How to find and interpret genomic variants in Next Generation Sequencing data

Sophia Derdak
Barcelona, May 3rd 2016
Next generation sequencing (previously: Second generation sequencing)

“next”?  
“after” Sanger...
Genome sequencing milestones

1977
Sanger Sequencing

1987
Sanger Sequencing using fluorescent dyes

2000
Sequencing-by-synthesis

2003
Human Genome Project completed

2006
Solexa Genome Analyzer launched

2009
Illumina Genome Analyzer IIx

2010
Illumina HiSeq launched

COMING UP:
- higher throughput
- clinical applications
- single cell/single molecule capacity
Illumina sequencing

INPUT:
- whole genome in fragments
- optional: selection of coding regions (“exome”)

SCAFFOLD:
- flowcell
- no beads
- no microwells

READOUT:
- fluorescent, base-by-base

Introduction

From flowcell to computer: Base calling

- The sequence of colors read for each cluster in each cycle are translated to nucleotide sequence

>100 Million Clusters Per Flow Cell

Base Calls
What are genomic variants?
Identification of genetic differences in comparison to a reference

Reference (haploid) TGGACCATCTGGTTGAGCATGTGGGGGTCAACTCCACATTCCCAGGGAGCCCCCGG

The true diploid genome of the sample TGGACCATCTGGTTGAGCATGTGGGGGTCAACTCCACATTCCCAGGGAGCCCCCGG

- ref/ref 0/0 homozygous reference
- ref/alt 0/1 heterozygous
- alt/alt 1/1 homozygous alternative
- ref/alt 0/1 heterozygous

~3.700.000 variant positions / 3.200.000.000 base position genome

>99% of the genomic positions are **not** variant positions
Genome sequencing: the experimental workflow

- Biological material
  - pool of cells
  - 2 homologous chromosomes per cell

DNA sample

DNA fragments
  - overlapping

PCR

amplified DNA fragments

dilution

load on sequencer

raw read data

variant calling

reference sequence

aligned read pairs

alignments
Mapping of reads to the reference sequence

(adapted from wikipedia)
From alignments to variants: the bioinformatic workflow

- bam processing
- remove duplicates
- realignment
- variant calling
- annotations
- filtering

+.bam alignment file

+.vcf variant call file

+.fasta reference genome file

+.pileup position file

- .bam alignment file
- inspect alignments
- + study-specific information

Genomic Variants in Next Generation Sequencing Data
Facultad de Farmacia de la UB, May 3, 2016
Sophia Derdak, CNAG
Variant calling: Variants in the sequencing data

Identification of genetic differences in comparison to a reference

Reference (haploid) TGGACCATCTGGTGAGCATGTGGGGGTCAACTCCACATTCCAGGGAGCCCCCCG

The true diploid genome of the sample TGGACCATCTGGTGAGCATGTGGGGGTCAACTCCACATTCCAGGGAGCCCCCCG

<table>
<thead>
<tr>
<th>Reference</th>
<th>TGGACCATCTGGTGAGCATGTGGGGGTCAACTCCACATTCCAGGGAGCCCCCCG</th>
</tr>
</thead>
<tbody>
<tr>
<td>ref/ref</td>
<td>0/0 homzygous reference</td>
</tr>
<tr>
<td>ref/alt</td>
<td>0/1 heterozygous</td>
</tr>
<tr>
<td>alt/alt</td>
<td>1/1 homozygous alternative</td>
</tr>
</tbody>
</table>

Aligned sequencing data derived from the sample

<table>
<thead>
<tr>
<th>Aligned sequencing data</th>
<th>TGGACCATCTGGTGAGCATGTGGGGGTCAACTCCACATTCCAGGGAGCCCCCCG</th>
</tr>
</thead>
<tbody>
<tr>
<td>ref/ref</td>
<td>0/0 homzygous reference 0% alternative allele</td>
</tr>
<tr>
<td>ref/alt</td>
<td>0/1 heterozygous 50% alternative allele</td>
</tr>
<tr>
<td>alt/alt</td>
<td>1/1 homozygous alternative 100% alternative allele</td>
</tr>
</tbody>
</table>

 ref/ref 0/0 homozygous reference 0% alternative allele
 ref/alt 0/1 heterozygous 50% alternative allele
 alt/alt 1/1 homozygous alternative 100% alternative allele

Variant calling: Variants in the sequencing data

Genomic Variants in Next Generation Sequencing Data
Facultad de Farmacia de la UB, May 3, 2016
Sophia Derdak, CNAG

List of variant positions
Bioinformatic tools for single nucleotide variant calling

- samtools + bcftools (Sanger Institute, UK, and Broad Institute, US)

- Genome Analysis Tool Kit (GATK) (Broad Institute, US)

- VarScan (Washington University)

- Platypus (Welcome Trust Center, UK)

- freebayes (Boston College, US)

Keep in mind that different software use different algorithms and thresholds and results may vary **A LOT**.

The coverage represents the number of times a base of the sample genome (or target region) is read during sequencing.

A higher coverage provides higher power for data analysis.

How to get a higher coverage:

- mainly by loading more sequencing units (indexes, lanes, entire flowcells) with the same library preparation

Typical coverage numbers (in CNAG projects):

- whole genome: 30x
- exome: 50-100x
- custom gene panel capture: >1000x
“I believe that we do not know anything for certain, but everything probably.”
Christiaan Huygens

- base calling (base qualities in the fastq files)
- contig order in the reference assembly
- reference sequence (not yet...)
- read alignment (mapping quality)
- variant position (variant and genotype quality)

- p-values
- probability likelihoods
- PHRED scores

Plato, ~400 BC
### raw vcf file ("all variants")

#### Mostly experiment-independent technical and quality filtering (well-covered positions with confident alternative allele)

### filtered vcf file ("good quality variants")

<table>
<thead>
<tr>
<th>CHR</th>
<th>POS</th>
<th>REF</th>
<th>ALT</th>
<th>GT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>148588972</td>
<td>G</td>
<td>C</td>
<td>0/1</td>
</tr>
<tr>
<td>1</td>
<td>154284894</td>
<td>A</td>
<td>G</td>
<td>0/1</td>
</tr>
<tr>
<td>1</td>
<td>203923829</td>
<td>A</td>
<td>G</td>
<td>0/1</td>
</tr>
<tr>
<td>1</td>
<td>243329075</td>
<td>T</td>
<td>C</td>
<td>0/1</td>
</tr>
<tr>
<td>2</td>
<td>102968362</td>
<td>T</td>
<td>C</td>
<td>0/1</td>
</tr>
<tr>
<td>2</td>
<td>122096456</td>
<td>G</td>
<td>A</td>
<td>0/1</td>
</tr>
<tr>
<td>2</td>
<td>242612151</td>
<td>C</td>
<td>T</td>
<td>1/1</td>
</tr>
<tr>
<td>3</td>
<td>56591283</td>
<td>TAAGCAGGGG</td>
<td>TAAGCAGGGGGAAGCAGGGG</td>
<td>0/1</td>
</tr>
<tr>
<td>4</td>
<td>146297387</td>
<td>CAAAAAAA</td>
<td>CAAAAAAADAAAA</td>
<td>0/1</td>
</tr>
<tr>
<td>6</td>
<td>116263181</td>
<td>T</td>
<td>C</td>
<td>1/1</td>
</tr>
<tr>
<td>8</td>
<td>96070181</td>
<td>T</td>
<td>C</td>
<td>1/1</td>
</tr>
<tr>
<td>9</td>
<td>129831659</td>
<td>T</td>
<td>A</td>
<td>1/1</td>
</tr>
<tr>
<td>9</td>
<td>131454120</td>
<td>C</td>
<td>T</td>
<td>1/1</td>
</tr>
<tr>
<td>10</td>
<td>29834095</td>
<td>G</td>
<td>A</td>
<td>1/1</td>
</tr>
<tr>
<td>11</td>
<td>18159254</td>
<td>A</td>
<td>G</td>
<td>1/1</td>
</tr>
<tr>
<td>11</td>
<td>35274829</td>
<td>A</td>
<td>G</td>
<td>0/1</td>
</tr>
<tr>
<td>12</td>
<td>21628320</td>
<td>C</td>
<td>T</td>
<td>0/1</td>
</tr>
<tr>
<td>15</td>
<td>42619508</td>
<td>C</td>
<td>T</td>
<td>1/1</td>
</tr>
<tr>
<td>15</td>
<td>75336729</td>
<td>A</td>
<td>G</td>
<td>0/1</td>
</tr>
<tr>
<td>16</td>
<td>84035844</td>
<td>T</td>
<td>C</td>
<td>0/1</td>
</tr>
</tbody>
</table>

chr chromosome

pos position on the chromosome

ref sequence in the reference genome

alt alternative sequence detected in the sample

gt genotype in the (diploid) sample
Variant calling: multi-sample analyses

Somatic variants

Inheritance
De novo variants

Affected vs. control group
Compare two samples of the same individual (e.g. tumor-normal)

vcf file ("good quality variants - all genotypes")

Definition of “somatic variant”, consider sample purity information
Select variants with genotype 0/0 in the normal and 0/1 in the tumor sample
Additionally, select alternative allele frequency thresholds for normal and tumor sample using AC

filtered vcf file ("somatic variants")

<table>
<thead>
<tr>
<th>CHR</th>
<th>POS</th>
<th>REF</th>
<th>ALT</th>
<th>GT_normal</th>
<th>AC_normal</th>
<th>GT_tumor</th>
<th>AC_tumor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1421916</td>
<td>T</td>
<td>C</td>
<td>0/0</td>
<td>37,1</td>
<td>0/1</td>
<td>44,7</td>
</tr>
<tr>
<td>1</td>
<td>179528803</td>
<td>A</td>
<td>C</td>
<td>0/0</td>
<td>53,0</td>
<td>0/1</td>
<td>53,16</td>
</tr>
<tr>
<td>1</td>
<td>59096853</td>
<td>C</td>
<td>T</td>
<td>0/0</td>
<td>19,1</td>
<td>0/1</td>
<td>21,7</td>
</tr>
<tr>
<td>2</td>
<td>115417479</td>
<td>C</td>
<td>T</td>
<td>0/0</td>
<td>11,0</td>
<td>0/1</td>
<td>21,5</td>
</tr>
<tr>
<td>2</td>
<td>166756497</td>
<td>T</td>
<td>A</td>
<td>0/0</td>
<td>20,1</td>
<td>0/1</td>
<td>23,8</td>
</tr>
<tr>
<td>3</td>
<td>53910122</td>
<td>G</td>
<td>A</td>
<td>0/0</td>
<td>28,0</td>
<td>0/1</td>
<td>29,7</td>
</tr>
<tr>
<td>7</td>
<td>151962062</td>
<td>A</td>
<td>G</td>
<td>0/0</td>
<td>12,0</td>
<td>0/1</td>
<td>19,5</td>
</tr>
<tr>
<td>9</td>
<td>136083801</td>
<td>A</td>
<td>G</td>
<td>0/0</td>
<td>10,0</td>
<td>0/1</td>
<td>16,5</td>
</tr>
<tr>
<td>11</td>
<td>89407177</td>
<td>C</td>
<td>T</td>
<td>0/0</td>
<td>15,0</td>
<td>0/1</td>
<td>31,6</td>
</tr>
<tr>
<td>12</td>
<td>129298780</td>
<td>G</td>
<td>A</td>
<td>0/0</td>
<td>19,1</td>
<td>0/1</td>
<td>30,6</td>
</tr>
<tr>
<td>15</td>
<td>20454042</td>
<td>A</td>
<td>T</td>
<td>0/0</td>
<td>22,1</td>
<td>0/1</td>
<td>24,5</td>
</tr>
<tr>
<td>15</td>
<td>21133683</td>
<td>T</td>
<td>C</td>
<td>0/0</td>
<td>12,0</td>
<td>0/1</td>
<td>4,6</td>
</tr>
<tr>
<td>17</td>
<td>15468718</td>
<td>G</td>
<td>A</td>
<td>0/0</td>
<td>11,0</td>
<td>0/1</td>
<td>16,5</td>
</tr>
<tr>
<td>17</td>
<td>15468728</td>
<td>T</td>
<td>C</td>
<td>0/0</td>
<td>11,0</td>
<td>0/1</td>
<td>19,5</td>
</tr>
<tr>
<td>17</td>
<td>36365191</td>
<td>A</td>
<td>T</td>
<td>0/0</td>
<td>11,0</td>
<td>0/1</td>
<td>18,4</td>
</tr>
<tr>
<td>17</td>
<td>66195635</td>
<td>T</td>
<td>G</td>
<td>0/0</td>
<td>12,0</td>
<td>0/1</td>
<td>12,5</td>
</tr>
<tr>
<td>19</td>
<td>43783125</td>
<td>C</td>
<td>A</td>
<td>0/0</td>
<td>10,0</td>
<td>0/1</td>
<td>24,5</td>
</tr>
<tr>
<td>19</td>
<td>43783146</td>
<td>A</td>
<td>T</td>
<td>0/0</td>
<td>13,0</td>
<td>0/1</td>
<td>29,8</td>
</tr>
<tr>
<td>20</td>
<td>29628070</td>
<td>T</td>
<td>C</td>
<td>0/0</td>
<td>13,0</td>
<td>0/1</td>
<td>22,4</td>
</tr>
<tr>
<td>21</td>
<td>11181025</td>
<td>G</td>
<td>A</td>
<td>0/0</td>
<td>39,0</td>
<td>0/1</td>
<td>36,10</td>
</tr>
</tbody>
</table>

CHR chromosome
POS position on the chromosome
REF sequence in the reference genome
ALT alternative sequence detected in the sample
GT genotype in the (diploid) sample, per sample
AC allele count, number of (ref, alt) bases, per sample
Compare three samples of a pedigree

vcf file ("good quality variants - all genotypes")

Apply model of inheritance: e.g. autosomal recessive
Select variants with genotype 0/1 in the parents and 1/1 in the daughter

filtered vcf file ("recessively inherited variants")

<table>
<thead>
<tr>
<th>CHR</th>
<th>POS</th>
<th>REF</th>
<th>ALT</th>
<th>GT_daughter</th>
<th>GT_father</th>
<th>GT_mother</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>200827638</td>
<td>A</td>
<td>G</td>
<td>1/1</td>
<td>0/1</td>
<td>0/1</td>
</tr>
<tr>
<td>1</td>
<td>22158157</td>
<td>A</td>
<td>G</td>
<td>1/1</td>
<td>0/1</td>
<td>0/1</td>
</tr>
<tr>
<td>2</td>
<td>171256597</td>
<td>A</td>
<td>C</td>
<td>1/1</td>
<td>0/1</td>
<td>0/1</td>
</tr>
<tr>
<td>2</td>
<td>208976955</td>
<td>A</td>
<td>C</td>
<td>1/1</td>
<td>0/1</td>
<td>0/1</td>
</tr>
<tr>
<td>4</td>
<td>48496368</td>
<td>A</td>
<td>G</td>
<td>1/1</td>
<td>0/1</td>
<td>0/1</td>
</tr>
<tr>
<td>7</td>
<td>14017007</td>
<td>C</td>
<td>T</td>
<td>1/1</td>
<td>0/1</td>
<td>0/1</td>
</tr>
<tr>
<td>8</td>
<td>143310815</td>
<td>G</td>
<td>A</td>
<td>1/1</td>
<td>0/1</td>
<td>0/1</td>
</tr>
<tr>
<td>8</td>
<td>41517860</td>
<td>G</td>
<td>A</td>
<td>1/1</td>
<td>0/1</td>
<td>0/1</td>
</tr>
<tr>
<td>11</td>
<td>47437403</td>
<td>C</td>
<td>T</td>
<td>1/1</td>
<td>0/1</td>
<td>0/1</td>
</tr>
<tr>
<td>11</td>
<td>59837097</td>
<td>C</td>
<td>T</td>
<td>1/1</td>
<td>0/1</td>
<td>0/1</td>
</tr>
<tr>
<td>12</td>
<td>9833628</td>
<td>C</td>
<td>T</td>
<td>1/1</td>
<td>0/1</td>
<td>0/1</td>
</tr>
<tr>
<td>13</td>
<td>36699762</td>
<td>G</td>
<td>A</td>
<td>1/1</td>
<td>0/1</td>
<td>0/1</td>
</tr>
<tr>
<td>13</td>
<td>52523808</td>
<td>C</td>
<td>T</td>
<td>1/1</td>
<td>0/1</td>
<td>0/1</td>
</tr>
<tr>
<td>14</td>
<td>38256944</td>
<td>T</td>
<td>C</td>
<td>1/1</td>
<td>0/1</td>
<td>0/1</td>
</tr>
<tr>
<td>15</td>
<td>79026001</td>
<td>C</td>
<td>A</td>
<td>1/1</td>
<td>0/1</td>
<td>0/1</td>
</tr>
<tr>
<td>16</td>
<td>10788129</td>
<td>G</td>
<td>T</td>
<td>1/1</td>
<td>0/1</td>
<td>0/1</td>
</tr>
<tr>
<td>16</td>
<td>1498197</td>
<td>A</td>
<td>G</td>
<td>1/1</td>
<td>0/1</td>
<td>0/1</td>
</tr>
<tr>
<td>19</td>
<td>49640002</td>
<td>G</td>
<td>T</td>
<td>1/1</td>
<td>0/1</td>
<td>0/1</td>
</tr>
<tr>
<td>20</td>
<td>10026357</td>
<td>T</td>
<td>C</td>
<td>1/1</td>
<td>0/1</td>
<td>0/1</td>
</tr>
<tr>
<td>22</td>
<td>23657980</td>
<td>G</td>
<td>A</td>
<td>1/1</td>
<td>0/1</td>
<td>0/1</td>
</tr>
</tbody>
</table>

CHR chromosome
POS position on the chromosome
REF sequence in the reference genome
ALT alternative sequence detected in the sample
GT genotype in the (diploid) sample, per sample
Compare three samples of a pedigree

vcf file ("good quality variants - all genotypes")

Apply model of inheritance: e.g. de-novo
Select variants with genotype 0/0 in the parents and 0/1 in the daughter

filtered vcf file ("de-novo variants")

<table>
<thead>
<tr>
<th>CHR</th>
<th>POS</th>
<th>REF</th>
<th>ALT</th>
<th>GT_daughter</th>
<th>GT_father</th>
<th>GT_mother</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13365778</td>
<td>A</td>
<td>G</td>
<td>0/1</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>1</td>
<td>144676632</td>
<td>T</td>
<td>C</td>
<td>0/1</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>1</td>
<td>144853029</td>
<td>C</td>
<td>G</td>
<td>0/1</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>1</td>
<td>16891333</td>
<td>T</td>
<td>C</td>
<td>0/1</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>3</td>
<td>143697451</td>
<td>T</td>
<td>G</td>
<td>0/1</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>3</td>
<td>72311749</td>
<td>G</td>
<td>T</td>
<td>0/1</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>3</td>
<td>9057481</td>
<td>C</td>
<td>A</td>
<td>0/1</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>5</td>
<td>39002519</td>
<td>C</td>
<td>T</td>
<td>0/1</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>6</td>
<td>33060143</td>
<td>A</td>
<td>G</td>
<td>0/1</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>6</td>
<td>37845185</td>
<td>C</td>
<td>A</td>
<td>0/1</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>7</td>
<td>1586741</td>
<td>G</td>
<td>T</td>
<td>0/1</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>8</td>
<td>49987965</td>
<td>A</td>
<td>C</td>
<td>0/1</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>9</td>
<td>39888209</td>
<td>C</td>
<td>A</td>
<td>0/1</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>12</td>
<td>92562268</td>
<td>G</td>
<td>T</td>
<td>0/1</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>12</td>
<td>94034134</td>
<td>G</td>
<td>T</td>
<td>0/1</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>13</td>
<td>19042019</td>
<td>G</td>
<td>T</td>
<td>0/1</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>15</td>
<td>84855648</td>
<td>C</td>
<td>A</td>
<td>0/1</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>16</td>
<td>46427389</td>
<td>T</td>
<td>C</td>
<td>0/1</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>17</td>
<td>10550780</td>
<td>A</td>
<td>G</td>
<td>0/1</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>22</td>
<td>20643742</td>
<td>C</td>
<td>A</td>
<td>0/1</td>
<td>0/0</td>
<td>0/0</td>
</tr>
</tbody>
</table>

CHR chromosome
POS position on the chromosome
REF sequence in the reference genome
ALT alternative sequence detected in the sample
GT genotype in the (diploid) sample, per sample
… a real world success story of finding the causative variant
Discard variants because:

- they have low technical quality
- they are known polymorphisms
- they do not have a protein-coding effect
chr17:7918347, T>C, 0/1

de Castro-Miró M et al. PLOS One 2014: Combined Genetic and High-Throughput Strategies for Molecular Diagnosis of Inherited Retinal Dystrophies.
Causative variant for inherited retinal dystrophy?

chr17:7918347, T>C, 0/1

Annotations at gene level
- GUCY2D
- Retina

- ENSEMBL Functional annotations: genes, transcripts, coding sequences
- UCSC genome browser, GeneCards...
  Tissue specificity of gene function

de Castro-Miró M et al. PLOS One 2014: Combined Genetic and High-Throughput Strategies for Molecular Diagnosis of Inherited Retinal Dystrophies.
chr17:7918347, T>C, 0/1

c.2747T>C
p.I916T

de Castro-Miró M et al. PLOS One 2014: Combined Genetic and High-Throughput Strategies for Molecular Diagnosis of Inherited Retinal Dystrophies.
Causative variant for inherited retinal dystrophy?

Look up a gene in a Genome Browser:
Variants inside candidate genes or genomic regions are interesting variants

HGMD:: Human Gene Mutation Database (Cardiff University and Biobase GmbH)

OMIM :: Online Mendelian Inheritance in Man (John Hopkins University)

Orphanet :: The portal for rare diseases and orphan drugs (INSERM, France)

ClinVar :: Information about relationships among variation and human health (NCBI)

Disease-specific databases and publications (e.g. COSMIC database for cancer)

Genetic linkage studies

Helpful, when studying a case with a previously described disease phenotype

The OMIM database is available and may be queried at: http://omim.org/
The Orphanet database is available at: http://www.orpha.net/consor/cgi-bin/index.php
The COSMIC Catalogue for somatic mutations in cancer is available at:
http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/
What else can genomic variants tell us?
One of the methods to assess complex disease is GWAS – Genome Wide Association Studies.

- Look for genetic polymorphisms (not necessarily coding!) that associate with the trait
- in 1000's of samples: cases and controls, perform statistical tests
- Results can be single position or hotspot region around a position: “Manhattan plots”

Pharmacogenomics is an emerging field that combines genetics with pharmacokinetics and pharmacodynamics of drugs.

- to understand genetic polymorphisms among patients
- to study the effect of these polymorphisms on the activity of the enzyme metabolizing the drug
- to develop more accurate drug dosing in order to avoid intoxication or insufficient drug action.

Using Genetics to Tailor Drug Therapy

- **People respond differently to the drug.**
  - Most people metabolize the drug quickly. Doses need to be high enough to treat their condition effectively.
  - Others metabolize the drug slowly and need lower doses to avoid toxic side effects of the drug.
  - A small portion of people metabolize the drug poorly. They have a higher chance to have serious side effects.

- **The range of responses is due to genetic changes, or variants.**
  - No variants
    - After a simple blood test, individuals can be given doses of medication that are tailored to their genetic profile.
  - One variant
    - Normal dose
  - Two variants
    - Dose for a poor metabolizer

- **Genes with variants affecting drug action**
  - **Warfarin** (inhibitor of blood coagulation)
    - VKORC1 and CYP2C9
  - **Irinotecan** (cancer)
    - UGT1A1
  - **Thiopurine drugs** (autoimmune disorders)
    - TPMT and ITPA

A clinically relevant incidental DNA variation can be defined as a verified DNA variation that has a proven medically relevant phenotype not directly related to the condition being studied for research.

It is an unforeseen clinical finding relevant to the individual research participant involved (and possibly to the family of the participant).

- originally coined in the field of radiology

- to be discussed in the field of bioethics

Should the participant (or the participant's physician) be informed about the incidental finding?

Does it make a difference whether the incidentally discovered genetic variant points at a disease with a therapy available or not?

Properly informed consent for the study participants must explain the possibility of finding an incidental DNA variation (especially in whole genome sequencing).

Health reports

- ancestry-related genetic reports
- uninterpreted raw genetic data
- oddities:

Does fresh cilantro taste like soap to you?

Yes
No
Not sure

5th CNAG Symposium on Genome Research: Single Cell Studies

19th May 2016
Auditori Antoni Caparrós
Torre D - Parc Científic de Barcelona

Free registration at:
www.cnag.eu

Speakers

Heather Lee, Babraham Institute
Christian Conrad, DKFZ & University of Heidelberg
Salvador Aznar-Benitah, Institute for Research in Biomedicine
Thomas Graf, Centre for Genomic Regulation
Eduard Batlle, Institute for Research in Biomedicine
Ramon Massana, Institute of Marine Sciences (CSIC)
Holger Heyn, Centro Nacional de Análisis Genómico