CSE 182 Project
The Diploid Genome Sequence of an Individual Human

When in Doubt, Spit It Out

A DATE WITH DNA K. C. Dustin and his wife, Debra Netschert, give saliva samples.

By ALLEN SALKIN
Published: September 12, 2008

it has sequenced the complete genome of an African man.
DNA sequence reads
Mapping short reads to a reference genome

Chromosome 4, Position 42156427

Reference Genome

Reads

VCF File

<table>
<thead>
<tr>
<th>Chrom.</th>
<th>Pos</th>
<th>Ref.</th>
<th>Alt.</th>
<th>GT</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>42156427</td>
<td>C</td>
<td>A</td>
<td>C/A</td>
</tr>
</tbody>
</table>
Abstraction of a causal mutation

- Some mutations may change the coding part of the gene. Others may be in non-genic region, or not change the code.
- Coding mutations may be deleterious or benign
Mutations versus polymorphisms

Polymorphisms are mutations that are ‘common’, or polymorphic in a population. SNPs refer to Single Nucleotide Polymorphisms
### VCF file and SNPs

<table>
<thead>
<tr>
<th>Chrom.</th>
<th>Pos</th>
<th>Ref.</th>
<th>Alt.</th>
<th>GT</th>
<th>rs#</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>42156427</td>
<td>C</td>
<td>A</td>
<td>C/A</td>
<td>rs24167</td>
</tr>
</tbody>
</table>
Duffy et al. (2007) conducted additional genotyping to clarify the role of OCA2 locus in the inheritance of eye color and other pigmented traits associated with skin-cancer risk in white populations. The highest association for blue/nonblue eye color (227220) was found with 3 OCA2 SNPs in intron 1, rs7495174 (T/C), rs4778241 (G/T), and rs4778138 (T/C). These 3 SNPs are in 1 major haplotype block (611409.0013), with TGT representing 78.4% of alleles. The minor population impact of the nonsynonymous coding region polymorphisms arg305-to-trp (611409.0011) and arg419-to-gln (611409.0012) associated with nonblue eyes (Rebbeck et al., 2002; Jannot et al., 2005) and the tight linkage of the major TGT haplotype within intron 1 of OCA2 with blue eye color and lighter hair and skin tones suggested that differences within the 5-prime proximal regulatory region of the OCA2 gene alter expression or mRNA transcript levels and may be responsible for these associations.
Project

• Goal: extract as much information as possible for Individual X
• Map the variants of the individual to specific OMIM entries.
• (Extra) If it is not a SNP, identify if it is coding or non-coding, and/or deleterious
• (Extra) Identify the genealogy of the person.
• Start early!