Biochemistry 412

New Strategies & Technologies for DNA Sequencing

New Applications

13 February 2009
Some Applications

Now....
Made feasible by the development of modern ultra-rapid and very cheap DNA sequencing methods.
Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing

Roman K Thomas¹,²,12, Elizabeth Nickerson³,12, Jan F Simons³,11,12, Pasi A Jänne¹, Torstein Tengs¹,², Yuki Yuza¹, Levi A Garraway¹,²,4, Thomas LaFramboise¹,², Jeffrey C Lee¹,², Kinjal Shah¹,², Keith O’Neill², Hidefumi Sasaki⁵, Neal Lindeman⁶, Kwok-Kin Wong¹, Ana M Borras⁷, Edward J Gutmann⁸, Konstantin H Dragnev⁹, Ralph DeBiasi¹,², Tzu-Hsiu Chen¹,², Karen A Glatt¹, Heidi Greulich¹,², Brian Desany³, Christine K Lubeski³, William Brockman², Pablo Alvarez², Stephen K Hutchison³, J H Leamon³, Michael T Ronan³, Gregory S Turenchalk³, Michael Egholm³, William R Sellers¹,², Jonathan M Rothberg³ & Matthew Meyerson¹,²,10,11

The sensitivity of conventional DNA sequencing in tumor biopsies is limited by stromal contamination and by genetic heterogeneity within the cancer. Here, we show that microreactor-based pyrosequencing can detect rare cancer-associated sequence variations by independent and parallel sampling of multiple representatives of a given DNA fragment. This technology can thereby facilitate accurate molecular diagnosis of heterogeneous cancer specimens and enable patient selection for targeted cancer therapies.
Amplifying and sequencing total DNA from an autopsy sample can cause you to miss things.

Figure 1 Failure of Sanger sequencing to detect clinically relevant EGFR mutations in a malignant pleural effusion specimen with low tumor content. (a) Photomicrograph of a hematoxylin and eosin-stained section of a paraffin-embedded fibrin clot from the pleural effusion fluid obtained at time of relapse. Four clusters of tumor cells showing features of adenocarcinoma, including rudimentary gland formation (arrows; < 50 total cells), are found within a mix of benign inflammatory and mesothelial cells. Scale bar, 250 μm. Sanger sequencing of exons 19 (b) and 20 (c) of EGFR from DNA isolated from the sample in a. Arrows indicate the sites of the mutations revealed by picotiter plate sequencing but not visible in the electropherograms.

>>>Note: it’s not really a failure of ‘Sanger sequencing’. Their approach works due to the fact that sequencing lots and lots of clonal DNA isolates (as opposed to a mixture) gives you a more sensitive means of detecting rare sequence variations.

Figure 4  Pyrogram of the raw data obtained from liquid-phase pyrosequencing. Proportional signals are obtained for one, two, three, and four base incorporations. Nucleotide addition, according to the order of nucleotides, is indicated below the pyrogram and the obtained sequence is indicated above the pyrogram.

Sensitive detection of a tumor mutation causing drug resistance and patient relapse

Resistance mutation (frequency \( \approx 2\% \) in sample)

wt allele (frequency \( \approx 98\% \) in total DNA sample)

Detection and characterization of novel pathogens that elude standard diagnostic methods

A New Arenavirus in a Cluster of Fatal Transplant-Associated Diseases

Gustavo Palacios, Ph.D., Julian Druce, Ph.D., Lei Du, Ph.D., Thomas Tran, Ph.D., Chris Birch, Ph.D., Thomas Briese, Ph.D., Sean Conlan, Ph.D., Phenix-Lan Quan, Ph.D., Jeffrey Hui, B.Sc., John Marshall, Ph.D., Jan Fredrik Simons, Ph.D., Michael Egholm, Ph.D., Christopher D. Paddock, M.D., M.P.H.T.M., Wun-Ju Shieh, M.D., Ph.D., M.P.H., Cynthia S. Goldsmith, M.G.S., Sherif R. Zaki, M.D., Ph.D., Mike Catton, M.D., and W. Ian Lipkin, M.D.

BACKGROUND
Three patients who received visceral-organ transplants from a single donor on the same day died of a febrile illness 4 to 6 weeks after transplantation. Culture, polymerase-chain-reaction (PCR) and serologic assays, and oligonucleotide microarray analysis for a wide range of infectious agents were not informative.

METHODS
We evaluated RNA obtained from the liver and kidney transplants in two recipients. Unbiased high-throughput sequencing was used to identify microbial sequences not found by means of other methods. The specificity of sequences for a new candidate pathogen was confirmed by means of culture and by means of PCR, immunohistochemical, and serologic analyses.

RESULTS
High-throughput sequencing yielded 103,632 sequences, of which 14 represented an Old World arenavirus. Additional sequence analysis showed that this new arenavirus was related to lymphocytic choriomeningitis viruses. Specific PCR assays based on a unique sequence confirmed the presence of the virus in the kidneys, liver, blood, and cerebrospinal fluid of the recipients. Immunohistochemical analysis revealed arenavirus antigen in the liver and kidney transplants in the recipients. IgM and IgG antiviral antibodies were detected in the serum of the donor. Seroconversion was evident in serum specimens obtained from one recipient at two time points.

CONCLUSIONS
Unbiased high-throughput sequencing is a powerful tool for the discovery of pathogens. The use of this method during an outbreak of disease facilitated the identification of a new arenavirus transmitted through solid-organ transplantation.

One last item:

An exciting developing area of genomics is....

....*metagenomics*!
Metagenomics entails the direct cloning and sequencing of genomes from organisms with *unknown or impractical* cultivation conditions.

Examples of some sources of DNA for metagenomic studies:

- Obligate pathogens or symbionts
- Mixed microbial populations in environmental samples
- Extinct organisms (paleogenomics)
<table>
<thead>
<tr>
<th>Species</th>
<th>Genome size</th>
<th>Host or habitat</th>
<th>Separation technique</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Treponema pallidum</em></td>
<td>1.1 Mb</td>
<td>Human, rabbit</td>
<td>Dissection, differential lysis</td>
<td>20</td>
</tr>
<tr>
<td><em>Rickettsia prowazekii</em></td>
<td>1.1 Mb</td>
<td>Human, chicken</td>
<td>Differential centrifugation</td>
<td>21</td>
</tr>
<tr>
<td><em>Mycobacterium leprae</em></td>
<td>3.3 Mb</td>
<td>Human, armadillo</td>
<td>Gradient centrifugation</td>
<td>22</td>
</tr>
<tr>
<td><em>Tropheryma whippel</em></td>
<td>0.9 Mb</td>
<td>Human</td>
<td>Differential centrifugation</td>
<td>23</td>
</tr>
<tr>
<td><em>Buchnera aphidicola</em> str. APS</td>
<td>0.6 Mb</td>
<td>Aphid (<em>Acyrthosiphon pisum</em>)</td>
<td>Dissection, differential lysis, filtration</td>
<td>24</td>
</tr>
<tr>
<td><em>Buchnera aphidicola</em> str. Sg</td>
<td>0.6 Mb</td>
<td>Aphid (<em>Schizaphis graminum</em>)</td>
<td>Gradient centrifugation</td>
<td>25</td>
</tr>
<tr>
<td><em>Wigglesworthia glossinidia</em> brevipalpis</td>
<td>0.7 Mb</td>
<td>Tsetse fly (<em>Glossina brevipalpis</em>)</td>
<td>Dissection, differential lysis</td>
<td>26</td>
</tr>
<tr>
<td><em>Biochmannia florianus</em></td>
<td>0.7 Mb</td>
<td>Carpenter ants</td>
<td>Differential lysis</td>
<td>27</td>
</tr>
<tr>
<td><em>Buchnera aphidicola</em> str. Bp</td>
<td>0.6 Mb</td>
<td>Aphid (<em>Baizongia pistaciae</em>)</td>
<td>Differential lysis, filtration</td>
<td>28</td>
</tr>
<tr>
<td><em>Wolbachia pipentis wMel</em></td>
<td>1.27 Mb</td>
<td>Fly (<em>Drosophila melanogaster</em>)</td>
<td>Differential lysis, pulsed-field gel electrophoresis</td>
<td>29</td>
</tr>
<tr>
<td><em>Wolbachia pipentis wAna</em></td>
<td>1.4 Mb</td>
<td>Fly (<em>Drosophila ananassae</em>)</td>
<td>None</td>
<td>30</td>
</tr>
<tr>
<td><em>Wolbachia pipentis wBm</em></td>
<td>1.1 Mb</td>
<td>Parasitic nematode worm (Brugia malayi)</td>
<td>EAC library screening</td>
<td>31</td>
</tr>
<tr>
<td><em>Phytoplasma asteris</em>, line OYM</td>
<td>0.9 MB</td>
<td>Plants and leafhoppers</td>
<td>Differential lysis, pulsed-field gel electrophoresis</td>
<td>32</td>
</tr>
<tr>
<td><em>Nanoarchaeum equitans</em></td>
<td>0.5 Mb</td>
<td><em>Ignicoccus</em> sp. co-culture</td>
<td>Differential centrifugation</td>
<td>59</td>
</tr>
<tr>
<td><em>Ferroplasma acidarmanus</em> type II</td>
<td>1.8 Mb</td>
<td>Acid-mine biofilm</td>
<td>None</td>
<td>45</td>
</tr>
<tr>
<td><em>Leptospirillum</em> sp. Group II</td>
<td>2.2 Mb</td>
<td>Acid-mine biofilm</td>
<td>None</td>
<td>45</td>
</tr>
<tr>
<td><em>Burkholderia</em> sp.</td>
<td>~3.8 Mb</td>
<td>Sargasso Sea</td>
<td>Filtration</td>
<td>4</td>
</tr>
<tr>
<td><em>Shewanella</em> sp.</td>
<td>~5 Mb</td>
<td>Sargasso Sea</td>
<td>Filtration</td>
<td>4</td>
</tr>
</tbody>
</table>

Figure 1 | Sequencing of ancient DNA. Genomic sequence from extinct organisms can be obtained from the DNA in ancient remains such as bone. Bones are milled into powder and immersed in a solution to extract the DNA. The damaged ends of the DNA molecules are then repaired enzymatically and cloned into a sequencing vector. The clones are sequenced using standard protocols, and the probable species of origin is determined by BLAST analysis. In the study of Pleistocene cave bears by Noonan et al., up to 5% of the clones found their closest match in the genome of the dog, a carnivore that is closely related to bears. Only a few (~0.05%) of the reads were of human origin, whereas 10–20% had significant matches only to environmental sequences.

Metagenomic strategy for sequencing ancient DNA: Neanderthal bone sample

*Note: performed using 454 Life Sciences’s pyrosequencing technology.

Figure 2 | Taxonomic distribution of DNA sequences from the Vi-80 extract. The taxonomic order of the database sequence giving the best alignment for each unique sequence read was determined. The most populous taxonomic orders are shown.

Neanderthal DNA sequence BLAST “hits” map more or less uniformly across the human genome.

Figure 4 | Location on the human karyotype of Neanderthal DNA sequences. All sequences longer than 30 nucleotides whose best alignments were to the human genome are shown. The blue lines above each chromosome mark the position of all alignments that are unique in terms of bit-score within the human genome. Orange lines are alignments that have more than one alignment of equal bit-score. To the left of each chromosome, the average number of Neanderthal bases per 10,000 is given. Lines (Neanderthal, blue; human, red) within each chromosome show the hit density, on a log-base 2 scale, within sliding windows of 3 megabases along each chromosome. The centre black lines indicate the average hit-density for the chromosomes. The purple lines above and below indicate hit densities of 2X and 1/2X the chromosome average, respectively. On chromosome 5, an example of a region of increased sequence density is highlighted. Sequence gaps in the human reference sequence are indicated by dark grey regions. Chromosomal banding pattern is indicated by light grey regions.

DNA offers clues to primitive humans

By Clive Cookson in Chicago

Published: February 13 2009 03:19 | Last updated: February 13 2009 03:19

In a spectacular celebration of Charles Darwin’s 200th birthday on Thursday, European scientists announced that they had completed a draft version of the Neanderthal genome.

Svante Pääbo and colleagues at the Max Planck Institute for Evolutionary Anthropology in Leipzig extracted enough DNA from fossilised bones to read about 60 per cent of the Neanderthal genetic sequence.

Neanderthals were the human species most closely related to people alive today; they inhabited Europe and western Asia until about 30,000 years ago. Comparison of Neanderthal genes with those of modern people and apes will illuminate the process of human evolution first proposed by Darwin.

Although analysis of the Neanderthal genome has only just begun, Dr Pääbo gave some early conclusions by video link from Germany to the American Association for the Advancement of Science meeting in Chicago.
Metagenomic environmental DNA sequencing* of seawater samples from the Sargasso Sea

*Note: done using "classic" Sanger shotgun sequencing approach.

Fig. 1. MODIS-Aqua satellite image of ocean chlorophyll in the Sargasso Sea grid about the BATS site from 22 February 2003. The station locations are overlain with their respective identifications. Note the elevated levels of chlorophyll (green color shades) around station 3, which are not present around stations 11 and 13.

Some sequences could be identified and mapped using BLAST onto the genomes of known marine bacteria (see below), other sequences were apparently from novel species.

Number of ORFs generated by genome sequencing projects (red: bacteria, orange: eukaryotic) and metagenomics projects (light green: microbial, dark green: viral). Data were taken from the GOLD database [65*].
Distribution of bacterial genera in human gut

*Individual & population differences*

*Figure 2.* Compositional view of human intestinal microbiomes. A compositional view of microbiomes based on the taxonomic assignment of protein-coding genes is shown. The stacked bars represent the compositions of each sample estimated from the results of BLASTP analysis with a 90% threshold identity. ‘Others’ includes the genera whose proportions were less than 1% in any of the samples. The data for the fecal samples from two American adults (‘Sub. 7’ and ‘Sub. 8’) are also shown.

Differences in human gut “microbiomes” are associated with obesity


Figure 1 | Correlation between body-weight loss and gut microbial ecology. a, Clustering of 16S ribosomal RNA gene sequence libraries of faecal microbiota for each person (in different colours) and time point in diet therapy (T0, baseline; T1, 12 weeks; T2, 26 weeks; T3, 52 weeks) in the two diet-treatment groups (fat restricted, FAT-R; carbohydrate restricted, CARB-R), based on UniFrac analysis of the 18,348-sequence phylogenetic tree. b, Relative abundance of Bacteroidetes and Firmicutes. For each time point, values from all available samples were averaged (n was 11 or 12 per time point). Lean-subject controls include four stool samples from two people taken 1 year apart, plus three other stool samples. Mean values ± s.e. are plotted. c, Change in relative abundance of Bacteroidetes in subjects with weight loss above a threshold of 2% weight loss for the CARB-R diet and 6% for the FAT-R diet.
<table>
<thead>
<tr>
<th>Disease or disorder</th>
<th>Association of gut microbiota with disease</th>
<th>Evidence of gut microbiota-targeted therapy</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronic peptic ulcer</td>
<td><em>Helicobacter pylori</em> infection is the pathogenic key to the development of most chronic peptic ulcers</td>
<td>Regimen of <em>H. pylori</em> eradication with antibiotics and proton-pump inhibitory agents</td>
<td>26,57</td>
</tr>
<tr>
<td>Antibiotic-associated diarrhoea</td>
<td>The suppression of antibiotic-sensitive bacteria and overgrowth of antibiotic-resistant species lead to intestinal dysfunction</td>
<td>Treatment with probiotics, such as the yeast <em>Saccharomyces boulardii</em>, together with antibiotics is effective in the prevention of antibiotic-associated diarrhoea</td>
<td>58,59</td>
</tr>
<tr>
<td>Ulcerative colitis</td>
<td>Abnormal immune response to commensal bacteria, and increased numbers of intestinal microorganisms, but reduced numbers of protective bacteria such as <em>Lactobacilli</em> and <em>Bifidobacteria</em></td>
<td>Short-term benefits were observed with antibiotic or probiotic/synbiotic therapy</td>
<td>42,60</td>
</tr>
<tr>
<td>Crohn's disease</td>
<td>Inadequate clearance of ingested microorganisms by dysfunctional intestinal macrophages (hypothesized mechanism)</td>
<td>Reinstating the balance of intestinal microflora with probiotics, prebiotics and/or antibiotics, such as the non-absorbable antibiotic rifaximin</td>
<td>61,62</td>
</tr>
<tr>
<td>Obesity</td>
<td>The relative abundance of the two predominant bacterial divisions, the <em>Bacteroidetes</em> and the <em>Firmicutes</em>, affect the efficiency of energy harvest from diet</td>
<td>It is suggested that manipulation of the commensal microbial composition could be a novel therapeutic approach for obesity</td>
<td>5,6</td>
</tr>
<tr>
<td>Diabetes</td>
<td>No gut microbiota-related mechanism is established, but it appears that diabetes is associated with the gut microbiota</td>
<td>Oral administration of probiotics shows a significant anti-diabetic effect in diabetic models</td>
<td>34–38</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>Conversion of dietary procarcinogens into DNA-damaging agents or generation of carcinogens by particular commensal bacteria are thought to be certain causes of colorectal cancer</td>
<td>Reduced prevalence of colon cancer was observed in interleukin 10 knockout mice by probiotic <em>Lactobacilli</em> administration, and strong antitumour activity was achieved by <em>Bifidobacterium longum</em> therapy in vivo</td>
<td>63,64</td>
</tr>
<tr>
<td>Idiopathic parkinsonism</td>
<td>Partial involvement of microorganisms such as <em>H. pylori</em></td>
<td><em>H. pylori</em> is important in the aetiology/pathogenesis of idiopathic parkinsonism and useful for disease categorization and subsequent treatment</td>
<td>65–67</td>
</tr>
</tbody>
</table>

Figure 2 | Schematic view of gut microbiota-targeted therapeutic strategy. Future therapeutic strategies for complex diseases will include the treatment of the gut microbiome, that is, 'drug the extended genome' strategy. The human microbiome may possess many potential targets, a number greater than the 3,000 targets from the human genome. As the gut microbiota affects metabolic phenotypes and enzyme induction states, which ultimately affects the outcome of interventions, the microbiome can be manipulated to alter interventional outcomes. Combinations of antibiotics, probiotics and prebiotics could be used in a gut microbiota-targeted therapy regimen to regulate the microbiome and, consequently, restore the homeostasis of gut ecology in the host. Such a therapeutic approach can be monitored and evaluated using ‘omics’ platform technologies, such as metagenomics and metabolomics, by capturing the holistic and dynamic biochemical variations associated with pathophysiological conditions of the host. The integration of metagenomic and metabolomic data into pharmacological and clinical results should be fulfilled to obtain comprehensive diagnostic and prognostic knowledge regarding the complex diseases.