The human genome: applications and implications

1) What is the human genome? History of the human genome sequencing project, & where we are now

2) Where is the human genome and how is it annotated? UCSC browser and ‘tracks’

3) Genetic testing

4) Ethics and the genome

Readings:

See links shown in class for more info

Also:
1. *DNA and insurance*: debate on the ethics of DNA testing by insurers
2. *Gene testing and anonymity*: debate on the overall value vs. risk of widespread genetic testing
3. *GINA and genomic medicine*: What effect has the Genetic Information Non-disclosure Act had on genetic testing and medicine?
4. *Genomics legal 2019*: genomics advances bring new legal challenges
5. *Genome injustice 2019*: Underrepresented groups push for more representation in genomics advances

The human genome project: some milestones

1986: "Human Genome Initiative" begins at US DOE
1992: Complete, low resolution linkage map of genome
1995: First complete genome (*Hemophilus influenzae*)
1999: Human chromosome 22 finished
2000: President Clinton announces completion of 'working draft' of human genome
2003: Human genome project declared complete
2004: Human gene count estimate: 20,000-25,000, function of more than half is unknown
2006: Human chromosomes 1, 3, 8, 11, 12, 15, 17 completed
2010: '1000 genomes consortium' to map human genetic variation

https://web.ornl.gov/sci/techresources/Human_Genome/index.shtml

http://web.ornl.gov/sci/techresources/Human_Genome/project/journals.shtml

The world's largest collaborative biology project

So whose genome was sequenced initially?

- Two groups worked to complete the genome assembly.

- The **publicly funded** group used DNA from an anonymized **group** of donors (two male, two female) from Buffalo, NY (where the DNA preparer was based)

- The **privately funded** group (Celera) used DNA from **five individuals** from an anonymized pool.

- Subsequently: the “1000 Genomes Project” included donors from diverse populations, for details see http://www.internationalgenome.org/about

https://dnalc.cshl.edu/view/15327-The-public-Human-Genome-Project-s-DNA-donors-Eric-Lander.html

Accessing the human genome: the UCSC browser
http://genome.ucsc.edu/

Current version: human reference sequence GRCh38 (a.k.a. hg38) produced in December 2013, with updates frequently added. Most recent is GRCh38.p13, from 2/28/19, although UCSC still uses .p12 from 12/21/17

https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.39

User guide to UCSC browser:
http://genome.ucsc.edu/goldenPath/help/hgTracksHelp.html

Numerous helpful links there

Statistics of hg38.p13 (2/28/19)

Number of regions with alternate loci or patches	358
Total sequence length	3,099,706,404
Total ungapped length	2,948,583,725
Gaps between scaffolds	349
Number of scaffolds	472
Scaffold N50	67,794,873
Scaffold L50	16
Number of contigs	998
Contig N50	57,879,411
Contig L50	18
Total number of chromosomes and plasmids	24
Number of component sequences (WGS or clone)	35,613

https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.39

Statistics of hg38.p12 (12/21/17)

Number of regions with alternate loci or patches	317
Total sequence length	3,257,319,537
Total assembly gap length	161,368,351
Gaps between scaffolds	349
Number of scaffolds	874
Scaffold N50	59,364,414
Scaffold L50	17
Number of contigs	1,535
Contig N50	56,413,054
Contig L50	19
Total number of chromosomes and plasmids	25
Number of component sequences (WGS or clone)	37,479

(from https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.38)

Getting to the data

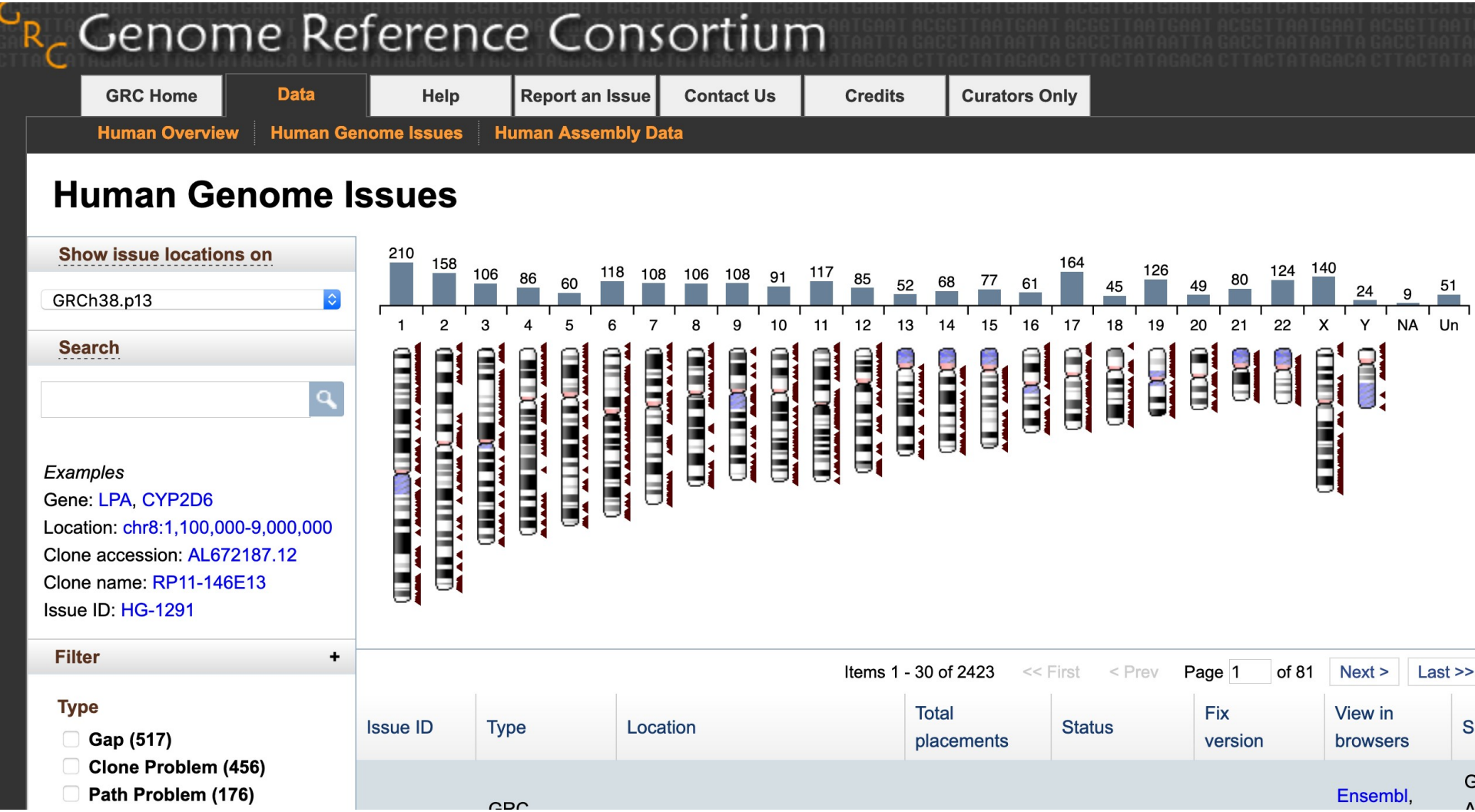
Chromosome lengths | Total lengths | Ungapped lengths | N50s | Gaps | Counts

Chromosome lengths are calculated by summing the length of the placed scaffolds and estimated gaps.

Chromosome	Total length (bp)	GenBank accession	RefSeq accession
1	248,956,422	CM000663.2	NC_000001.11
2	242,193,529	CM000664.2	NC_000002.12
3	198,295,559	CM000665.2	NC_000003.12
4	190,214,555	CM000666.2	NC_000004.12
5	181,538,259	CM000667.2	NC_000005.10
6	170,805,979	CM000668.2	NC_000006.12
7	159,345,973	CM000669.2	NC_000007.14
8	145,138,636	CM000670.2	NC_000008.11
9	138,394,717	CM000671.2	NC_000009.12

https://www.ncbi.nlm.nih.gov/grc/human/data

A few issues remain to be resolved:
https://www.ncbi.nlm.nih.gov/grc/human/issues



Human Genome Issues

Show issue locations on

GRCh38.p13

Search

Examples

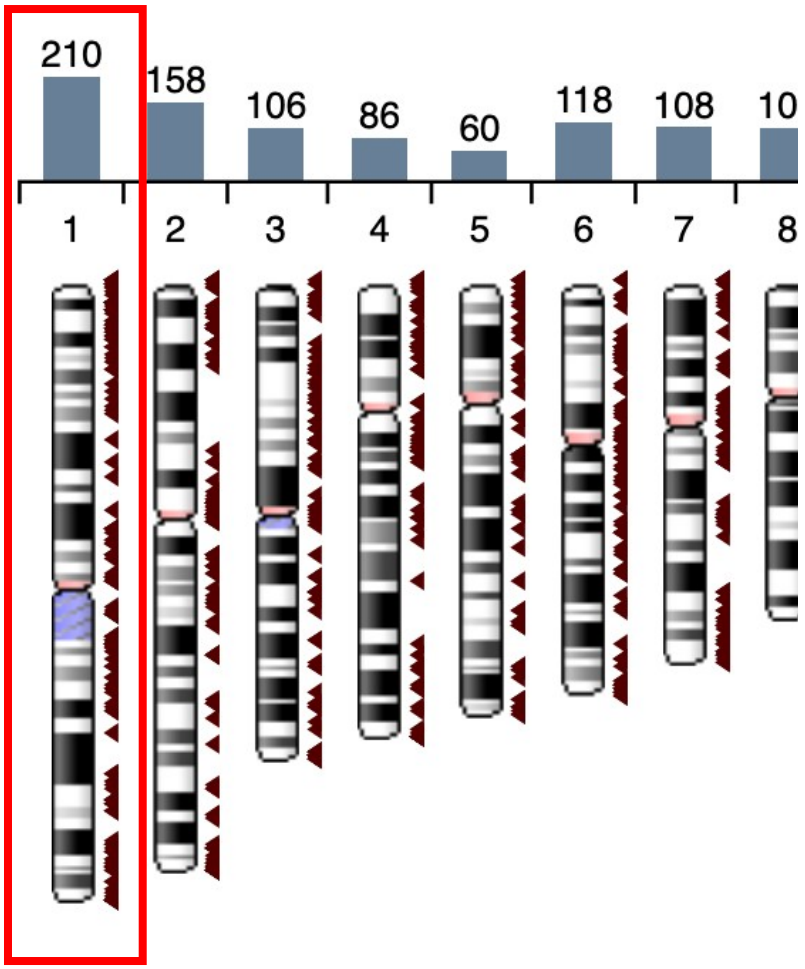
Gene: LPA, CYP2D6

Location: chr8:1,100,000-9,000,000

Clone accession: AL672187.12

Clone name: RP11-146E13

Issue ID: HG-1291



Click on one of the chromosomes

Human Genome Issues

Show issue locations on

GRCh38.p13

Search

Examples

Gene: LPA, CYP2D6

Location: chr8:1,100,000-9,000,000

Clone accession: AL672187.12

Clone name: RP11-146E13

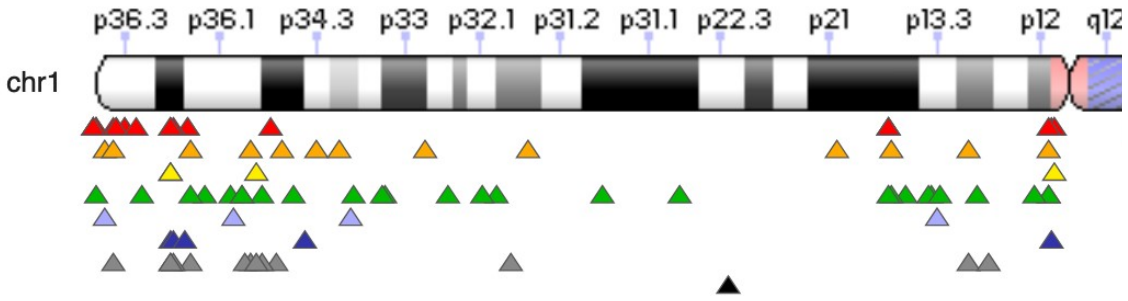
Issue ID: HG-1291

Filter +

Type

- **Gap (70)**
- **Clone Problem (24)**
- **Path Problem (14)**
- **Variation (45)**
- **Localization Problem (15)**

More...



Gap	62	Under Review	9
Clone Problem	22	Awaiting Elec Data	2
Path Problem	13	Awaiting Exptl Data	1
Variation	45	Stalled	1
Localization Problem	7	Awaiting External Info	3
Missing sequence	9	Resolved	
GRC Housekeeping	24		
Unknown	3		

Items 1 - 30 o

Issue ID	Type	Location	Total placen
HG-2581	Variation	chr1:19,218,148-19,370,277	1

What about this issue?

Human Genome Issue HG-2581

Summary:	The Reference does not represent the coding allele for GeneID: 246181 (AKR7L)
Description:	The Reference does not represent the coding allele for GeneID: 246181 (AKR7L).
Status:	Resolved (GRC Resolved- No Change)
Type:	Variation
Last updated:	2020-10-07
Affects version:	GRCh38
Fix version:	GRCh39
Resolution:	The mismatch needed to represent a coding allele for AKR7L (rs190747734) falls below allele frequency of .05, making in a rare allele.

Patches and alternate loci

No patches or alts are associated with HG-2581.

Find a gene: TFIIB, CFTR, GCDH, or ACE2
http://genome.ucsc.edu/cgi-bin/hgGateway

- Choose the assembly you want (use the latest release, HG38)
- Type “TFIIB”, “CFTR”, “GCDH” or “ACE2” in search term bar
- Following the search, click on the first or second line
- For GCDH:
  - It’s on chromosome 19, left arm
  - spans nearly 8,787 bp
  - 11 exons
  - disease association (turn on OMIM genes track in “Phenotype and literature” bar, and hit refresh button)

Tracks: choose to visualize annotations to the gene
Example: OMIM, are there disease associations?
OMIM links: description of gene and its disease linkage

What are the mutations in GCDH associated with glutaricaciduria in Pennsylvania Amish?

- Click on dark green OMIM bar (once the “OMIM genes” track has been turned on)
- Choose the OMIM link 608801
- Under Molecular Genetics heading, “…a single mutation was found as the cause of glutaric acidemia in the Old Order Amish of Lancaster County, Pennsylvania (A421V; 608801.0002), Biery et al. (1996).
- https://www.youtube.com/watch?v=N2ox8g4uQqc&feature=youtu.be

What am I looking at? What does it mean?

- Each track has it's own description and options that can be changed
- The top 'track' is the Gencode Track

- For a description of how a track works:
  - Example: for Gencode, go to the "Genes and Gene Predictions" bar below
  - Expand it by clicking '+' if it isn't already
  - 'Gencode v36' should be in 'pack' mode
  - Click on the **Gencode v36** link

- Note that non-coding and splice variants are shown by default

How many versions of the genome are there?

- Human to human variation in sequence
  - The 1000 genomes initiative http://www.internationalgenome.org/1000-genomes-browsers
  -
  - Personal genome project (voluntary) http://www.personalgenomes.org/
- Variation within a person
  - Accumulation of mutations with age?
  - Mutations associated with disease, e.g. cancer genomes (Cancer Genome Atlas: https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga
)

Some applications of human genome sequences

Personalized genomic medicine, family planning

- Will I get a disease? What should I do if so
- Will my children suffer genetic disorders

Medical research: genetic basis for disease and effective treatments

Geneology and human lineages: how have human populations evolved and migrated?

Want your genome sequenced for science?

http://www.personalgenomes.org/

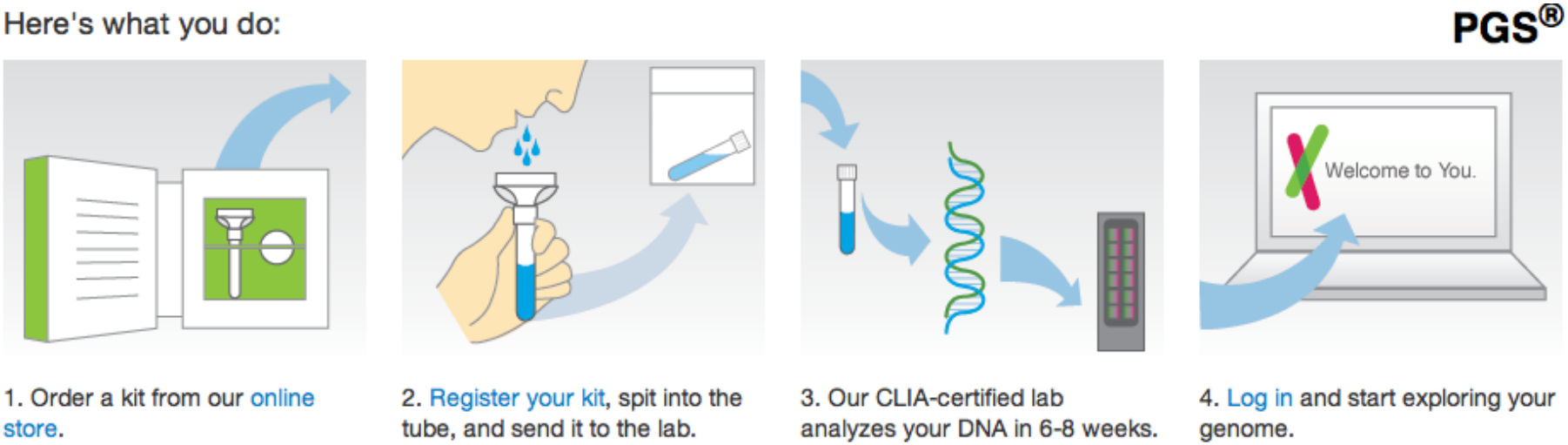
The mission of the Personal Genome Project is to encourage the development of personal genomics technology and practices that:

- are effective, informative, and responsible
- yield identifiable and improvable benefits at manageable levels of risk
- are broadly available for the good of the general public

Family member wanting to get genome sequenced for science: what if another family member prefers genetic privacy?

- Participants in the Personal Genome project https://www.youtube.com/watch?v=mVZI7NBgcWM

How to purchase your “personal genome”



What you get: a characterization of your “SNPs”, or “single nucleotide polymorphisms”. SNP chips are used.

There are differences in human genomes from one person to another, which can give information about ancestry. (Some of these changes correlate with disease states)

The SNP information can difficult to interpret. Effects of many mutations vary as a function of context.
Also, the genome service will want to sell your information

How SNPs are detected

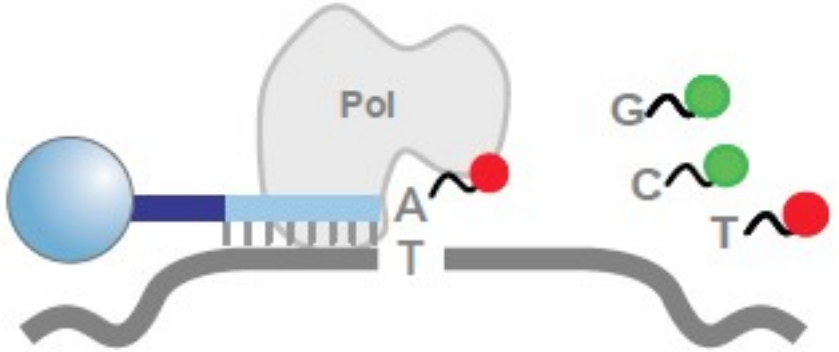
1)DNA is purified, then amplified (whole genome amplification)

2) DNA is fragmented and hybridized to probe DNA on array

3)Hybridized probe is extended with fluorophore-containing nucleotide

Step 1. Selectivity

Hybridization of unlabeled DNA fragment to 50mer probe on array



Step 2. Specificity

Enzymatic single base extension with labeled nucleotide

http://res.illumina.com/documents/products/datasheets/datasheet_omni_whole-genome_arrays.pdf

Diagnostics and personal genome services

DIRECT-TO-CONSUMER GENETIC TESTS:
"Misleading Test Results Are Further Complicated by Deceptive Marketing and Other Questionable Practices"

US Govt. Accountability Office (GAO) report (2010)

Contradictory Risk Predictions for Prostate Cancer and Hypertension

Gender	Age	Condition	Company 1	Company 2	Company 3	Company 4
Male	48	Prostate cancer	Average	Average	Below average	Above average
		Hypertension	Average	Below average	Above average	Not tested

Source: GAO.

Personal genome services and the FDA

Prior to 2010, several companies marketed “personal genome services” (no doctor’s order required) for learning disease susceptibility

In 2010, the FDA notified 17 personal genome service companies that their services are essentially medical devices, and thus require review and approval

Since these tests are unlikely to accurately predict disease risk, most companies folded

23 and Me held out until 2013, when it received a warning letter from the FDA, and switched to “ancestry genetic report”

In 2017: 23 and me received approval to notify customers of genetic disease risks for 10 conditions
http://www.latimes.com/business/la-fi-23andme-reports-20170414-htmlstory.html

Direct-to-consumer genetic tests and the FDA

1st FDA approval (2015): Bloom's Syndrome carrier test (23&Me)

- Carrier screening tests are medical devices, and classified as "Class II", since higher risk, requiring greater regulatory controls to ensure device safety and effectiveness (example: condoms are Class II devices)

- Test doesn't require a licensed practioner, but must include:
  - explanation of what results might mean to prospective parents
  - instructions for accessing a board-certified clinical molecular geneticist or equivalent
- Additional tests have been approved since then:
  - https://www.fda.gov/medical-devices/vitro-diagnostics/direct-consumer-tests
  - http://www.latimes.com/business/la-fi-23andme-reports-20170414-htmlstory.html

What DNA tests can and can’t tell you about ancestry

- https://www.vox.com/videos/2019/4/16/18410869/dna-genetic-ancestry-tests

Genetic privacy: The Human Genome Project Ethical, Legal, and Social Issues (ELSI)

http://www.ornl.gov/sci/techresources/Human_Genome/elsi/elsi.shtml
and https://www.genome.gov/Funded-Programs-Projects/ELSI-Research-Program-ethical-legal-social-implications

In May 2008: GINA (Federal Genetic Information Non-discrimination Act) passed

(http://www.ginahelp.org/GINAhelp.pdf)

- Health insurance companies may not treat people differently based on genetic code
- Employers **cannot**
  - demand genetic tests
  - discriminate against who they hire or how much they pay on the basis of genetic information
  - disclose genetic information in their possession except under specific and specially controlled circumstances.

Some issues with GINA

- Some kinds of insurance are not included in GINA, including disability, life, or long-term care insurances
- May clash with established state policies
- Doesn’t specify regulations for “Personal Genome Services”

Two articles published in 2019 (PDFs on D2L)

How is the law responding to issues surrounding genetic testing?

And

How can human genome information be studied and used equitably?

The Belmont Report (1979)

- Ethical Principles and Guidelines for the Protection of Human Subjects of Research

- Inspired in part by the ethical violations reported in the Tuskegee Syphilis Study (https://www.cdc.gov/tuskegee/timeline.htm)
  - Syphilis was left untreated in group of black men, to study progression of the disease
  - The men were willing participants, but were not informed of the study or its purpose

- Three fundamental ethical principles were defined in response to prevent further ethical failures

Belmont Report Principles

1) The principle of Respect for Persons acknowledges the dignity and autonomy of individuals, and requires that people with diminished autonomy be provided special protection. This principle requires that subjects give informed consent to participation in research

2) The principle of Beneficence requires us to protect individuals by maximizing anticipated benefits and minimizing possible harms

3) The principle of Justice requires that we treat subjects fairly

So, research on human subjects requires adherence to these guiding principles:

Consent, beneficence, and fair treatment

These guidelines are clearly relevant when considering how human DNA sequences should be used in the clinic, in research, and elsewhere

A study of human lineages: a 90-year old lock of hair from an indigenous Australian man yields complete genome sequence

- Finding: indigenous Australians are descendents of the first humans to leave Africa (other Asian populations came from a second migration)
- approval was given by representatives of indigenous group from the region where man would have lived
- What about other indigenous individuals?
- Other proposed studies have been severely restricted by indigenous Australians (ie. Consent not given)
- Archaeological specimen: is consent required? For how long?
- How must human body parts and specimens held in museums (like mummified remains) be treated?
- https://www.nature.com/news/2011/110928/full/477522a.html

  - (SEE ALSO: Genome injustice 2019, posted on D2L)

The human genome: applications and implications

1) What is the human genome? History of the human genome sequencing project, & where we are now

2) Where is the human genome and how is it annotated? UCSC browser and ‘tracks’

3) Versions of the genome

4) Ethics and the genome