لجنة Dentin 2020

Genetics

Writer : Tala Shehadeh

doctor : Dr.Osama Smadi

corrector: Type here Heba Albajjai

Chromosomes, Genes & DNA

- Somatic cells are diploid - 46 chromosomes,22 pairs autosomes; 1 pair sex chromosomes.
- Each pair of autosomes is homologous, Contains the same genes in the same order. 1 is maternal, the other is paternal.
- Chromosome are composed of deoxyribonucleic acid (DNA),Genome contains 3 billion base pairs (haploid), ~1% encode proteins.
- Genes are located on chromosomes

Genomic Advancements

- The last 10-20 years has seen advances of 'genomic technologies'.
- Thanks to the production of new sequencing technologies ,such as next generation sequencing, we are able to sequence the entire gene in just under one day.

❖ Next Generation Sequencing (NGS)

- Whole Genomes (sequencing the entire gene).
- Whole Exomes (WES) .(sequencing a part of the gene ,sequence only the exons of the coding genes).
- Targeted Gene panel.(sequence part of the genes depending on the needs of the individual who is looking at the gene from a disease point of view or a population point of view).

- **Whole Genome Sequencing (WGS)**
  - Decrease in size of technology.
  - Improvement in IT and bioinformatics.

  With this new technology there is a need for bioinformatics technologies which enable the analysis of the large data size that anyone can get from a single gene and enables us to look at the sequence base by base ,gene by gene, and this enables us to read the genome and understand the content of it with respect to the genes and their variants ,their location and the differences between each individual and another.
  - Decrease in genomic technology costs.

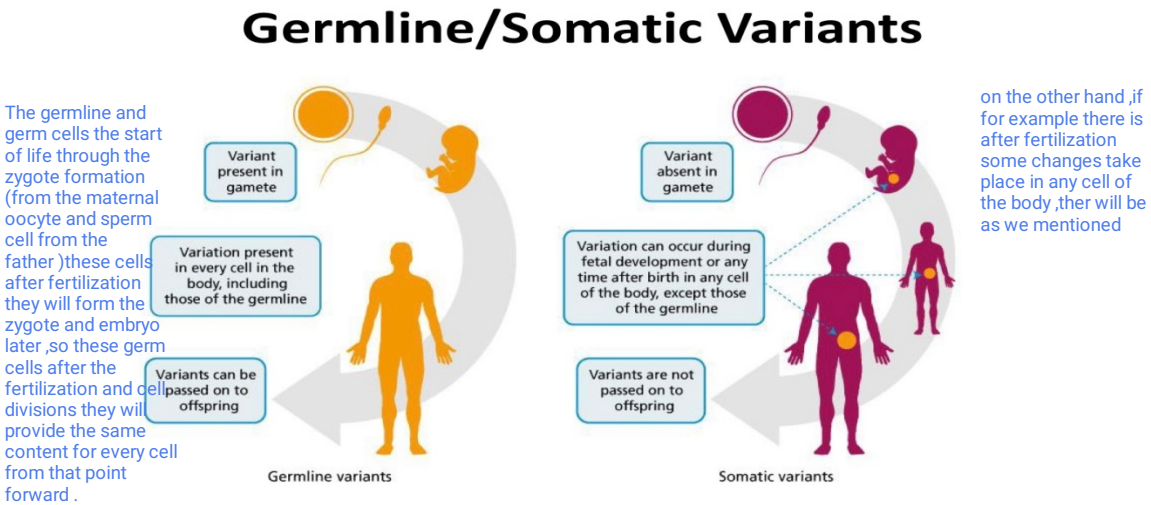
- ✓ What is Genetics? Genetics refers to the study of individual genes and their roles in inheritance. for health and disease purposes.
- ✓ What is Genomics? Genomics refers to the study of the entire genes, their interactions and functions.

Your genome is one whole set of all your genes plus all the DNA between your genes.

How many genes do we have? The exact number is still in fact a debate , so the last words there is between 20000 and 22000 genes that scientists believe there are in human genome.

Germline Meanings:

- In biology and genetics, the germline is the sequence of germ cells with their genetic material that get passed to the offsprings.
- Reproductive cells (sperm or the egg), are part of the germline. (they are the origin of life)
- Cells that are not in the germline are called somatic cells. Example cells of the liver/skin.
- If there is a mutation in the germline, it will be/can be passed to offspring, however changes in a somatic cell it won't.(the change will only stay in that cells and the cells resulted from its division ,for example ,it's a cancer cell with a mutation , this will only stay within that cell and the cells that resulted from it.



The germline and germ cells the start of life through the zygote formation (from the maternal oocyte and sperm cell from the father)these cells after fertilization they will form the zygote and embryo later ,so these germ cells after the fertilization and cell divisions they will provide the same content for every cell from that point forward .

on the other hand ,if for example there is after fertilization some changes take place in any cell of the body ,ther will be as we mentioned

Genetic diseases Classification

Three Groups:

1. Multifactorial inheritance (polygenic)
2. Monogenic (Mendelian)
3. Chromosomal aberrations

Disorders with multifactorial inheritance (polygenic).

(this type of diseases are common diseases and are influenced by multiple gens and contribution to advancement for this diseases by environmental factors)

- influence of multiple genes plus interplay with environmental factors. (contributes to the development of the disease for example diet , physical activity, smoking, ETC….)
- relatively frequent (population disorder).
- Examples:
  - Diabetes mellitus
  - Hypertension
  - Certain congenital heart diseases
  - Some types of cancer (ovarian, breast, colon)

- Can run in families-(sometimes there is a heritable component to these diseases despite the multifactorial and multigenic factors).

1st degree relatives about 5-10%;

2nd degree relatives 0.5-1%

Monogenic (Mendelian) Disorders (diseases resulting due to defect in single gene).

- Mutation of one gene: Mendelian inheritance,There are more than 5,000 diseases. (they are heritable diseases)
- Three types according to their way of inheritance:

i. Autosomal dominant
ii. Autosomal recessive
iii. X-linked

Autosomal Dominant Disorders

One affected gene is enough foe the phenotype to appear.

- Both Homozygotes And Heterozygotes Are Affected (shows the phenotype).
- Usually Heterozygotes (Inherited from one parent)
- Males And Females are Affected
- Transmitted From One Generation To The Other at 50% rate

Autosomal recessive

Two affected genes must be for the phenotype to appear.

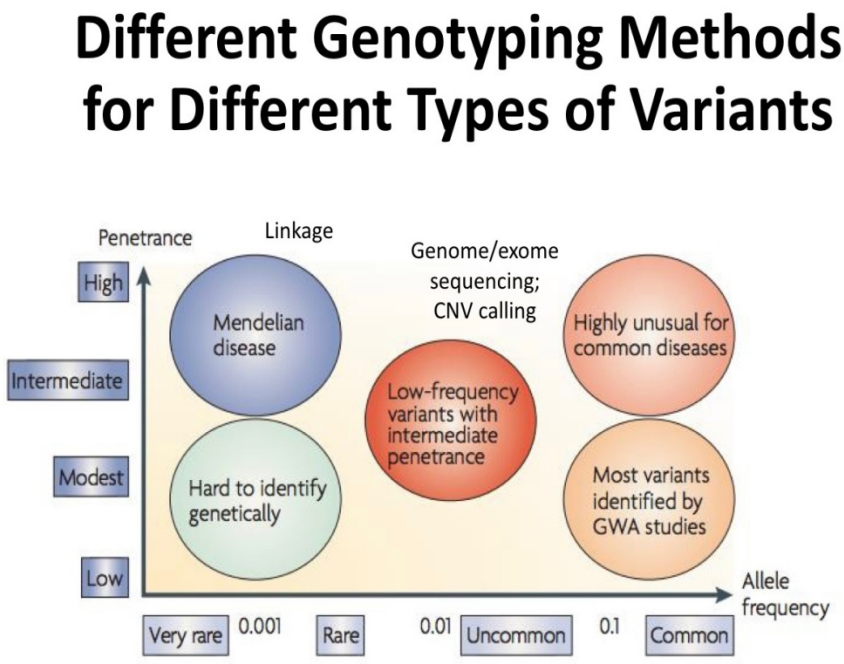
- The majority of Mendelian disorders.
- Homozygotes are affected, heterozygotes (parents are carriers)
- 25% of descendants can be affected
- **if the mutant gene occurs with low frequency - high probability in consanguineous marriages.(relatives marriage) the disease is more frequent in middle eastern countries like Jordan because there is higher rates of relative's marriage .** in Arab world the the consanguineousrate rate is 50% high compared to the western population or Asian population.
- Onset of symptoms often in childhood. Essentially from birth and in early life stages.
- Frequently enzymatic defect. (many of these diseases are associated with enzymatic disorders)

X-linked diseases

- Transmitted as heterozygous from mother to sons.
- Daughters: 50% carriers, 50% healthy.(XX, one from the mother and one from the father).
- Sons: 50% diseased & 50% healthy.(XY, the X from the mother and the Y from the father).
- Examples:
  - Hemophilia A (defect of Factor VIII).

- Hemophilia B (defect of Factor IX).
- Muscle dystrophy (Duchen disease).

Different Genotyping Methods for Different Types of Variants



here we are presenting some gene variants and how frequently they happen also with respect to their Penetrance or association with expression of disease and phenotype

- Diseases vary from being very rare (in general population meaning)at 0.001 percent to being very common at 0.1 percent. (the frequency of diseases can be very rare 0.001 percent but this percent can change and increasing to one hundred and the frequency can keep increasing depending on nature of variants)
- Depending on the nature of the variant they will have either a low penetrance or high penetrance(association with expression of the diseases)
- Penetrance :the extent to which a particular gene or set of genes is expressed in the phenotypes of individuals carrying it.(from google).
- Mendelian disorders usually associated with a very rare variance with high penetrance and these historically were found before we came up with the advanced genomic tools that we discussed, there were other tools such as linkage and

(like next next generation sequencing)

some gene maping and cloning experiences that helped discovering these variants and there associated genes. (leading to such rare diseases like Mendelian type)

- On the other hand , if we talked about multifactorial disorders they are impacted by several genes usually, or could be genes that interplay together along with environmental influence (variation gens) and they exist at higher frequency , and to look for these disorders and to understand them and to try to find out if which genes are associated with these multifactorial disorders , one need to look to the entire genomic content such as through the exome and whole genome sequencing .Through that type of technology ,there are not just the gene variants, there is also the copy number ,variations that are associated with such diseases and can be looked at through computational tools, and comparing the studied individuals with the disease with those without the disease and will enable us to discover the critical set of genes that play a role in such polygeneic multifactorial diseases.

Where does Variation come from?(how they happen)

so,these genetic changes we keep referring to,how they happen?

- Mutation
  - ⇒ random changes to DNA
    - errors in mitosis(after fertilization) & meiosis(before conceiving and fertilization, in germ cells) so we will have some errors that heppen during genomic DNA replication and these changes can accumulate and get passed through germ cells to the embryo later on .
    - environmental damage(for examplr reactive oxygen molecules , radiation, some other factors that influence the happening of these mutations)
- Reproduction is another source of genetic variation through:
  - ⇒ mixing of alleles through:
    - recombination of alleles
      new arrangements in every offspring
    - new combinations = new phenotypes =new diseases

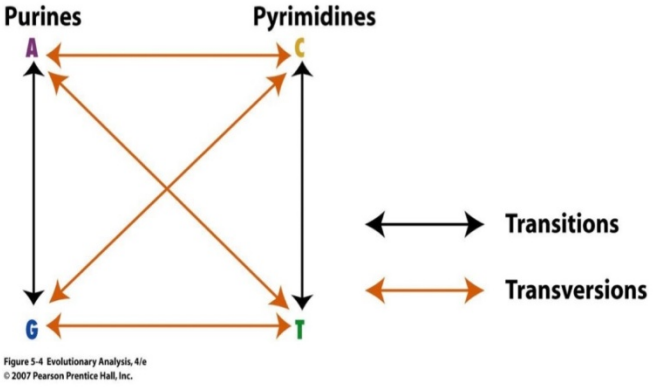
we subjected to these changes of DNA distribution and content and new combinations can lead to new phenotypes and some types of diseases.

⇒ spreads variation
- offspring inherit traits from parent

Significance of Mutations

- Most mutations are neutral – have little or no effect (benign ,non pathogenic)on the expression of genes or function of proteins, and these changes their impact or effect shown in the proteins that they will made.
- Harmful mutations – leads to defective proteins –disrupt normal biological functions. and can lead to deep seating diseases (أمراض مزمنة).and these mutations are associated with:
  - ⇒ Cause genetic disorders, for example sickle cell anemia.
  - ⇒ Associated with many types of cancer (somatic changes).

if the mutation happen in the somatic cells some changes can be very harmful and lead to major disease such as cancer that is resulting from somatic changes in somatic cells.



Transitions are more common than transversions because DNA repair enzymes can recognize wrong insertion representing a a transition better than a transversion

- When the changes are between purines purely or pyrimidines purely ,for example G to A or C to T they are called transitions , and when these changes are between purines and pyrimidianes or vice versa they are called transversionsfor example:A to C.

Gene Mutations create changes in DNA sequence, amino acid sequence, and changes in the encoded proteins

changes that we discussed they can be in different forms:

Point Mutations

- Mutations that affect one nucleotide are called point mutations because they occur at a single point in the DNA sequence.

Substitutions

⇒ substitute one nucleotide for another

⇒ change one of the amino acids in a protein, so there is a change in the resulting protein.

in the end because of that change the protein will be changed because of the changing the amino acid (most point mutations will result in defective proteins but sometimes it can be silent mutation so no changes in the end)

From Genes to Proteins via mRNA

- Proteins consist of polypeptide chains made of amino acids
- There are 20 amino acids and they are encoded by 61 different codons that are resulting from different combinations of the three base bairs from nucleotides that make codons. in addition to 61 combinations of codons there is 3 stop codons :they will lead to termination of translation.
- Genetic code

⇒ 64 combinations of 3 bases called codons

⇒ There are three stop codons:
(UAA, UGA, UAG)

- Genetic code is degenerate(means that one amino acid may be encoded by more than one of these codons).(the codons are considered degenerate)
- Genetic code is universal

there is another form of nucleotides changes we call them :

Frameshift Mutations

♦ Insertions or deletions

⇒ Shift the “reading frame” of genetic code.

⇒ May change every amino acid that follow the point of mutation altering the protein so it is unable to perform its normal function.

⇒ The translation will result in abnormal protein.

⇒ Bases are read in groups of 3 (each 3 neucleotides are called codon) so when one neuocleotide is deleted or inserted it will alter the reading of all the codons following , resulting in either stop codon or abnormal proteins that lacks the desired function.(all following codons will be read incorrectly).

there is two types :1-in frame insertion or deletion;there will be in the end complete translation of the mRNA with some changes in encoded amino acids after the point of insertion or deletion.
2- out of frame mutations :there will be changed in order of codons and will lead to termination of the translation and if the change happens early on the gen it will result in completly abnormal protein and can lead to disease .

the last change we will discuss is :

Chromosomal Mutations

1.Deletions

Involve loss of all or part of a chromosome

2. Duplications/Additions

Produce extra copies of parts of a chromosome

e.g. Fragile X syndrome(leades to mental retardation)

within the chromosome there will be insertion of certain nucleotides repeats ,for example fragile X syndrome: multiple trinucleotides repeats that get inserted over and over at the end there are hundreds of foreign base bairs in the gen so abnormal phenotype

3. Inversions

Reverse direction of parts of chromosomes

within the chromosome there will be cutting and rejoining of part of chromosome and the chromosome distribution will be totally disrupted

4. Translocations

When part of one chromosome breaks off and attaches to Another(can be associated with diseases such as leukemia and lymphoma and so on).

5. Non Disjunction –

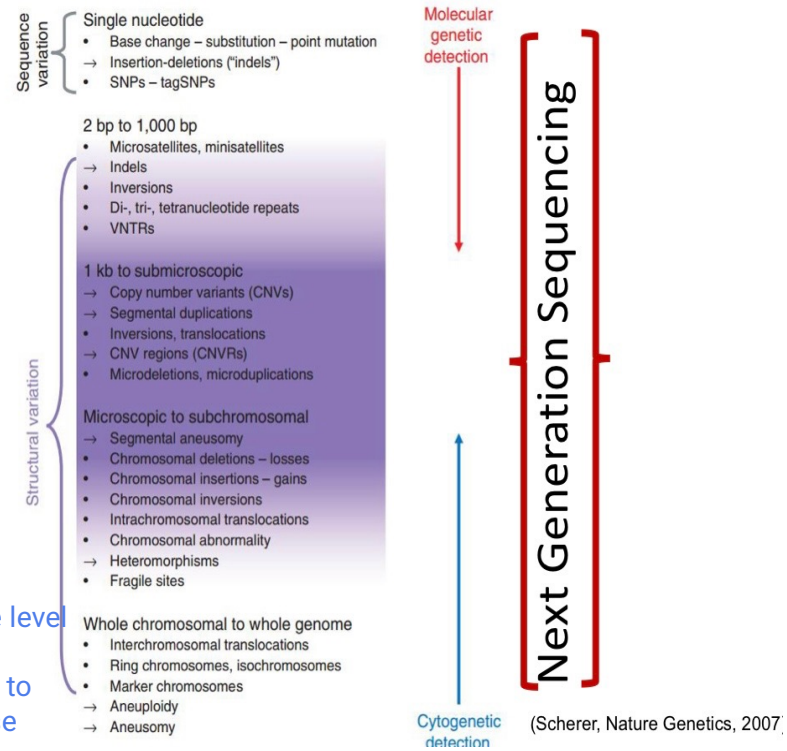
during meiosis, a pair of chromosomes do not separate & a gamete has one more chromosome while the other has one less

the changes that we discussed so far they also have names:

Human Genetic Variation

- Single nucleotide polymorphisms (SNPs)(point mutations)(the majority of changes)
- Tandem repeat Sequences
  - ⇒ Microsatellites (<8 bp)
  - ⇒ Minisatellites (VNTRs(variable neuceotide repeats); 8-100 bp)
- Copy number variants (CNVs; 1Kb – 1Mb) M= mega
- Insertions – deletions (indels; 100bp – 1Kb)
- These size limitations are arbitrary.

Genetic variation across size spectrum



(Scherer, Nature Genetics, 2007)

(whether it is gross changes such as the level of entire chromosome or simply single nucleotide change ,you will have access to different technologies that can find these variations.)

- When you study the changes , when you are looking at a single gene you can do real time PCR probe detection for the variance which simply are looking to distinguish between two allels.

And if it is at the entire chromosome level one will depend on more comprehensive genomic tools such as cytogenetics and next generation sequencing approaches and microarrays ,ETC…

So all of these tools enable us to look at the variations in the genetics of us as humans based on our needs and the scope of diseases as well.

SNPs (single nuecleotide polymorphisim)

If the single nuecleotide mutation occurs in more than 1% of the population it is called SNP, If it is in 1 or ess than 1% of the population it is called mutation.

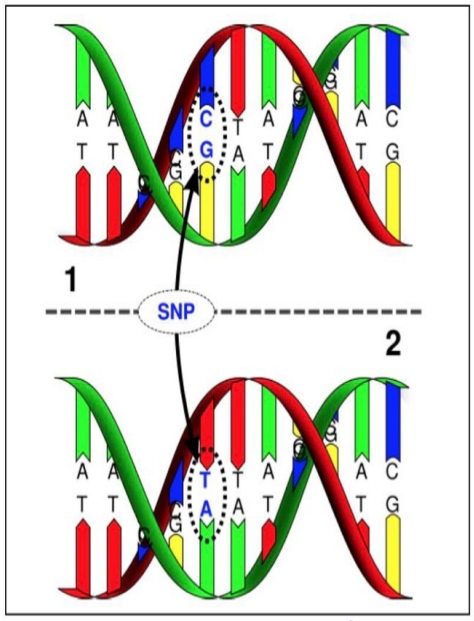
- 10 million SNPs in any given human genome & counting and they occur every 1000 to 3000 base pairs in average
- The most common type of genetic variation (change)

- 2 alleles; e.g., A → T
- Occurs across the entire genome & in stable regions(happen anywhere in genome)
can happen in slot region such what happen in cancer.
- Many SNPs are in linkage disequilibrium (LD)

⇒ SNPs close together are more likely to travel together in a block these SNPS sometimes are neighboring to each other meaning that there are adjacent to each other and if someone looking for disease for example diabetes multifactorial multigenic disorder when discover the SNPs that are associated with the disease .

⇒ Can use 1 'tag' SNP per block – cost effective for **GWAS**(genome wide association studies , a type of study that looks for SNPs)

⇒ We have the opportunity to focus on one of the representative SNPs out of this neighboring SNPs that when they are associated with each other exist in block and they are what we call linkage dis equilibrium, meaning they go hand in hand and if you know which SNP exists in this location you can assume that the rest of the surrounding regions also has a similar genetic distribution with respect to homozygosity and heterozygosity .And this kind of feature was used in the genome wide studies when they call the tag SNPs which they use in microarrays technology instead of looking to the 10 million SNPs one can only look at 10000 of these SNPs or up to 100000 that will be representative of the 10 million and also will cover the entire genome for the purpose of these disease studies.



C-allele: 70% frequency

C = major allele

T-allele: 30% frequency

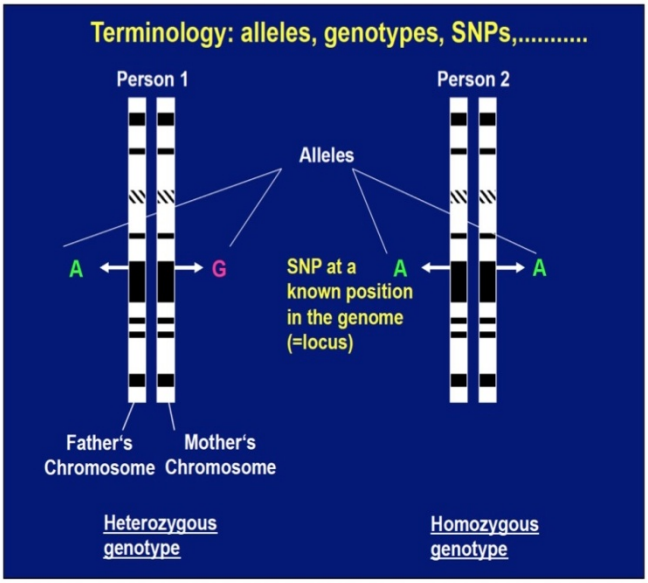
T = minor allele

the majority of population have the C complementary to G in the other strand in the locus and it is the normal healthy state ,if any changes occur in that location for example T instead of C as variant or mutation the second variant (T) called minor allele

after all advancement the announcement of human genome happened in early 2000 the scientists built the genome as whole and from this genome they be able to study healthy individual to map all the gens and read all the genome base by base ,and when assumed is healthy and they have for example one nucleotide see in one position in any of the genome in entire that will be a reference so refers to as major allele .

- Major allele means that the majority of the population will have this allele in their genome in this specific location.
- Minor allele means that the minority of the population will have this allele in their genome in this specific location.

if we look to major and minor alleles at population level not individual there will be a percentile,there is a certain SNP that a population have frequency of 85 % while type there will be the difference s50% will carry the variants.



If allele **G** is associated with risk for disease, it is the **risk allele.**

Allele A the **protective** allele (WT)

these variants represented in different sitting or just change the nucleotide selection and here wo looking to two different individuals,suppose the allele G associated with a risk of disease

Slide courtesy of Sven Cichon

- If the individual has the risk allele they will have the disease (it is the changed , variant allele)
- If the individual has the protective allele or the wild type ,this means they won't get the disease (it is the non changed form of the allele) it is the most common found in population.

you have to understand alleles in general ,they are either be protective when they WT form but when it change they called mutant or risk allele and these WT and risk alleles are terms that commonly used in genetic studies and compared individual to each other.

You can use the following databases to know the sequence of genetic material in order to study diseases.

when people study the genome,this was global and universal exercise that was contributed to by many scientists and researchers ,they were collected data from so many population that have been studied and put genetic variation as a bases ,here is examples listed .and what interested in looking more of human diseases and variations we can use these data bases as reference and better understanding for these genetic diseases

Genetic Variation Databases

Database	Content	Address
dbSNP	SNPs covering the human genome	http://www.ncbi.nlm.nih.gov/projects/SNPs
HapMap	Catalog of variants from HapMap Project	http://hapmap.org
1000 Genome Project	Extension of HapMap – aim to catalog 95% of variants with 1% freq	www.1000genomes.org
UCSC Genome Bioinformatics	Reference human genome sequence with annotation	http://genome.ucsc.edu
Ensembl	Genome browser, annotation, comparative genomics	http://www.ensembl.org/index.html

Genetic Variation Databases

Database	Content	Address
GeneCards	Database of human genes linked to relevant databases	http://www.genecards.org
PharmGKB	SNPs involved in drug metabolism	http://www.pharmgkb.org
DGV	Database of Genomic Variants, including CNV	http://projects.tcag.ca/variation
SCAN	SNP & CNV annotation based on gene function & expression	http://www.scandb.org/newinterface
OMIM	Online Mendelian Inheritance in Man – over 12,000 genes	http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim

Database	Content	Address
HuGE navigator	Human genome epidemiology knowledge base	http://hugenavigator.net/HuGENavigator/home.do

if you look at disease the key material to have is always DNA sample

Collecting DNA

- Sources of DNA
  - ⇒ Blood samples (most common).
  - ⇒ Buccal brushes
  - ⇒ Saliva samples it has million of cells that have DNA
  - ⇒ Dried blood spots ,usually used for newborn disorders.

at time of birth one can collect several drops of blood from the newborn and place it on card that absorb the blood

- Depends on
  - ⇒ Conditions at time of collection
  - ⇒ Resources available to process samples
  - ⇒ What other biological samples will be collected
  - ⇒ Long & short term storage
  - ⇒ Quality control

depends on need and genetic analysis

the DNA fortunately if it is stored correctly it can stored for long time up to tens of years and will be in good quality that can used in genetic studies

Saliva vs. Blood Samples

saliva as alternative for blood sample and today's used in common

Saliva is a practical sample ~~coection~~ collection option , considering:

i. Lower cost: compared to blood collection
ii. More convenient & acceptable to patients
iii. Increases compliance
iv. Lower mean yield of DNA with comparable Quality(it has lower quantity of DNA but same Quaity).
v. No difference in success from high throughput Genotyping.

transport easily from one location to another for storage and genetic analysis

if you have doing study and use saliva as source in fact you use non invasive method of sample collection as better better compliance and participation by volunteers.

the quantity of DNA compared to blood is less however the quality is essentially the same

وأن ليس للإنسان إلا ما سعى
اللهم صلِ وسلم وبارك على سيدنا محمد
💙