-changing Variants (per/individual)	5	10,000-12,000 by Genome Project Consortium 2010
Variants (per/individual)	5	20,000-23,000 by Genome Project Consortium 2010
Point Mutation	3, 10, 37, 55	= Bases Mutation Types of point mutation: 1. Silent mutation → SNP 2. Missense mutation → SNP (conservative/non-conservative) 3. Nonsense mutation → SNP 4. Frameshift mutation (indel, duplication), sometimes not regarded as a type of point mutation. 1: Synonymous, mostly! 2,3: Non-synonymous 
Silent Mutation	4	Changed codon, same AA
Missense Mutation	4	Changed codon, different AA, but can be conservative. But not a stop codon.
Conservative Mutation	4	AA with similar chemical properties
Nonsense Mutation	4	= Stop Mutation
Frame-shift Mutation		Not 3 fold. Often lead to non-sense mutation.
SNP	4, 19, 35, 37, 39, 49, 51, 59, 60,	= Single Nucleotide Variants Definition: SNP is a substitution of a single nucleotide that occurs at a specific position in the genome. Certain definitions require an appreciable frequency (typically > 1%), because too rare SNPs will be easily lost over generations. (a result of) single base substitution. Week6: Single base mutation in DNA

61，73，74 Single nucleotide sequence variants

Frequency of SNP is very high:
1. 1 in 31 bp in non-coding regions | 1 in 124 bp in coding regions
2. On avg, one in 300 bases.
3. Occur 0.5-10 per every 1000 base pairs

It is the most common genetic variation
Most simple form of genetic polymorphism

90% of sequence variants are differences in SNPs
90% 的杂合性是由常见、共有的变异导致的
>1,000,000 (1 million) SNPs identified
SNP Variants frequency: most are very low



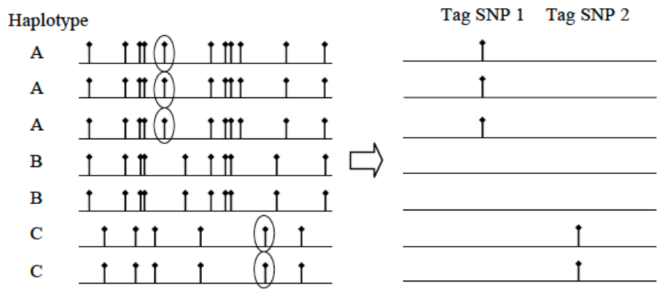
99.9& nucleotide bases are the same in all people

Not uniformly distributed

Tag SNP 的来源
By typing an adequate density of SNPs, one can identify tags that capture the vast majority of common variation in a region.

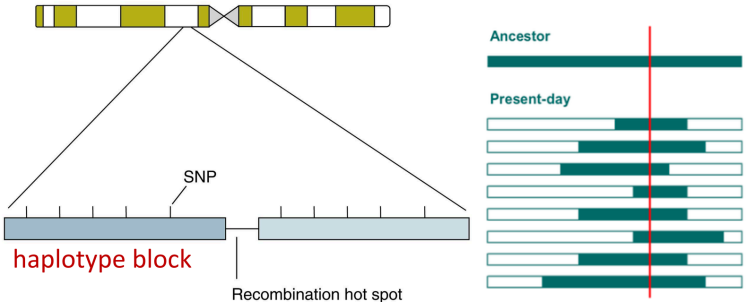
Tag SNP (SNP Marker): The genome is organized into 10-20 kb haplotype blocks (haplotype) that are transmitted almost intact from generation to generation, separated by "hot spots" where recombination is most likely to occur.

SNPs within a block (haplotype) tend to be in linkage disequilibrium with one another; therefore, a single SNP ("tag SNP") can be used as a marker for the entire group within a block.



How exactly do we make use of SNP marker?
Two strain of C.elegans: N2 and CB. They are two different alleles made up of 48 SNPs. Luckily, we identified SNPN2 and SNPSB that can and cannot be cleaved by the same REase (Dral). And they can be used to identify the allele (a form of haplogroup in this case) in that way.

A haplotype (haploid genotype) is a set of linked SNP that tend to always occur together.
Haplotype: a collection of specific alleles (or SNPs, or satellites or any genetic markers) for a cluster of tightly linked loci on a chromosome
A **haplogroup** is a group of similar haplotypes that, share a common ancestor with a SNP mutation, which can be used to define genetic populations.
The distribution of different allele combinations are at very close markers → reflects the ancestral haplotype.



The rsID (Reference SNP cluster ID) is an accession number used by researchers and databases to refer to specific SNPs.

Why we want to figure out TagSNP/neighboring SNP/haplotype?
Because causal variant may not be genotyped due to sequencing oversight. But the effect of them can be inferred by genotyping neighboring SNPs. This is because neighbors must be in linkage disequilibrium with the causal variant of interest.

Haplotype Map of Human Genome/International HapMap Project

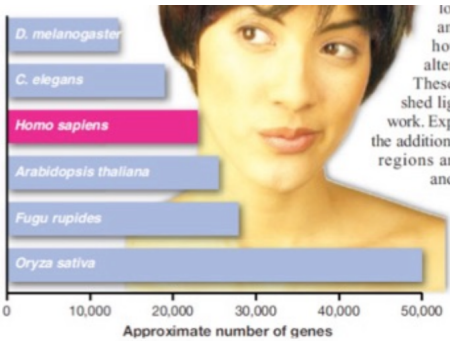
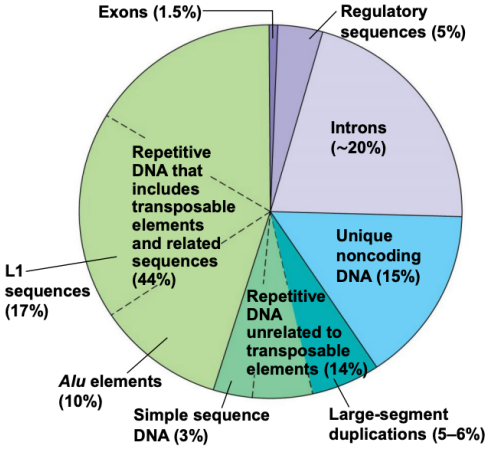
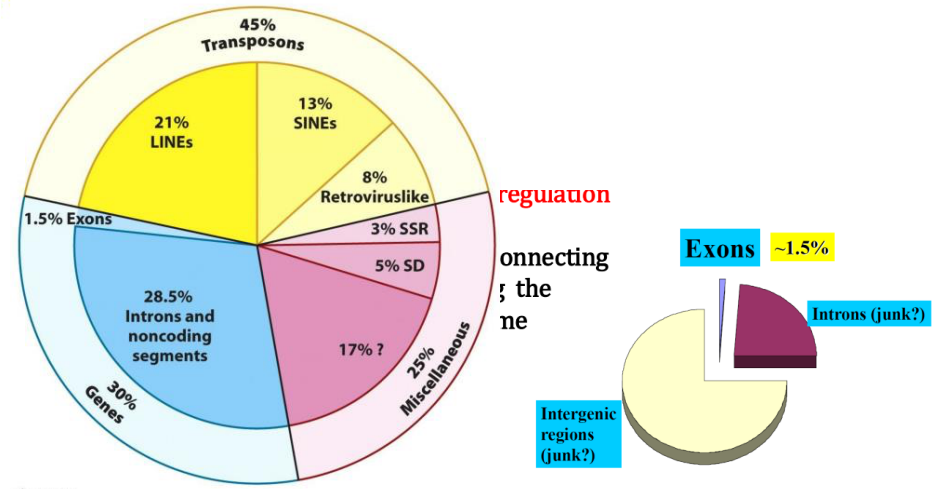


Goals:
• Define haplotype "blocks" across the genome | Genome中的block
• Identify reference set of SNPs: "tag" each haplotype | Block中的tag
• Enable unbiased, genome-wide association studies

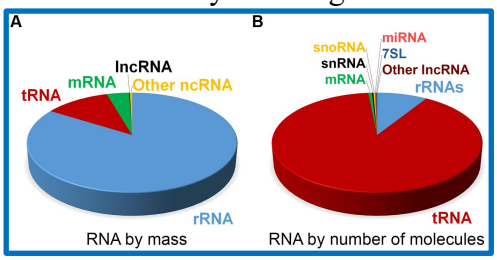
90% chromosomes differs due to the differences in few common haplotypes.

可以被NGS一下子检测到

Splice Acceptor	4	In the end of intron, usually the 3' end of RNA. 起到 Splice Donor 切割后连接外显子的作用 Cryptic splice acceptor: 除非发生了 mutation，不然不应该被使用

Splice Donor	5	In the start of intron, usually the 5' end of RNA. RNA splicing 中 exon 和 intron 连接的起始点
Genome Made-up (Composition of human genome)	4, 51, 82	~1.5% of the genome codes for proteins (whole exome sequencing) Genes comprise less than 5% of the genome. Only 1.5% of the total genome codes proteins Human has so few genes yet so "advanced":    What is the function of the “junk”? • Regulatory roles in gene regulation • Structural roles such as connecting adjacent genes, influencing the structure of the chromosome • Some regulate gene expression • Some encode RNA: to regulate gene expression, stabilize the genome • Historical junk (mobile elements, gene fragments, evolutionary building blocks?)

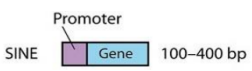
RNA encoded by human genome: coding-RNA & non-coding RNA

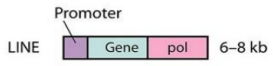
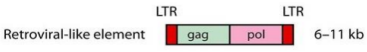
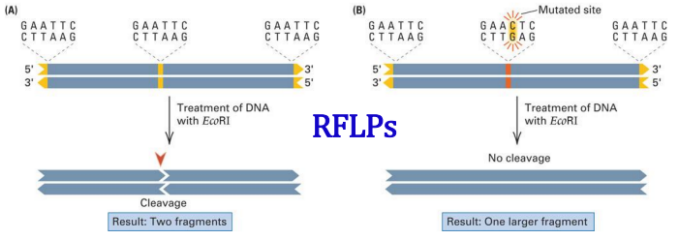


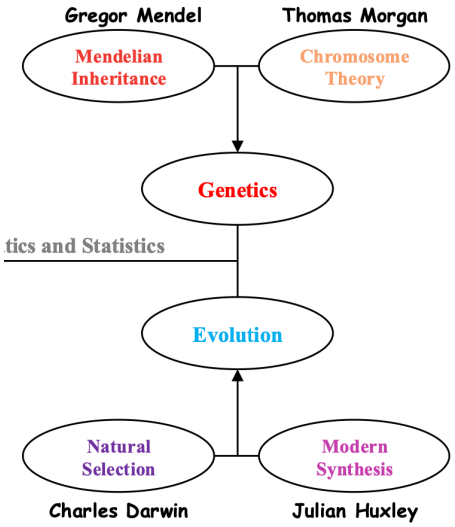
Non-coding RNA		Length (nt)	Species	Function
Ribosomal RNA (rRNA)		120~4700	All	Translation
Transfer RNA (tRNA)		70~100	All	Translation
Small nuclear RNA (snRNA)		70~350	Eukaryote	Splicing, mRNA processing
Small nucleolar RNA (snoRNA)		70~300	Eukaryote, archaea	**RNA modification, rRNA processing**
miRNA	Small ncRNA	21~25	Eukaryote	Translational regulation
siRNA		21~25	Eukaryote	Protection against viral infection
piRNA		24~30	Eukaryote	Genome stabilization
Long ncRNA		several hundreds~ several hundred thousands	Eukaryote	Transcription, splicing, transport regulation

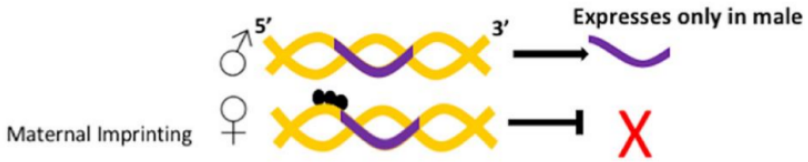
Reference Sequence	5	Used to annotate a mutation. NCBI. Intron is not coding sequence. Only extron. g.1346A>C | p.Arg54Gly | g.6666_6667delAT c.745delT | c.145+1T
Chromosome Structure Abnormalities Annotation	44	del(4)(p15.3) dup(4)(q31.1q35) inv(5)(q23q33) | inv(5)(q23p2) i(X)(q10) r(22)(p11.2q13.3) 46,XY,t(2;8)(q33;q24.1) 46,XY,der(14;21)(q10;q10),+21 45, XX, rob(13q, 15q)
Human Genome Project, HGP	50, 72, 82	1990-2003 Status: The world's largest collaborative biological project Sequencing the human genome → build reference sequence Mission: To understand the human genome. To understand the role human genome plays in both health and disease. Aim: 1. Identify sequence of bases on all 23 human chromosomes. (Completion of a high quality version of the human sequence) 2. Identify genes on the chromosome sequences. 3. Locate the position of the genes on the chromosomes. 4. Mapping and sequencing of a set of five model organisms (e.g., yeast, worm, mice, etc). 5. Create physical and genetic maps of the human genome Method:

		Shotgun sequencing, Sander method People: **International Human Genome Sequencing Consortium** • Sequencing machines run 24/7 • Many tasks performed by robots Two separate efforts--one public, one private--were published back-to-back in Nature and Science in 2001 (over 3GB → more than 500x the size of the bacteriophage). HGP 的开始标志着 pre-genomic era 的结束和 genomic era 的开始。2012 年开始，进入 post-genomic era
1000 Genome Project Consortium	5	2008-2010 Focus on the variation in human genome.
ENCODE Project	52	Identifying all functional elements in the human genome sequence
Germ Cells & Somatic Cells	3, 6, 12, 37	Genetic Variation can happen in germ cells and somatic cells. DNA damage can happen in germ cells and somatic cells Mosaicism: A is a germline mosaicism 生殖细胞的一部分: the mutations occurs during A's own germ cell' development and may involve multiple sperm or egg cells. The % to be affected is not sure. 即使个体本身表型正常（因为体细胞未受影响），但仍然可能将突变传递给后代。父母表型正常，但多次生育出患有同一遗传病的子女时，可能是**生殖系嵌合**导致的。 A is a somatic mosaicism 体细胞的一部分: the mutation occurs during A's own early embryo's development, so the offspring has multiple cells with or without the mutation. The mutation is also possible to be present in the germline. 体细胞嵌合可能导致局部或全身的异常表现。 Germline mutation 某几个生殖细胞: 发生在精子或卵子或这些细胞的前体细胞中，这些突变会传递给下一代。 Somatic mutation 某几个体细胞: 发生在体细胞中，这些突变不会传递给下一代。
Chromosome Rearrangement	8	Mismatch 不等于 Chromosome Recombination
Repeated sequences = Non-coding sequence	51	at least 50% of the human genome 有些 non-coding 是 unique, 不重复的 Significance: chromosome structure and dynamics
Tandem repeats	51	copies which lie adjacent to each other, either directly or inverted
Satellite DNA	51	HIGHLY repetitive sequence Typically found in centromeres, telomeres, sub-telomeres, heterochromatin Minisatellite DNA: 10-60 bp; found in many places in the genome

		including the centromeres Microsatellite DNA: less than 10 bp; includes telomeres
Chromosome-level Repeats	42	***Chromosome Architecture*** ***Centromere*** = alpha-satellite, is a highy repeated DNA sequence 171bp ***Telomere***, is a repeated DNA sequence 10-15kb (GGGTTA) ***Sub-telomere***, is a repeated DNA sequence
Low-copy Repeats (LCRs)	8	= Segmental Duplications 5% of the human genome consists of SDs In humans, chromosome Y and 22 have the greatest proportion of SDs: 50.4% and 11.9% respectively ***Definition:*** DNA segments with near-identical sequence. (可以是两段) The reason for recurrent mutations. Mismatch with LCRs within the same gene or the gene on the homologous DNA. Have a role in reshaping or rearranging the genome to create entirely new genes 属于 CNV
Transposon **= Interspersed Repeats** **= Transposon-derived repeats**	42	= 转座元件 能在基因组中自由移动和复制 都是本身就含有 repeats Short DNA segments that are scattered throughout the genome account for the chromosome banding pattern obtained with stains G-banding: Giemsa Staining Band ***Short interspersed elements (SINEs) 13% -逆转座子*** • GC rich, G-light band (less condensed) • Gene-rich • Rich in Alu repeated sequences  **Use LINE reverse transcriptase to copy their RNA to DNA** ***Long interspersed elements (LINEs) 21% -逆转座子*** • AT rich, G-dark band (more condensed) • Gene-poor • Rich in L1 repeated sequence

		 Encodes transcriptase, which copies their RNA to DNA ***Long terminal repeat (LTR) retrotransposons 8 % -逆转座子***  Reverse transcription occurs in cytoplasm, with the DNA copy then being transported to the nucleus ***DNA transposons 8% -转座子***  Encodes transposase which cut the transposon out of the genome
Copy Number Variants (CNVs)	37, 55	Example: CAG triple repeats in HD Feature: related to INDELs but consist of CN of larger segments of the genome, ranging from 1000 bp to mega bp. Effect: The genomic content of any two people can differ by as much as 50-100 Mb (~1% of the typical human genome size) because of CNVs. Counted as individual loci, SNP loci are the most common in frequency. Counted as base pairs, CNV loci contribute actually greater to the genome.
Polymorphism = Genetic Polymorphism	8, 49	Unformal definition: the loci that exhibits a *high frequency* of variation from individual to individual. Formal definition: the loci with >= 2 alleles where at least one minor allele has a frequency of at least 1%. 本质（of varied loci): alleles that have different DNA sequences. Including… 1. SNP polymorphism 2. Restriction fragment length polymorphisms (RFLPs) Changes in DNA fragment length with or without cleavage sites in it; Detected by restriction enzymes.  3. Simple-sequence repeats (SSRs) A genetic polymorphism resulting from short DNA sequence's copy number difference. (CNVs)
Balanced Polymorphism	58	= Heterozygote Advantage
Mendel's Laws = Law of	8	***Apply in:*** single gene disease, obvious dominance and recessive (complex diseases, mitochondrial inheritance, co-dominance, non-throughout

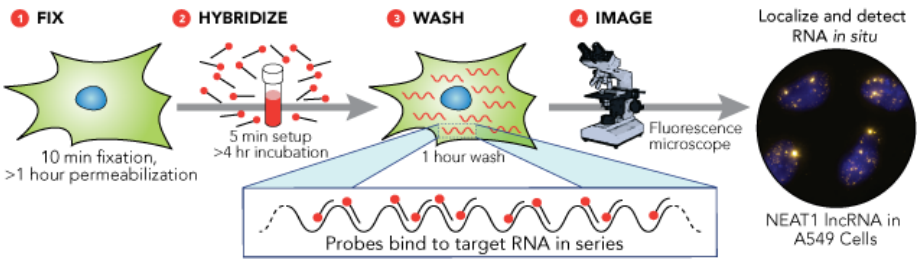
Mendelian Inheritance = Mendel's Law of Inheritance		penetrance…) ***Content*** First law: Law of gene segregation Second law: Law of independent assortment Another law: Law of dominance. ***Some Remaining Concerns with Respect to Mendel's law:*** 1. **Are Mendel's data too good to be true?** 孟德尔的数据过于接近理论预期的比例（如 3:1 或 9:3:3:1），可能存在数据修正或选择性报告的可能性。然而，这并不意味着孟德尔的实验本身是伪造的，而可能是统计上的偏差。 2. **Is Mendel's description of his experiments fictitious 虚拟的?** 孟德尔实验的一些细节被认为过于简化或缺乏具体信息（如环境条件、重复实验次数等）。 3. **Did Mendel detect but not mention linkage?** 孟德尔可能已经观察到基因 linkage，但由于他选择研究的性状大多位于不同的染色体上或距离较远，所以未表现出明显的连锁。 4. **Did Mendel support or oppose Darwin?** 孟德尔在他的文献中未直接提及达尔文，也没有明确支持或反对达尔文的理论。然而，两者的研究是相关的。达尔文强调自然选择在进化中的作用，而孟德尔则提出了遗传变异的规律。后来的研究将两者的理论结合起来形成了 Modern Synthesis。 
Penetrance = Expressivity	12, 38	The % of individuals with a particular genotype that express an associated phenotype. Penetrance means the proportion of individuals carrying a particular allele of a gene that also have the associated phenotype Determinants are some age-related cumulative factors: Environmental modifiers Genetic modifiers Epigenetics "Incomplete penetrance" means that if A has cytotoxic mutation, it will not absolutely expressed.
Genomic	13, 45	An epigenetic phenomenon: some genes are expressed only from the

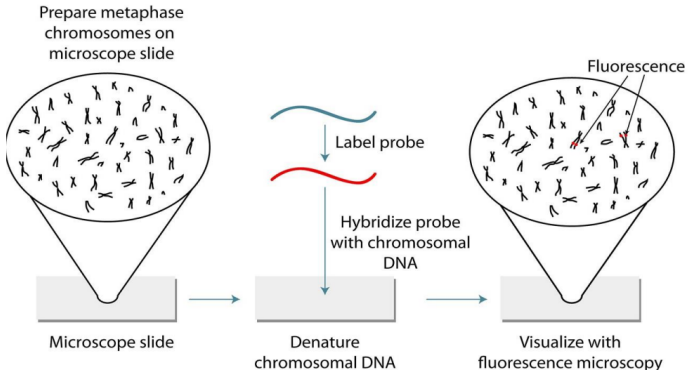
Imprinting		maternal or paternal copy. Expression: "Male or female undergoes imprinting center deletion"  Will lead to Angelman Syndrome and Prader-Willi Syndrome.
Multifactorial Disorder = Complex Diseases	31, 37, 64	A type of genetic disease. Cancer is a genetic disease. Caused by a combination of environmental factors (triggers) and mutations in multiple genes. Week4: Many variants of small effect contribute to disease risk, along with many environmental factors. Week13: Risk and progression of these diseases can have an inherited genetic component but is also influenced by environmental factors Inclusion: 1. Infectious (not 100% environment) 2. Cancer 3. Metabolic (obesity, diabetes) 4. Cardiovascular (hypertension) 5. Respiratory (asthma) 6. Psychiatric (depression, 精神分裂) Run in family, but do not have a clear-cut pattern of inheritance. Multifactorial in origin, complex in inheritance pattern.
Genetic Disorder = Genetic Disease = Inherited Disease	31	A disease caused by abnormality in DNA: single-gene disorder, mitochondrial diseases, chromosome abnormality, COMPLEX DISEASE autosomal recessive disorder is the most common genetic disorders.
Genotyping	34	Genotype: The set of alleles possessed by an individual at one locus; the genetic composition of an individual at one or many loci. 检测 DNA 中特定 loci 的多态性，借此确定一个对象的基因型
Finding the gene	61	How to link genetic variation to phenotypic variation? Linkage Analysis + Genetic Association Study
Genetic Association Study	34	Find candidate genes/genome regions that contributes to a complex disease. Case-control study GWAS Week7:

		Genetic association studies (e.g., association mapping, QTL mapping and tag SNPs) Genome-wide association studies (GWAS).

GWAS = Genome-wide association studies	35, 37, 60，61，62, 75	**Background:** Many common diseases have genetic components as well as quantitative phenotypic traits.
		
		In One Word: Connect genotypic variation with phenotypic variation.
		
		Feature: make every gene in genome a candidate gene
		
		Aim: go through genetic variants across the entire genome, looking for any that associates with disease. To answer the question "Do genetic variants occur more or less commonly in people with a disease than people without the disease?".
		
		Significance:
		1. Help to predict the likelihood of developing a particular disease
		2. The identification of risk SNPs can help with the development of treatments.
		3. The ultimate goal of GWAS is to determine factors that can be used to make predictions about who is at risk for a disease, what are the biological underpinnings of disease susceptibility and creating new prevention and treatment strategies.
		
		Focus: Typically focus on SNPs. That is, to identifying SNPs associated with particular diseases. Those being confirmed are called risk SNPs.
		Week7: By analyzing SNPs on a whole-genome scale to identify genetic associations with clinical conditions or phenotypic traits.
		
		Style: no prior assumption about candidates: non-candidate driven, phenotype-first.
		
		History: The first successful GWAS was published in 2005.
		
		Prerequisite:
		1. Completion of HGP.
		2. Fast DNA sequencing with low cost.
		3. Biobanks, for enough sample.
		4. Identification of “tag SNP”, to reduce the number of SNPs to be tested.
		5. Assumption that common variants present in 1–5% of the population account for the genetic contribution to common disease.
		
		Steps:
		`For each Single-nucleotide polymorphism (SNP)`
		`  Compute frequency on cases and controls`
		`  Compute odds ratio`
		`  Calculate the p-value of the odd ratio`
		
		Limitation: the SNP associations identified so far have explained only a small fraction of the genetic contribution to most common disorders.

		1. The forementioned assumption may be false. 2. Only SNP were studied. 3. Inconsistency due to false positive, false negative, population differences **Be Critical:** 1. Initial skepticism is warranted 2. Replication, especially with low p values, is encouraged 3. Large sample sizes are crucial 4. Replication (additional independent samples) is necessary! 5. Discovery vs. replication samples: discovery samples need to be corrected for multiple testing! 6. For GWAS with ~1M SNPs, it may need a p-value < 10^{-7}! 7. At the same time, effect size is typically very small. 8. Hence, a large sample size or meta-analyses across samples may be needed. **Rare diseases:** 1. Increase sequencing depth while controls must be re-sequenced (panel sequencing = targeted sequencing → find rare variants within disease-related genes but no accidental findings) with equal vigor! 2. Rare variants must be grouped for analysis BEFORE knowing the association study results. **Future efforts:** 1. Assay larger amounts of genetic variation through resequencing. 2. Include rare and complex variations. 3. Increase study size, investigate additional ancestry groups 4. Integrate with additional physiological data.
Polygenetic score = polygenic risk score	35,40	A number that condenses information from a person's multiple genetic variants into a score that measures the individual's genetic predisposition to specific diseases or complex traits. The calculation uses all these variants to arrive at a polygenic score. for polygenic diseases or MI with a polygenetic contribution part. 23 and Me's Type II Diabetes report calculates the combined effects of >1000 of these variants Also: Breast cancer
Cell Cycle Control System	40, 67	Composition: 3 transitions "trigger" 1. Late G1 start transition (G1/S Checkpoint): G1/S-Cdk & G1/S Cyclin complex is environment good? 决定是否进入细胞周期 Week 8: TP53, ATM

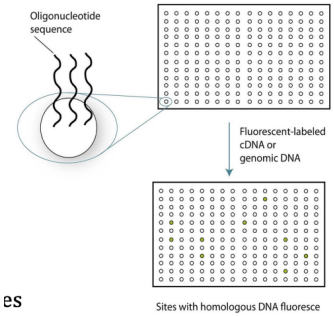
		2. S Phase Checkpoint: Week 8: epair damaged DNA 3. G2/M transition: S-Cdk & S Cyclin complex DNA replication ok? Environment good? 决定是否可以进入分裂期 Week 8: repair damaged DNA 4. Meta-to-ana transition: M-Cdk & M Cyclin complex 染色体粘到纺锤丝上了吗？ 决定是否可以进入后期
Gene Copy	41	IN NON-DERMCELLS un-dividing cells: 2c dividing (after S phase before cytokinesis): 4c
"n"	41	n = complete sets of chromosome "倍性" | n = "ploidity" 除非是像 meiosis I 的**同源染色体分离** 不然都是不会变的
Karyotyping	42, 87	高中的洋葱秋水仙素实验 Phytohemagglutinin (PHA as a [mitogen]) → Colchicine (mitotic inhibitor) → hypotonic solution → methanol + acetic acid Karyotyping = find out the karyotype Genotyping = find out the genotype Karyotype 不等于 genotype Karyotyping is 1. identification of karyotype, 2. pairing and 3. ordering all the chromosomes, thus providing a genome-wide snapshot of an individual's total chromosomes. Karyotypes are prepared using standardized staining procedures that reveal characteristic structural features (G-banding pattern) for each chromosome, which can be used to detect chromosomal abnormalities.
FISH **= Fluorescence in situ Hybridization** **= in situ Hybridization**	42, 80	**A common method for gene expression studies in pre-genomic era.** Target: intracellular RNA Walk-through  1. Detect chromosomal abnormalities at designated locations Resolution higher than G-banding but still low: in the megabase range



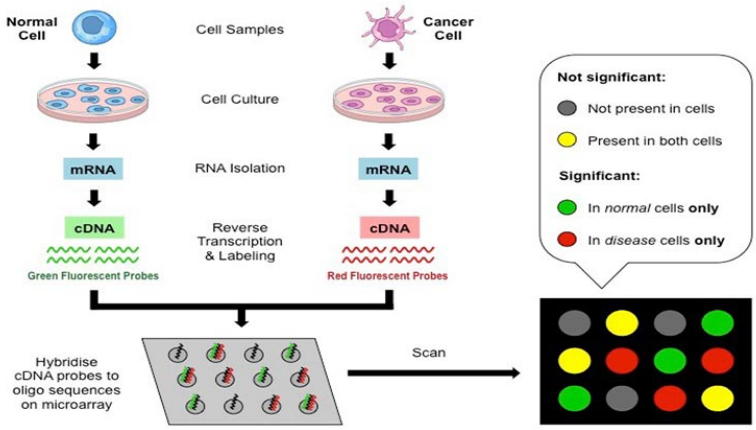
For example: detect sub-telomere region of chromosome 1's short arm to see whether it is deleted in disease condition

2. Chromosome Painting
Identify marker chromosome (structurally abnormal, unidentified EXTRA piece of chromosomal material)
For example: Use red, green and blue probe to detect Chromosome 1,2,3 in the genome. And see that there are 3 dots of green fluoresce, which in abnormal.

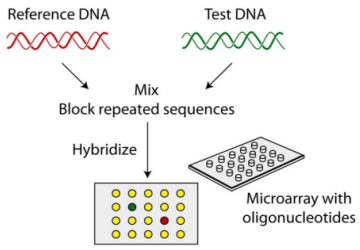
Microarray 43, 52 For: simultaneously detect the expression of numerous genes from a sample

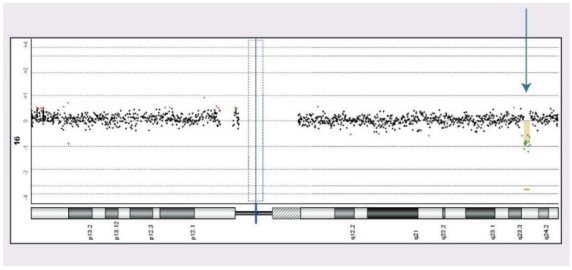


es

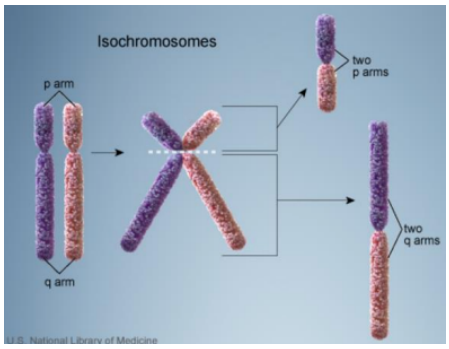
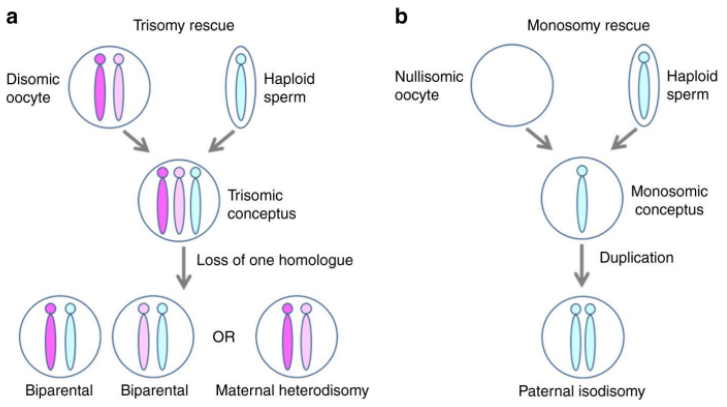


An application of microarray is in Comparatives Genomic hybridization, a sub-microscope molecular cytogenetic technology. (Array CGH)





Deletion of a small region on the longer arm of chromosome 16

Chromosome Ideogram	43	region 7p1 → band 7p13 → sub-band 7p13.2 deletion 46, XX, del(7)(p13)
Isochromosome	44	Duplications of either short or long arms. Produced by mis-division of the centromere 
Robertsonian Translocation	45	Exchange between two acrocentric chromosomes The two long arms join together, and the two short arms are lost 45 chromosomes, but carriers are usually normal as well. Reason: the short arms of all five pairs of acrocentric chromosomes (13,14,15,21,22) contain similar ribosomal RNA genes The loss of two of the ten is not critical
Uniparental Disomy	45	胚胎通过“拯救”过程避免染色体数目异常，但可能导致两条染色体 都来自单一亲本。 Result (bad): heterodisomy, paternal idodisomy 
Genetic Testing = Genetic Screening	45, 87	Genetic testing is a type of medical test that identifies changes in chromosomes, DNA/RNA, proteins and metabolites in order to detect heritable disease-related genotypes, mutations, phenotypes, or

karyotypes.

Methods Include:
1. Chromosomal genetic tests = Karyotyping: whole chromosomes or long lengths of DNA
2. Molecular genetic tests = gene tests: study single genes or short lengths of DNA
3. Biochemical genetic tests: amount or activity intensity of proteins;

Types Include:
1. Carrier Screening: if he/she is a carrier for a specific autosomal recessive diseases. Ethnicity/ancestry-based carrier screening → expanded carrier screening.
Molecular genetic tests + Biochemical genetic tests

2. Prenatal Screening: performed during pregnancy to determine whether a baby is likely to have specific birth defects.
Chromosomal genetic tests + Molecular genetic tests + Biochemical genetic tests

3. Newborn Screening: testing all babies in their first days of life for certain disorders that can hinder their normal development.
Heel Stick + Hearing Screen + Congenital Heart Screen

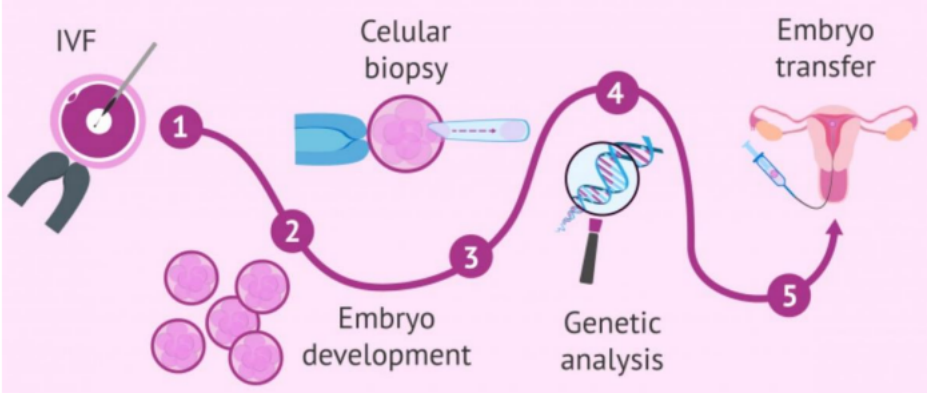
1st screen (regardless of feeding): 18 ~ 48 h
2nd screen: 7 ~ 14 days
3rd screen (sick and premature): 1 month

Blood cell disorders
Inborn errors of amino acid metabolism
Inborn errors of organic acid metabolism
Inborn errors of fatty acid metabolism
Miscellaneous multisystem diseases

出生前检测指标 (Indication):
• Advanced maternal age (>35 years)
• Recurrent pregnancy Loss or stillbirth
• Family history of chromosomal disorders
• Abnormalities in fetal anatomy detected by ultrasound

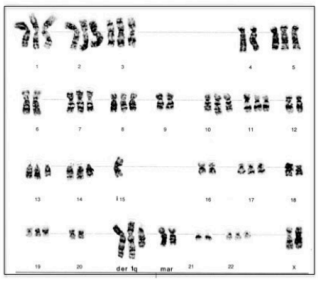
出生后检测指标 (Indication):
• Multiple congenital anomalies
• Unexplained severe development delay or intellectual disability
• Unexplained infertility

Future:
1. NGS
2. From Amniotic Fluid testing to Non-invasive Prenatal Testing
3. IVF's Preimplantation Genetic Diagnosis/Testing

		
Morgan's Recombination Law	59	Two loci from the same chromosome could be linked or separated.
Genetic Linkage	47, 48	***Discovery:*** AaBb x aabb 在这种情况下，会发生交叉互换和不会发生交叉互换在后代中形成的比例是不同的，并且是否会交叉是由基因连锁的因素决定的。 ***Object:*** 基因组合 ***Definition：*** 在同一条染色体上的基因 (A和B)在减数分裂中倾向于一起传递而不是独立分离的现象。这个理论的提出是创新的，毕竟根据孟德尔A和B应该表现出来独立分离(second law). ***Relevant Concepts*** **1. Genetic Distance (Genetic Map Distance/Map Distance)**: = the average number of crossovers between the two loci = RF = RF x 100% = map units aka cM Physical distance NOT always correlated with map distance. Because very little recombination takes place in HETEROchromatin; a small distance in the genetic map in HETEROchromatin corresponds to a large physical distance on the chromosome. A common notation for RF is θ, where θ varies from 0 (no recombination at all) to 0.5 (independent assortment). If two loci are so close together that θ = 0 between them, they must be completely linked. If they are so far apart that θ = 0.5, they are expected to be assorting independently and are unlinked. **2. Genetic Map**: A map constructed with Genetic Distance ***Perquisites for***

		A和B独立分离/A和B符合孟德尔第二定律： 1. A和B在不同的染色体上 or 2. A和B在同一条染色体上，但是距离很远 ***Measurement:*** $$\text{Recombination frequency (RF)} = \frac{\text{Recombinants}}{\text{Total offspring}} \times 100\%$$ 1 % = 1 map unit → 1 cM (centiMorgans) In other words, 1 cM = 1% chance of recombination between markers. In other words, RF between a pair of genes is a function of distance. 取值范围: [0, 50%] ***Inspire:*** Gene Mapping Method (**Linkage Analysis**) → Gene Map (发明了在染色体上定位不同基因位置的方法) Linkage Analysis depends on RF as a measure.
Linkage Analysis	48	***Definition:*** A method of mapping genes using the RF between loci as a measure of genetic distance. ***Framework:*** 1. Linkage: Narrow search to a small chromosomal region - Affected relatives co-inherit markers in a region more often than by chance - Monogenic disorders: successful - Multigenic disorders: less successful 2. Association: Choose and test common variants in genes - Candidate genes - Well-suited to common alleles of modest penetrance or… 2. Association: Find and test rare variants in genes - Candidate genes - Resequencing to find rare variants - Very expensive ***Feature:*** Candidate gene approach. 从表型出发，如果在 framework 中的 1 出发时就遗漏了真正的“钥匙”，那么后面时全白费的。 ***In human*** **Aim:** To find the chromosomal segments that co-segregate with the ailment 轻病 phenotype **Steps:**

		1. Establish a pedigree 2. Make several estimates of RF **No recombination** **Tight linkage between a pair of loci** **50% recombination** **An unlinked pair of loci** **10% recombination** **A linked pair of loci** 3. Calculate a LOD score for each estimate 4. The estimate with the highest LOD score will be considered the best estimate A variety of DNA sequence polymorphisms can be used as markers for linkage analysis. 当研究者试图LOCATE未知的致病基因时，可以通过marker来划定基因的染色体区域。如果一个标记和marker紧密连锁，它们在遗传时经常一起传递。
Linkage Equilibrium	59, 60	当两个遗传标记在群体中是完全独立的，它们的组合频率等于各自的单独频率的乘积时，称它们处于连锁平衡。
Linkage Disequilibrium (LD)	59, 60	Definition: 当两个遗传标记之间的组合频率偏离了独立分布的期望频率（即不满足连锁平衡条件）时，称它们处于连锁不平衡。 Linkage disequilibrium is the nonrandom association of alleles at closely linked loci in a given population. Quantitative Definition: An allele A with frequency p(A). At a closely linked locus, the frequency of the B allele is p(B). The observed frequency of the haplotype AB is p(AB). If the association between A and B is completely random and statistically independent, p(A)*p(B) = p(AB) LD between A and B if p(AB) differs from p(A)*p(B). Influential Factors: Genetic linkage and recombination rate, Mutation, Gene flow, Selection, Genetic drift, Inbreeding. Why matters? LD enables genetic association studies (e.g., association mapping, QTL mapping and tag SNPs) and genome-wide association studies (GWAS).
Genetic Markers	49	A gene or DNA sequence with **a known location** on a chromosome that can be used to identify **individuals or species and** study the relationship between a **disease** and its genetic **cause**. It can be those repeated sequence. May be short: a sequence surrounding a SNP. May be long: minisatellites.

Chromosome Instability	64	= ploidity change (2) + Telomerase shortening + chromothripsis Cancer Cell  Normal Cell  For examples: Deletions; Duplications; Translocations; Inversions; etc….