REVEAL THE TRUE BIOLOGY

Technologies to Capture Biological Differences Among Individuals
Technologies Emerging to Observe the Functioning of Single Molecules in Real Time

- Replicating DNA
- Transcribing Genes (RNA)
- Translating RNA into Proteins

Complex Behavior
These technologies are enabling scoring of very large-scale, high-dimensional data on individuals for low cost.

- Modified and unmodified DNA
- Modified and unmodified coding and non-coding RNA
- Phosphorylated and unphosphorylated proteins
- Metabolites
DNA variation is just one dimension among many that define living systems

(■ - DNA, □ - RNA, △ - Protein, ▽ - Metabolite)
Single molecule real time observation system
Building a new, super high resolution microscope

Confinement to $20 \times 10^{-21}$ liters

Science, Vol 299, Jan 31 2003, pp682-686
Multiplexing
Validation of ITD mutations in FLT3 as a therapeutic target in human acute myeloid leukaemia

Catherine C. Smith, Qi Wang, Chen-Shan Chin, Sara Salerno, Lauren E. Damon, Mark J. Levis, Alexander E. Perl, Kevin J. Travers, Susana Wang, Jeremy P. Hunt, Patrick P. Zarrinkar, Eric E. Schadt, Andrew Kasarskis, John Kuriyan & Neil P. Shah

Affiliations | Contributions | Corresponding author

Received 25 August 2011 | Accepted 05 March 2012 | Published online 15 April 2012

Effective targeted cancer therapeutic development depends upon distinguishing disease-associated ‘driver’ mutations, which have causative roles in malignancy pathogenesis, from ‘passenger’ mutations, which are dispensable for cancer initiation and maintenance. Translational studies of clinically active targeted therapeutics can definitively discriminate driver from passenger.
Acute Myeloid Leukemia (AML)

- Aggressive disease in which too many myeloblasts (immature white blood cells that are not lymphoblasts) are found in the bone marrow and blood

- Estimated 12,950 people (6,830 men and 6,120 women) were diagnosed with and approximately 9,050 men and women died of AML in 2011\(^1\)

- Extremely poor prognosis

- No notable treatment improvements over the past several decades

- Ambit Biosciences recently developed a FLT3 kinase inhibitor - the AML chemotherapeutic AC220 (quizartinib)

FLT3-ITD is Important in AML

• Activating in-tandem duplication (ITD) mutations in FLT3 (FLT3-ITD) are detected in approximately 20% of AML patients and associated with a poor prognosis

• Potential resistance mutations located > 800 bp away from ITD region
<table>
<thead>
<tr>
<th></th>
<th>Detect ITD</th>
<th>Phase Distant Mutations</th>
<th>Low Frequency Mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short read (ILMN, Ion Torrent®, etc.)</td>
<td>☠️</td>
<td>☠️</td>
<td></td>
</tr>
<tr>
<td>Medium read (454®, 454 Jr)</td>
<td></td>
<td>☠️</td>
<td></td>
</tr>
<tr>
<td>Sanger/CE</td>
<td></td>
<td></td>
<td>☠️</td>
</tr>
</tbody>
</table>

Only way to do identify secondary mutations in ITD+ cells now is with Sanger and it requires cloning and two sequencing reactions.
PacBio® Long Read Technology Permits Deep Sequencing of the FLT3 Locus by Eliminating the Need for Cloning

Single molecule long reads show that a secondary mutation is on the same molecule as the ITD in one pass without cloning.
<table>
<thead>
<tr>
<th>Subject Number</th>
<th>Mutation</th>
<th>Native Codon</th>
<th>Alternative Codon</th>
<th>Pre-Treatment</th>
<th>Relapse</th>
<th>Normal Control #1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Observed</td>
<td>Total Number</td>
<td>Observed</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Alternative</td>
<td>of ITD+</td>
<td>Alternative</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Codon Frequency in ITD+ Sequences</td>
<td>Sequences Sampled</td>
<td>Codon Frequency in ITD+ Sequences</td>
</tr>
<tr>
<td>1009-003</td>
<td>D835Y</td>
<td>GAT</td>
<td>TAT</td>
<td>0.21%</td>
<td>482</td>
<td>8.4%</td>
</tr>
<tr>
<td></td>
<td>D835V</td>
<td>GAT</td>
<td>GTT</td>
<td>0.00%</td>
<td>482</td>
<td>3.3%</td>
</tr>
<tr>
<td></td>
<td>D835F</td>
<td>GAT</td>
<td>TTT</td>
<td>0.00%</td>
<td>482</td>
<td>10.2%</td>
</tr>
<tr>
<td>1011-006</td>
<td>D835Y</td>
<td>GAT</td>
<td>TAT</td>
<td>0.00%</td>
<td>196</td>
<td>41.0%</td>
</tr>
<tr>
<td>1011-007</td>
<td>F691L</td>
<td>TTT</td>
<td>TTG</td>
<td>0.18%</td>
<td>561</td>
<td>6.2%</td>
</tr>
<tr>
<td></td>
<td>D835Y</td>
<td>GAT</td>
<td>TAT</td>
<td>0.00%</td>
<td>930</td>
<td>3.0%</td>
</tr>
<tr>
<td></td>
<td>D835V</td>
<td>GAT</td>
<td>GTT</td>
<td>0.43%</td>
<td>930</td>
<td>29.6%</td>
</tr>
<tr>
<td>1005-004</td>
<td>F691L</td>
<td>TTT</td>
<td>TTG</td>
<td>0.00%</td>
<td>496</td>
<td>29.6%</td>
</tr>
<tr>
<td>1005-006</td>
<td>D835Y</td>
<td>GAT</td>
<td>TAT</td>
<td>0.00%</td>
<td>171</td>
<td>39.5%</td>
</tr>
<tr>
<td></td>
<td>D835F</td>
<td>GAT</td>
<td>TTT</td>
<td>0.00%</td>
<td>171</td>
<td>2.7%</td>
</tr>
<tr>
<td>1005-007</td>
<td>D835Y</td>
<td>GAT</td>
<td>TAT</td>
<td>0.00%</td>
<td>57</td>
<td>4.0%</td>
</tr>
<tr>
<td></td>
<td>D835V</td>
<td>GAT</td>
<td>GTT</td>
<td>0.00%</td>
<td>57</td>
<td>47.4%</td>
</tr>
<tr>
<td>1005-009</td>
<td>D835Y</td>
<td>GAT</td>
<td>TAT</td>
<td>0.00%</td>
<td>19</td>
<td>50.6%</td>
</tr>
<tr>
<td>1005-010</td>
<td>F691L</td>
<td>TTT</td>
<td>TTG</td>
<td>0.00%</td>
<td>387</td>
<td>25.3%</td>
</tr>
</tbody>
</table>
Accurate single molecule long reads span the ITD region and kinase domain

- Unique ability to sequence through repeats, phase SNPs, and detect low frequency mutations
- Critical to understanding AML and FLT3

PacBio’s unique capabilities continue to unlock research problems previously unanswered or misunderstood

- Prior research incorrectly suggested FLT3-ITD was NOT a likely viable drug target

UCSF team now identifying additional promising candidate compounds that could target these particular drug-resistant mutations
Assaying internal molecular states and microenvironments is critically important, but what about the external macroenvironment?

- Your own DNA
- Your microbiome DNA
- DNA from bugs in the environment
E. Coli Outbreak in Germany (June 2011)

Farmers dumped produce at the German consulate in Valencia to protest claims that the E. coli outbreak started in Spain.

<table>
<thead>
<tr>
<th>Country</th>
<th>HUS Cases</th>
<th>HUS Deaths</th>
<th>EHEC Cases</th>
<th>EHEC Deaths</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany</td>
<td>814</td>
<td>27</td>
<td>2773</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>856</td>
<td>28</td>
<td>2841</td>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>
Rapid identification of the German E. Coli outbreak strain and understanding its increased virulence

- Addressing:
  - How did this strain become so virulent: Enteroaggregative E. coli or Enterohemorrhagic E. coli
  - How best to treat

Origins of the E. coli Strain Causing an Outbreak of Hemolytic–Uremic Syndrome in Germany
Phylogenetic relatedness among 55 E. coli strains
Leading to a view of how these bugs are related to one another.
Acquisition of the stx-phage, other virulence factors, and antibiotic resistance plasmids via horizontal gene transfer most likely led to the increased virulence with effects on treatment as well.
Can we dig more deeply into those data, going beyond the basic A, G, C, and T building blocks?

Inter-pulse Distance (IPD)

Clark et al (2011) Nucleic Acids Research
Recent studies have well demonstrated the ability of single molecule real-time sequencing to detect all known base modifications (base mods).

**Epigenetic markers:**
- 5-mC
- 5-hmC

**Identity markers:**
- (host-pathogen interactions)
- 4-mC
- 6-mA
- Glucosyl-5-hmC
- Base J

**DNA damage:**
- 8-oxoG
- N²-BPDE-G
- Thymidine dimer

**Other:**
- dU
- 5Br-dU
- Ribonucleosides
So, we reanalyzed the German E. coli outbreak strain data to identify kinetic variation events: Big signature identified.
Making sense of the kinetic signature, searching for patterns

- Let’s focus just on the A residue detections:
- ~2.2M A sites tested
- Roughly 50,000 A sites detected
  - GATC: 38,821
Given all of this “A” activity, we partnered with NEB to explore the methylase world in these bugs.
Methyltransferase predicted at position 2,273,609

**IPD ratio circos plot:**

**Log-likelihood analysis:**

**Additional information:**
- Predicted specificity from bioinformatics: CTG(m6A)TG
- Activity: Unknown
- Predicted specificity from experimentation: Not determined

**Consensus:** ACC(m6A)CC

Based on top 22 sites (FDR <=0.001)
Methyltransferase predicted at position 5,184,126

Additional information:
- Predicted specificity from bioinformatics: CTGC(m6A)G
- Activity: Yes
- Predicted specificity from experimentation: CTGC(m6A)G
- Evidence 1: Resistance to PstI cleavage (CTGCAG)

Consensus: CTGC(m6A)G

Based on top 28 sites (IPD ratio)
A complete mapping of the methyltransferases in the E. coli outbreak strain led to the identification of an outbreak specific MTase.

The methyltransferase and restriction enzyme here were found to target CTGCAG motifs and SPECIFIC TO THE OUTBREAK STRAIN.
Strong expression differences between the outbreak strain and highly related strains without the CTGCAG methyltransferase; Do the CTGCAG modifications play a role?

56 CTGCAG Genes
Pathway Enrichment

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Adjusted p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell projection</td>
<td>6.01e-7</td>
</tr>
<tr>
<td>Flagellum</td>
<td>8.79e-5</td>
</tr>
<tr>
<td>Translation</td>
<td>4.53e-4</td>
</tr>
<tr>
<td>Pilus</td>
<td>1.13e-4</td>
</tr>
</tbody>
</table>

847/5380 = 0.15 (15% genes up regulated)

56/117 = 0.32 (32% of CTGCAG genes up regulated)

CTGCAG Genes > 2-fold enriched for being up regulated
(Fisher Exact Test P = 1.11e-16)
Action of these epigenetic changes induced by the virus

- Infected rabbits, an animal model for ecoli infection in human, with

- All animals infected with the outbreak strain that infected humans got sick

- None of the animals infected with the modified version of the outbreak strain of E. coli got sick
Summarizing our observations

Potential to Affect Virulence
PacBio RS Quick Turnaround Time Provides a Path for Real Time Pathogen Surveillance
Diverse Range of Specific and Non-Specific Viruses Targeted

- Specific targets included (but not limited to):
  - Human Rhinovirus
  - Human Respiratory Syncytial Virus
  - Human Metapneumovirus
  - Human Coronavirus
  - Human Parainfluenza Virus
  - Influenza A viruses
  - Paramyxovirinae
  - Respirovirus
  - Rubulavirus
  - Pneumovirinae
  - Parechovirus
  - Enterovirus
  - Rotavirus
  - Caliciviruses
  - Astrovirus
  - Adenovirus
  - Human Herpesvirus 5
  - Human Herpesvirus 3

- Rhinovirus
  - Common Cold
- Herpes viruses
  - HHV5 Mononucleosis
  - HHV3 Chickenpox/Shingles
- Astrovirus
  - Gastroenteritis (in children)
- Adenovirus
  - Upper respiratory tract infection
- Influenza A
  - Flu
- Parainfluenza virus
  - Lower respiratory tract infection
- Rotavirus
  - Severe diarrhoea (“stomach flu”)
Enabling a More Comprehensive Understanding of Personal Genomes Via the Construction of a Disease Weather Map

Environmental Swabs

- Train Stations (BART, Cal Train)
- Airports (SFO, SJC, OAK)
- Emergency Rooms
- Sewage Treatment Plants
- Universities

Prepare Samples
- Convert RNA to cDNA
- Purify DNA
- Fragment DNA
- Add 3’ A Tail
- Ligate Adapters

Sequence

Assemble, Identify, and Quantify Genomes

Project findings onto map (detections and trends)
Starting with low-resolution maps as a proof of concept
Detection of Viral Pathogens from Inanimate Surfaces

- High-traffic surfaces at Pacific Biosciences:
  - Front door handle
  - Common laboratory bench top
  - Break room refrigerator door handle
  - Slide projector remote control
  - Lavatory toilet flush handle
  - Lavatory door handle
  - Laboratory telephone handle
  - Cubicle desk surface
  - Money

Sampled every week for a period of one month
Anonymous donors, submitting nasopharyngeal swabs every two weeks over ~2.5 months
Sequencing Read Distribution from Sewage

- Lactococcus phage Q54 1
- Lactococcus phage BK5-T 1
- Lactococcus phage TP901-1 1
- Lactococcus phage 1706 2
- Pseudomonas phage LIT1 1
- Lactococcus phage KSY1 2
- Pleurotus ostreatus virus 1 1
- Melon necrotic spot virus 9
- Carnation mottle virus 2
- Enterobacteria phage GA 2
- Acinetobacter phage AP205 2
- Enterobacteria phage Qbeta 1
- Human enterovirus B 1
- Porcine kobavirus swine/K-30-HUN/200
- Aichi virus 4
- Southern bean mosaic virus 1
- Ryegrass mottle virus 1
- Rubus chlorotic mottle virus 1
- Eggplant mosaic virus 1
- Turnip yellow mosaic virus 1
- Nemesia ring necrosis virus 1
- Potato virus S 1
- Clover yellow mosaic virus 7
- Potato virus X 1
- White clover mosaic virus 8
- Pepino mosaic virus 1
- Garlic virus D 1
- Garlic virus A 2
- Garlic virus C 1
- Cucumber green mottle mosaic virus 1
- Tobacco mosaic virus 8
- Pepper mild mottle virus 12
- Tomato mosaic virus 12
- Paprika mild mottle virus 2
- Bell pepper mottle tobamovirus 3
- Cucumber mottle virus 1
- Rehmannia mosaic virus 1
- Brome mosaic virus 1
- Cucumovirus 1
- Astrovirus MLB1 1
- Escherichia phage K1H 2

- bacteria
- eukaryota
- ssRNA viruses
- dsDNA viruses
- archaea

Common virus causing common cold; hand foot mouth disease; aseptic meningitis
Oyster-associated non-bacterial gastroenteritis
Common virus infecting tobacco plants
Common in field-grown bell, hot, and ornamental pepper species
Common virus infecting tomato and pepper plants
Wash Your Hands!!!

Metagenomic Hit Distribution for Inanimate Surfaces and Nasal Swabs:

**Surfaces:**
- fridge door
- desk surface
- toilet flush handle

**Nasopharyngeal swabs:**
Detection of Viral Pathogens – Influenza on Surfaces

<table>
<thead>
<tr>
<th>Person</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Sampling period

- **front door 1**
- **fridge door 2**
- **projector remote 1**
- **projector remote 2**
- **restroom door 1**
- **restroom door 2**
- **toilet flush handle 1**
- **desk 2**
- **desk 3**
- **desk 8**
- **$1$**
- **$1$**

**Viruses:

- **H1N1**
- **H1N2**
- **H2N2**
- **H1N1/H3N2**
- **H1N1/H2N2**
- **H1N1/H3N2**
- **H1N1/H5N1**
- **H1N1**
- **H3N2/H1N1**
- **H1N1/H3N2**

**Colors:**
- **Green:** sampled, no hits
- **Red:** influenza
- **Orange:** human respiratory syncytial virus
- **Blue:** human metapneumovirus
- **Pink:** paramyxovirus
- **Black:** not sampled

**Controls:**
- **Positive control**
- **Negative control**
The ultimate is to get to a high resolution map: What’s going on in your neighborhood?
Coming soon to a neighborhood near you?

Mom, can I go over to Billy’s house after school?
My god, she’s going to check disease weather map

Just a minute, dear...
Give me the street-level view of Billy’s house please.
Disease Weather Map: Street Level View

Long-term Metagenomic Report
- Viral pathogens (1yr)
  - No diarrheal viruses
  - Cold viruses detected 20 days
  - Influenza detected 3 days
- Disease susceptibility profile
  - Bacterial composition indicates pro-athero, pro-diabetes environment
  - Food viruses detected indicate high-fat Western diet, minimal vegetables

Daily Metagenomic Report
- Current pathogen composition within 95% CI of year mean
- No cold, flu, etc. viruses detected
Disease weather map saves the day

Okay, Johnny, you can go over to Billy’s house but come home for dinner.

Uhhhhghggg ....
Acknowledgements

Pacbio

Steve Turner
Andrew Kasarskis
NGAT Group
Jonas Korlach
Tyson Clark
Khai Luong
Kristi Spittle

Robert Sebra
Seven Lin
Eric Olivarez
Kevin Travers

Informatics
Jason Chin
Ali Bashir
Aaron Klammer
Michael Brown

Mount Sinai

Andrew Kasarskis
Jun Zhu
Bin Zhang
Michael Linderman
Gaurav Pandey
Rui Chang
Shaun Purcell

Partners

Sage Bionetworks
Stephen Friend
Et al.

UCSF
Neil Shah
Cathy Smith
W. Michael Korn
Joe De Risi Lab

UCLA
Lusis Lab

Harvard
Matt Waldor lab
John Mekalanos lab
U of Maryland
Dave Rasko lab
U of Wisconsin
Attie Lab

NEB
Rich Roberts
Tamas Vincze
U of Minnesota
Gang Fang
Vipin Kumar