DNA Databases

Chapter 18
CODIS

- COmbined DNA Index System
- 1998 – launched by the FBI
- USA’s Nationwide database
  - Holding DNA profiles
  - From convicted offenders
  - Case samples
  - Missing persons
- Links all 50 states together
CODIS

• Search a DNA profile nationwide the same way police can search a nationwide fingerprint database

• End of 2003
  – Over 1.5 million STR profiles

• United Kingdom was the first to set up a nationwide DNA database
  – 1995
  – National DNA Database (NDNAD)
Why do DNA databases work?

• Why do fingerprint databases work?

• More than 60% of violent criminals are repeat offenders
  – In less than 3 years

• Information sharing is necessary between different agencies and geographically regions
Uses of DNA databases

1. Identify suspects in violent crime cases
   • DNA may be only evidence present
   • Match to offender without any leads

2. Make associations between unsolved cases across different jurisdictions
   • Crimes in different states can be linked

3. Locate missing persons
   • Identify by DNA the person/body
Value of DNA databases

• Databases are only valuable if they contain many DNA profiles
  – Effectiveness increases as databases get larger

• Huge backlog exists within the US
  – Samples are sitting in police evidence rooms without being processed
  – Matches cannot be made until samples are in the database for all to access
Cost of DNA databases

- It costs money to genotype and process all these samples
- However, crimes also cost money
  - For victims, insurance, police investigations
- A study was done to compare the costs of DNA databases vs. the crimes
- Conservative estimate of cost of a violent crime:
  - $111,238
Cost of DNA databases

• In US there are about 366,000 violent crimes per year
• 34% are “no-suspect”
  – Meaning there is no other evidence available
• 47% success rate of recovering DNA
• Two-thirds (2/3) are repeat offenders
• Average rapist commits 8 assaults prior to apprehension/arrest
## Cost of DNA databases

- 42% hit rate on “no-suspect” assault when the offender is in the database

<table>
<thead>
<tr>
<th>Calculation</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>366,460 x 34%</td>
<td>no suspect assault cases</td>
</tr>
<tr>
<td>124,596 x 2/3</td>
<td>assaults caused by repeat offenders</td>
</tr>
<tr>
<td>83,056 x 7</td>
<td>future assaults that are preventable if DNA is databased after 1st crime</td>
</tr>
<tr>
<td>581,392 x 47%</td>
<td>unnecessary victims of assaults (because DNA can be recovered)</td>
</tr>
<tr>
<td>276,626 x 42%</td>
<td>estimated assaults that could be solved with DNA databases</td>
</tr>
<tr>
<td>116,183 x $111,238</td>
<td>$12.9 Billion overall cost of preventable crimes</td>
</tr>
</tbody>
</table>

NOT having a DNA database costs $12.9 Billion per year!
Establishing a national database

Requires:

• Commitment on the part of the state and local governments to provide samples
• Common set of DNA markers, protocols and nomenclature
• High quality standards for every laboratory
• Standard software and computer formats

So that every DNA profile can be trusted and compared between labs
Technology

1. Collection of known samples
   • From convicted offenders
2. Genotyping and analyzing these samples
   • Inputting them into the database
3. Genotyping of unknown samples obtained from crime scenes
4. Comparing the DNA profiles and checking for a match
Who is entered into the database?

- Each state has their own laws for which crimes will require a DNA sample to be taken – usually violent crimes
- States are currently moving towards collecting from any felony
- Most states require the criminal be convicted before their DNA is taken
- Trend is towards broader coverage of criminals
<table>
<thead>
<tr>
<th>Offense</th>
<th>Number of States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex Crimes</td>
<td>50</td>
</tr>
<tr>
<td>Murder</td>
<td>50</td>
</tr>
<tr>
<td>Any violent crime</td>
<td>47</td>
</tr>
<tr>
<td>Felony Burglary</td>
<td>44</td>
</tr>
<tr>
<td>Felony drug crimes</td>
<td>35</td>
</tr>
<tr>
<td>Any felony</td>
<td>30</td>
</tr>
<tr>
<td>Juvenile violent crimes</td>
<td>32</td>
</tr>
<tr>
<td>Misdemeanors</td>
<td>23</td>
</tr>
<tr>
<td>Arrestees</td>
<td>4</td>
</tr>
</tbody>
</table>
Who is entered into the database?

• Current Vermont law:
  – DNA sample is taken from anyone arrested for any felony
  – If suspect is found innocent then DNA sample and records of DNA profile are destroyed
  – If suspect convicted then DNA entered into CODIS

• DNA from the crime scene are also entered into CODIS
  – Under a separate indexing
CODIS

• Actually began in 1990 as a pilot project
  – Serving just 14 labs

• During the 90’s the database grew until it was actually large enough to be useful

• As of April 2004:
  – 175 labs around the US
  – Plus 31 labs in foreign countries
  – Contains over 1.5 million DNA profiles
CODIS

Three primary sample indexes:
1. Convicted offender samples
2. Forensic casework samples
3. Missing persons samples

Also contains a population file:
• Containing DNA profiles of anonymous persons
  – Used to estimate allele frequencies
  – Calculate population statistics
DNA Profiles

Over the years the markers genotyped to produce a DNA profile has changed:

• RFLP loci
• HLA-DQA1, PolyMarker and D1S80

Now contains 13 STR core loci

– Having the 13 Core loci is now a requirement for the DNA profile to be entered into CODIS
– 13 Core loci produce 1 in 100 Trillion random match probability
What CODIS doesn’t have

CODIS contains none of the following:

• No names
• No personal identification
• No criminal history
• No case related information

Only contains a unique identifier and which index the sample is derived from
## Example

<table>
<thead>
<tr>
<th>Sample Info</th>
<th>Sample #</th>
<th>Category</th>
<th>Tissue Type</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>F130</td>
<td>1/1</td>
<td>Convicted Offender</td>
<td>Stain</td>
<td>Caucasian</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Marker</th>
<th>Value 1</th>
<th>Value 2</th>
<th>Date</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMEL</td>
<td>X</td>
<td>Y</td>
<td>2/3/1999</td>
<td>17:53:30</td>
</tr>
<tr>
<td>CSF</td>
<td>10</td>
<td>12</td>
<td>2/3/1999</td>
<td>17:53:30</td>
</tr>
<tr>
<td>D13S317</td>
<td>3</td>
<td>3</td>
<td>2/3/1999</td>
<td>17:53:30</td>
</tr>
<tr>
<td>D16S539</td>
<td>6</td>
<td>8</td>
<td>2/3/1999</td>
<td>17:53:30</td>
</tr>
<tr>
<td>D18S51</td>
<td>1</td>
<td>8</td>
<td>2/3/1999</td>
<td>17:53:30</td>
</tr>
<tr>
<td>D21S11</td>
<td>2</td>
<td>2</td>
<td>2/3/1999</td>
<td>17:53:30</td>
</tr>
<tr>
<td>D3S1358</td>
<td>14</td>
<td>15</td>
<td>2/3/1999</td>
<td>17:53:30</td>
</tr>
<tr>
<td>D7S820</td>
<td>12</td>
<td>19</td>
<td>2/3/1999</td>
<td>17:53:30</td>
</tr>
<tr>
<td>D8S1179</td>
<td>8</td>
<td>9</td>
<td>2/3/1999</td>
<td>17:53:30</td>
</tr>
<tr>
<td>FGA</td>
<td>12</td>
<td>13</td>
<td>2/3/1999</td>
<td>17:53:30</td>
</tr>
<tr>
<td>TH01</td>
<td>3</td>
<td>3</td>
<td>2/3/1999</td>
<td>17:53:30</td>
</tr>
<tr>
<td>TPOX</td>
<td>17</td>
<td>19</td>
<td>2/3/1999</td>
<td>17:53:30</td>
</tr>
<tr>
<td>VWA</td>
<td>21</td>
<td>25</td>
<td>2/3/1999</td>
<td>17:53:30</td>
</tr>
</tbody>
</table>
CODIS Match

• When CODIS finds a match
• The two labs are connected to each other
  – The lab that submitted the criminal’s DNA profile
  – The lab that submitted the case samples
• Only the local labs have the identifying information and case history
• Match is confirmed with further testing
  – If match stands up a warrant is issued
Levels of CODIS – Example

National Level

State Level

Local Level

Levels of CODIS

- CODIS is built of local, state and national labs (the FBI) all working together
- Each state has a single state lab that functions as the State DNA Index (SDIS)
- The local labs feed into the SDIS
- The state labs feed into the NDIS
- FBI is responsible for maintaining the single repository at the national level
Requirements

Before a state can have its DNA profiles included in CODIS must complete the following:

1. Prove that the lab adheres to strict quality assurance standards
2. Also adheres to nomenclature and formatting standards
3. A minimum number of at least 10 of the core loci must be in each DNA profile
CJIS WAN

• Criminal Justice Information Services Wide Area Network
• Like the internet but access is completely blocked to anyone outside the CODIS laboratories
• Completely secure network that connects all CODIS labs to each other and to the database itself
Convicted Offender Samples

- Come into the labs as either liquid blood or buccal swab
- Enough DNA to set up a cell line if needed
- Can be processed quickly and in an automated fashion
- Most of this work is currently being outsourced to private companies in order to catch up on backlog
Forensic Casework Samples

• Much more complex samples
• May be blood or semen stains, liquid or dried samples
• Variety of formats and tissues
• Process of DNA extraction is much more complicated and usually cannot be automated easily
• Quality Assurance regulates that all samples are handled to same quality standards
Database for Missing Persons

• DNA may be obtained from relatives
  – Mitochondrial or Y-DNA
• Or from missing individual’s possessions
• Database can be matched to unidentified persons or bodies that are discovered
• Without a national database this sort of search would require fingerprints or some other identifying features that may not be on record
Database for Mass Disaster

• In the cases of a mass disaster bodies can be identified against a database

For example:

• After September 11th families of people who were missing submitted their DNA or the person’s possessions to a overall database

• When bodies/parts were recovered they could be matched to database
Important Issues for DNA databases

1. Maintaining Privacy
2. Maintaining Quality Control
3. Search and Match Algorithms
4. Sample collection from convicted offenders
5. Working unknown suspect cases
6. Measuring success of the database
Maintaining Privacy

The possibility that DNA or a DNA profile can be used against a person needs to be addressed:

1. If anyone working for CODIS is caught distributing information they are penalized

2. DNA markers are in non-coding regions of the genome

   - This means they will not predict anything about genetic predisposition to diseases
Maintaining Privacy

The possibility that DNA or a DNA profile can be used against a person needs to be addressed:

3. No identifying information is contained with the DNA profile
   – Everything is coded
   – The information is broken into two parts

• Like the NOC list from Mission Impossible
Maintaining Quality Control

• With all these different labs conducting the genotyping and inputting the DNA profiles, quality control must be maintained across all CODIS labs.

• Assurance of quality control comes from:
  – Requiring all CODIS labs to follow strict standards
  – Frequent audits of all labs
  – Regular proficiency tests of analysts
Search and Match Algorithms

• Need to be able to search rapidly
  – New algorithms are always being developed
• Match between different labs
• Different labs use different genotyping kits
  – Different primer binding sites
  – Different product sizes
• Can lower stringency of a match
• Can set microvariants as .x so that all microvariants will be equivalent
Sample collection from convicted offenders

- Incarcerated felons are not always cooperative to having their blood or saliva sample taken
- Used to fight sample collection in court
  - Now laws are clear and precise as to who must submit DNA
- Still may try to fight the actual sample collection
Working unknown suspect cases

• Working with forensic samples is time consuming and difficult

• May require:
  – Extraction from different substrates
  – Separating mixtures
  – Testifying in court

• However without the forensic sample the database is worthless
  – Therefore worth money and time invested
Measuring success of database

- Usually success is calculated by counting the number of “hits”
- Hit = match is found within the database
  - Probability of random match is calculated
- Case to offender hits:
  - A forensic sample is matched to a convicted offender within the database
- Case to case hits:
  - Two or more forensic samples are matched to each other through use of database
DNA Database laws

• Databases work because criminals are often repeat offenders
• The earlier a DNA profile is put into the database the more crimes can be successfully matched
• Serial crimes can be linked together through the database
• Ultimately want to apprehend criminals and prevent future crimes
Crimes for Inclusion in database

- As of June 1998 all 50 states require convicted offenders of certain crimes to submit DNA to the database.
- Yet each state has different specific crimes for inclusion:
  - All 50 agree that convictions for murder and sexual assault require DNA submission.
  - Trend is towards any felony convictions.
  - A few states will take DNA of a suspect.
State Laws

• Some states specify exactly what sample is needed
• Other simply require any biological sample
• Example – California:
  – Requires two specimens of blood, a saliva sample and fingerprints
• South Carolina:
  – “Suitable sample from which DNA may be obtained”
DNA Databases around the world

• Many countries are now using or starting national databases
• Overlap is critical for international crimes
• Between the UK and the CODIS
  – Eight STR’s overlap
• Most countries will continue to buy STR genotyping kits from two US companies
  – Overlap will continue
Any Questions?

• Review Chapters 12, 13, 15, 16, 18

• Email me at least 5 questions you have about these chapters

• Next class will be review for Exam

• Exam Three – March 17th