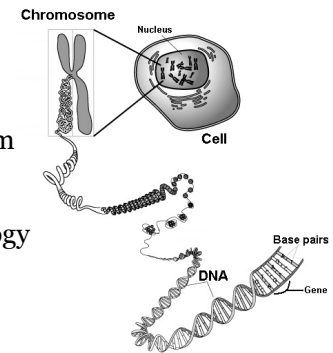


1. Mystery of the devious defecator
2. 'In the news' sign-up and expectations
3. Genetic data collection
4. Lab on genome size and gene/protein content

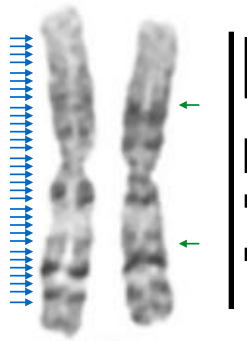
- Last time:
  - What is DNA?
  - What information can we get from DNA?
  - Central dogma of molecular biology
- Remarks:
  - Genetics isn't everything, environment is also important



## How is genetic data collected?



<https://www.youtube.com/watch?v=qUaFYzFFbBU>



**Sequencing**  
contiguous regions

**Genotyping (SNPs)**  
100,000s positions

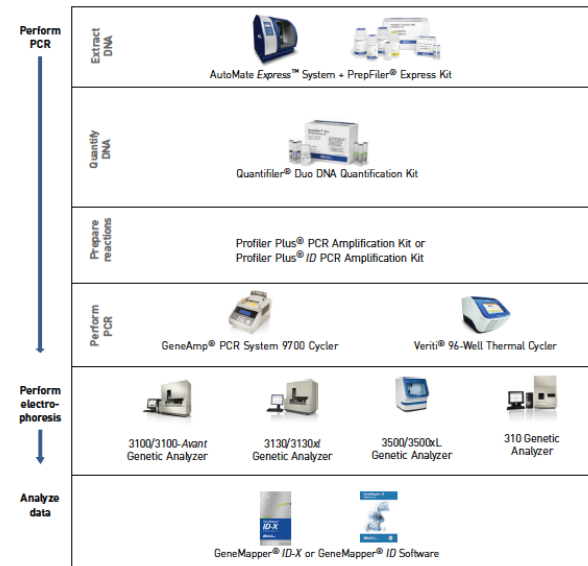
**DNA fingerprinting (STRs)**  
13 positions (FBI)

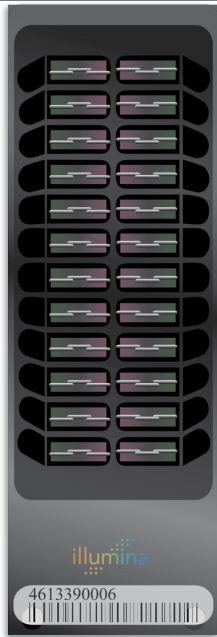
## Two basic approaches

1. Sequencing: ...CACAGTGGTACGT...
2. Probing for specific patterns:
  - single nucleotide polymorphisms (C or G at a particular site)
  - short tandem repeat counts (12 copies of TACT in a row)

The instruments for  
extracting this data uses  
quite advanced technology

Beyond the scope of this course





From 23andMe website:

- “The BeadChip is a small glass slide with millions of tiny beads on its surface. Attached to each bead are probes - bits of DNA complementary to sites in your genome where SNPs of interest are located. Your DNA will stick to the probe that matches whichever SNP you happen to have.”
- “In the genotyping process, a person’s DNA is chopped up into pieces and washed over the chip, where each fragment sticks to any probes that are complementary to it. Then fragments of DNA that have been specially tagged are introduced to the chip in such a way that they stick to any probes that are themselves paired perfectly with sample DNA. At that point, the tagged DNA fragments are triggered to glow indicating which version of each SNP is present in the sample.”

<https://www.23andme.com/more/genotyping/>

## Sample genotyping output

rsid	chromosome	position	allele1	allele2
rs4477212	1	82154	T	T
rs3131972	1	752721	G	G
rs12562034	1	768448	A	G
rs11240777	1	798959	G	G
rs6681049	1	800007	C	C
rs4970383	1	838555	A	A
rs4475691	1	846808	T	T
rs7537756	1	854250	G	G
rs13302982	1	861808	G	G
rs1110052	1	873558	T	G
rs17160698	1	887162	T	T
rs3748597	1	888659	C	C
rs13303106	1	891945	A	G
rs28415373	1	893981	C	C
rs13303010	1	894573	A	A
rs6696281	1	903104	C	C
rs28391282	1	904165	G	G
rs2340592	1	910935	A	G
rs13303118	1	918384	T	G
rs2341354	1	918573	A	G
rs6665000	1	924898	A	A
rs2341362	1	927309	C	C
rs9777703	1	928836	T	T
rs1891910	1	932457	G	G
rs9697457	1	934345	G	G

## Sample sequencing output

```
>ENA|AJ437030|AJ437030.1 Human immunodeficiency virus type 1 mRNA for gag polyprotein (gag gene), clone g20s1
ATGGGTGCGAGAGCGTCAGTATTAAAGCGGGGAAAATTAGATAAAATGGGAGAAAATTCGG
TTAAGGCCAGGGGGGAAGAAACATTATAAATTAAACATATAGTATGGGCAAGCAGGGAG
CTAGAACGATTTCACCTTAATCCTGGCCTTTTAGAGACATCAGAAGGCTGTATACAGATA
CTAAACAGCTACACACCTCCCTTCAGACAGGACAGAGAGAAATTAATCTTACATAAAT
ACAGTAGACACCTCTATGTGTACATCAAGAGTAGATGTAAGAGACCAAGCAAGGCT
TTAGACAAGATAGAGGAAGAACAACAACAAAGTAAGCAAAAGGCACAGCAGCAGCT
GATACAGGAACAGCAACACAGGCTAGCCAAAATTACCCTATAGTGCAACATTCAA
GGGCAAAATGGTACATCAGGCCCTATCACTAGAACTTTAAATGCATGGGTAAAAGTAGTA
GAAGAAAGGCTTTTCAGCCAGAAATTAACCCATGTTTCAGCAGCTTATCAGAAGGAGCC
ACCCACACAGATTAAACACCATGCTAAACACAGTGGGGGACATCAAGCAGCCATGCAA
ATGTTAAAGAGACCAATCAATGAGGAAGCTGCAGATGGGATAGATTACATCCAGTACAT
GCAGGGCTTATTCAGCCAGGCGAGATGAGAGAACCAAGGGGAAGTGCATAGCAGGAAGT
ACTAGTACCTTCAGGAACAAATAGGATGGATGACAGTACTCCACCTACCCAGTAGGA
GAAATCTATAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGAAATGTATAGCCCT
ACGAGATTCGGACATAGACAGAGGCGCAAGGAAAGGAAACCTTTAGAGACTATGTAGATCGG
TTCTATTAAGACCTTAAGAGCCGAGCAAGCTTCACAGATGTAAATAATGAGTACAGAA
ACCTTTGTTGCCAGATGCCAATCCAGATTGTAAAGACTATCTTAAAGCATTTGGACCA
GCAGCTACACTAGAAAGAAATGATGACAGCATGTCAAGGAGTGGGGGACCCAGCCATAAA
GCAAGAAATTTGGCTGAAGCAATGAGCCAGCAACAAATTCAAATATCATATGATGCAG
AGAGGCAATTTAGGAATCAAAGAGGATTTTAAAGTGTTCATTTGTGCAAGGAAGGA
CACATAGCCAAAATTCAGGGCCCTTAGGAAAAAGGCTGTTGGAGATGTGAAAAGAA
GGACACCAATGAAGATTGTACTGAAGACAGGTTAATTTTATAGGAAAATCTGGCCT
TCCCAAGGGAAGGCCAGGGAATTTCTCAGAGCAGGCCAGAGCCAAAGCCCAACCA
GAAGAGAGCTTCAGGTTCCGGGAGGAGACAACTCCCTCCAGAGCAGGAGCCGATA
GACAAGGAAGCTGTATCTTTAGCTTCCTCAAACTACTCTTTGGCAACGACCCATCGTCA
CAGTGA
```

## Sample DNA profiling output

