Building a bridge from discovery to medicine
FIMM: International research institute in Meilahti

"Science"
UNIVERSITY OF HELSINKI
Host

"Health & Medicine"
NATIONAL INSTITUTE FOR HEALTH AND WELFARE

"Technology"
VTT

Nordic EMBL partnership

Public-private funding partnership at the FIMM launch (2007-2013)

Orion-Farmos Research Foundation
Profiles of the FIMM group leaders

**Human genomics**
- Autoimmunity
- Breast cancer
- Viral infections

**Medical systems biology**
- Puberty
- Biobanks

**Computational biomedicine**
- Prevention
- Networks
- Statistics
- Diagnostics
- Imaging

**Translation**
- Lung cancer
- Drugs
- Pers. medicine
- Leukemia

4 computational
6 genomics
6 systems biology
6 translational
Genomics Empowering Personalized Medicine

2010-2011:

National Biobanks of Finland (4% of the population, 200,000 DNA samples)
- 40,000 genotyped
- Half a million genetic variants
- 100 Finns fully sequenced
⇒ 8 M imputed genotypes

⇒ 100 publications on GWAS

YEARS 2012-2013:
- 7000 Finnish genomes sequenced
- Complete list of genomic variants
- All imputed variants, incl. rare ones from 200,000 samples

- Links to systematic health data:
  Detailed registry links and long-term follow-up:
  - ALL diseases, medication use
  - Interactions
Sisu: Sequencing Initiative SUomi to promote implementation of personalized medicine in the health care (National grand challenge)

THL Biobanks (>200,000 samples)

Genomics (<1000 € genomes)

Registries:
- Diseases
- Medication
- Outcome

Data storage

Modeling

Implementation in the health care

Comprehensive health follow-up
Cost per Megabase of DNA Sequence

From $10,000 to $0,1 /Mb cost in 10 years

10% deflation per month
Figure 2 shows the trend in storage capacity needed to store the volume of biological data at EMBL-EBI (in Terabytes; a terabyte is a million million bytes). This trend is expected not only to persist but to become steeper still, posing a serious challenge to existing bioinformatics infrastructures in Europe.
“Extending” the FIMM computing cluster by cloud computing arrangement with CSC

- 972 cores in operation
- ~23 328 CPU hours (~3 CPU years) / day

208 cores @FIMM
768 cores @CSC (on dedicated hardware)
The next solution: large-scale data- and computational centers (CSC - ELIXIR)

Kajaani Data Center

CSC – IT Center for Science Ltd. is building one of the most eco-efficient data centers in the world. The location is Kajaani, in Northern Finland. The Kajaani Data Center is a proven solution based on technology, modern, reliable infrastructure and ecological efficiency for data needs in research and development in public and private sector. The Funet Network (Finnish University and Research Network) ensures excellent networking capabilities around the world.

Unique Synergy Benefits
The Data Center is being built in an old paper mill by the River Kajaani. The CSC Data Center is 8000 m² in total, of which CSC will first occupy 4000 m² with an option for another 4000 m² for future growth. The Data Center project has started in 2010 and is estimated to be up and running by the early 2012. The infrastructure is already in place.
Towards individualized medicine

Cancer genomics & next-gen sequencing

Bioinformatics for biomarkers and diagnostics

Molecular pathology, imaging & informatics

From translational genomics towards individualized management and treatment of patients
To obtain a comprehensive description of genomic, transcriptomic and epigenomic changes in 50 different tumor types and/or subtypes which are of clinical and societal importance across the globe.

Systematic studies of over 25,000 cancer samples at the genomic, epigenomic, and transcriptomic levels
Cancer genomics: Direct therapeutic possibilities illustrating the impact

**PLX4032**

**BRAF**

The NEW ENGLAND JOURNAL of MEDICINE

**Inhibition of Mutated, Activated BRAF in Metastatic Melanoma**

Keith T. Flaherty, M.D., Igor Puzanov, M.D., Kevin B. Kim, M.D., Antoni Ribas, M.D., Grant A. McArthur, M.B., B.S., Ph.D., Jeffrey A. Sosman, M.D., Peter J. O'Dwyer, M.D., Richard J. Lee, M.D., Ph.D., Joseph F. Grippo, Ph.D., Keith Nolop, M.D., and Paul B. Chapman, M.D.

**Plexxikon Drug Shows Stunning 81 Percent Response in Melanoma**

By Donna Young

Washington Editor Bioworld Today
Identification of the transforming *EML4–ALK* fusion gene in non-small-cell lung cancer

Manabu Soda\(^1,2\), Young Lim Choi\(^1\), Munehiro Enomoto\(^1,2\), Shuji Takada\(^1\), Yoshihiro Yamashita\(^1\), Shunpei Ishikawa\(^5\), Shin-ichiro Fujiwara\(^1\), Hideki Watanabe\(^1\), Kentaro Kurashina\(^1\), Hisashi Hatanaka\(^1\), Masashi Bando\(^2\), Shoji Ohno\(^2\), Yuichi Ishikawa\(^6\), Hiroyuki Aburatani\(^5,7\), Toshiro Niki\(^3\), Yasunori Sohara\(^4\), Yukihiro Sugiyama\(^2\) & Hiroyuki Mano\(^1,7\)

*EML4-ALK* fusion oncogene has been reported in approximately 4% of NSCLC.

Clinical activity of the oral ALK inhibitor PF-02341066 (Crizotinib) in ALK-positive patients with non-small cell lung cancer (NSCLC). J Clin Oncol 28:18s, 2010 (suppl; abstr 3):
Pfizer Wins Approval For Xalkori, Lung Cancer Drug That Heralds Age Of Expensive, Personalized Medicines

Pfizer announced that the Food and Drug Administration has approved Xalkori, generically known as crizotinib, a new medicine that can have a dramatic impact in a small minority of lung cancer patients.

A companion diagnostic test — the Vysis ALK Break Apart FISH Probe Kit (Abbott Molecular, Inc.) — was approved concurrently. It is designed to detect rearrangements of the ALK gene, which are found in about 4% to 5% of patients with NSCLC.
Like many targeted therapies, crizotinib comes with a hefty price tag. According to Pfizer, the drug will cost $9,600 per month, or $115,000 per year.

Crizotinib is only effective in about 5% of patients whose tumors have a mutation in a gene called ALK. The real cost of the drug is $9,600 plus 25 ALK tests, because that’s how many patients will need to be screened for one to actually get the drug

$250 \Rightarrow$1200 USD $\times$ 25 = $30,000 for diagnosis
Challenges in implementing cancer genomics in the clinic

Diagnostic testing of individual mutations one at a time is not cost-effective and should eventually be replaced by comprehensive mutation profiling (Next-Gen Sequencing)

NGS will not do it all: Most tumors do not display “directly actionable” mutations

For many that do, there are no drugs on the market

Cancer cells become resistant to single targeted therapeutics and relapse after initial response, combinations needed
From translational genomics towards individualized management and treatment of patients

Cancer genomics & next-gen sequencing

1) Exome seq in LGL leukemia
2) Exome seq for copy number changes in cancer
3) RNAseq
Example 1) Recurrent missense mutations in the STAT3 gene in LGL leukemia provide insights to pathogenetic mechanisms and suggest potential diagnostic and therapeutic applications

T-cell large granular lymphocyte (T-LGL) leukemia is an orphan lymphoproliferative disorder characterized by expansion of mature, clonal CD3+CD8+ cytotoxic T lymphocytes

1) T-LGL exome seq
Example 2) Exome sequencing data applied to assess gene copy number changes

Female vs male – RPKM values

Anna-Maija Sulonen
Henrik Edgren
Pekka Ellonen
Henrikki Almusa
Olli Kallioniemi
Janna Saarela
Sequencing of cancer transcriptomes (RNA-seq)

- Quantitation of gene expression
- Quantitation of exon expression
- Exploration of alternative splicing
- Identification of novel exons
- RNA editing
- Transcriptional read-through
- Detecting expressed mutations
- Detecting non-coding RNAs
- Detecting UTRs
- Fusion-gene detection

Sophisticated bioinformatic pipelines to process and mine the data

Example: extended 5’UTRs in the androgen receptor subject to miRNA regulation

Östling et al., Cancer Res, 2011
28 fusion gene candidates in breast cancer, 27 (96%) validated by RT-PCR, Sanger seq

24/27 one partner associated with a copy no. transition

Most in-frame

Exclusive expr. of the fusion (activation)

Many involve known fusion partners in other cancers (ACACA, RARA, NOTCH1 and NUP214)

RNAi validation

Biomarker potential
From translational genomics towards individualized management and treatment of patients

Cancer genomics & next-gen sequencing

Bioinformatics for biomarkers and diagnostics

Molecular pathology, imaging & informatics

Towards individualized medicine
Bioinformatics: creating the hairball problems
Computational genome-scale molecular pathology

"Virtual biobank" with 20,000 tissue samples

- 43 normal tissue types
- 227 cell types
- 68 different cancer types
- 105 cancer subtypes

Clinical data

Affymetrix microarray data on expression of 22,000 genes

www.genesapiens.org

In vivo role of genes in health & disease

Biomarkers + targets

Interpretation of new molecular profiling data

Molecular biology data on each gene

Kilpinen et al. Genome Biology, 2008
# IST-Online: Reference database of Affy microarray analyses of cancer (20 064 tissue samples)

<table>
<thead>
<tr>
<th>Type of Cancer</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL</td>
<td>2033</td>
</tr>
<tr>
<td>AML</td>
<td>1298</td>
</tr>
<tr>
<td>Adrenal gland cancer</td>
<td>36</td>
</tr>
<tr>
<td>Bladder cancer</td>
<td>200</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>3114</td>
</tr>
<tr>
<td>CLL</td>
<td>663</td>
</tr>
<tr>
<td>CML</td>
<td>103</td>
</tr>
<tr>
<td>Cervical cancer</td>
<td>95</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>1024</td>
</tr>
<tr>
<td>Glioma</td>
<td>652</td>
</tr>
<tr>
<td>Head and neck cancer</td>
<td>110</td>
</tr>
<tr>
<td>Kidney cancer</td>
<td>399</td>
</tr>
<tr>
<td>Liver cancer</td>
<td>143</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>1493</td>
</tr>
<tr>
<td>Lung carcinoid</td>
<td>40</td>
</tr>
<tr>
<td>Lymphoma</td>
<td>703</td>
</tr>
<tr>
<td>Melanoma</td>
<td>208</td>
</tr>
<tr>
<td>Mesothelioma</td>
<td>46</td>
</tr>
<tr>
<td>Myeloma</td>
<td>129</td>
</tr>
<tr>
<td>Neuroblastoma</td>
<td>285</td>
</tr>
<tr>
<td>Other GI cancer</td>
<td>372</td>
</tr>
<tr>
<td>Other neuroectoderm.</td>
<td>121</td>
</tr>
<tr>
<td>Other urogenital tumor</td>
<td>41</td>
</tr>
<tr>
<td>Ovarian cancer</td>
<td>635</td>
</tr>
<tr>
<td>Pancreatic cancer</td>
<td>78</td>
</tr>
<tr>
<td>Peritoneal cancer</td>
<td>47</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>460</td>
</tr>
<tr>
<td>Sarcoma</td>
<td>370</td>
</tr>
<tr>
<td>Skin, squamous</td>
<td>30</td>
</tr>
<tr>
<td>Testicular cancer</td>
<td>206</td>
</tr>
<tr>
<td>Thyroid cancer</td>
<td>70</td>
</tr>
<tr>
<td>Uterine cancer</td>
<td>176</td>
</tr>
</tbody>
</table>

*19/01/2012*
KLK3 (PSA) profile across all normal and tumor tissues (n = 20 064)
ACPP (PAP) gene expression profile across all normal and tumor tissues (n = 20,064)
Tumor/patient-centric view of cancers based on whole-genome gene expression profiling.
Computational genome-scale molecular pathology: Analogy to BLAST analysis for DNA sequences

Global nucleotide sequence comparisons

The BLAST Search Algorithm
Computational genome-scale molecular pathology: Analogy (1) to BLAST analysis for DNA sequences

Global nucleotide sequence comparisons

The BLAST Search Algorithm

query word \((W = 3)\)

| Query: GSWVLTTGQSLALILNMCQKGIWLKIQPLMFHRRERILVEAFVEDLRQTIQEDL |
|---|---|---|---|---|---|---|---|---|
| POG 10 | PEG 15 | PRO 14 | PKG 14 | PNG 13 | PDD 13 | PRG 13 | PIG 13 |
| FOG 12 | POG 12 | PGG 12 | PGG 12 | PDD 13 | PDD 13 | PDD 13 | PDD 13 |
| 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 |

neighborhood words

neighborhood score threshold \((T = 13)\)

query: 325 SLALLMKCTGQSLALILNMCQKGIWLKIQPLMFHRRERILVEAFVEDLRQTIQEDL

High-scoring Segment Pair (HSP)

Global gene expression comparisons

Alignment of gene expression profiles from test samples against a reference database: New method for context-specific interpretation of microarray data


Sami Kilpinen et al.,

AGEP

Query sample

Gene Expression data

Reference database

- tissue 1
- tissue 2
- tissue 3
Computational genome-scale molecular pathology
AGEP: Alignment of gene expression profiles
(Kalle Ojala et al., Genomic Medicine, 2011)

Metastasis identified as originating from colorectal cancer

Cancer of the unknown primary (CUP)

88% classification accuracy in leave-one-out classification of primary tumors from a database of 5577 primary tumors representing 56 different tumor types

78% classification accuracy from external, diagnostically challenging tumor samples, including both metastases (n = 128) and primary, poorly differentiated tumors (n = 59).
From translational genomics towards individualized management and treatment of patients

Cancer genomics & next-gen sequencing

Bioinformatics for biomarkers and diagnostics

Molecular pathology, imaging & informatics

Towards individualized medicine
Biobanks and molecular phenotyping of cancers: towards future diagnostics

Johan Lundin
People in Lundin group

Nina Linder
Mikael Lundin
Juho Konsti
Riku Turkki
Tiina Lehtimäki
Hélène Rotkirch
Ville Ojansivu

Pathologists enthusiastic over new multitouch microscope

The 23rd European Congress of Pathology was held at Helsinki Exhibition & Convention Centre in the beginning of September. Thousands of pathologists gathered together to hear the latest news in the fields of research, technology and diagnostics.

Senior lecturer Johan Lundin from the Institute for Molecular Medicine Finland (FIMM) presented the new multitouch microscope, now adapted to a giant-sized display.
Cloud Computing Image Analysis for future cancer diagnostics

Image Capture

Tissue Sample Multi-color IHC

Microscopy scanner

Image Transfer

Internet

Reference database

Secure Compute cloud FIMM+CSC

FIMM

Clinical correlates

Image analysis

ECP 23/Johan Lundin / Biobanks

www.fimm.fi
Google Goggles
Use pictures to search the web.

Get Google Goggles
Android (1.6+ required)
Download from Android Market.

Send Goggles to Android phone

New! iPhone (iOS 4.0 required)
Download from the App Store.

Send Goggles to iPhone

Google Goggles in action

Click the icons below to see the different kinds of objects and places you can search for using Google Goggles.

Text  Landmarks  Books  Contact Info  Artwork  Wine  Logos
Masterworks of Art - Frida Kahlo and Diego Rivera

Frida Kahlo (1907 – 1954) was a legendary Mexican painter whose striking artworks reflected a lifetime of unbearable pain, ...
Tumor diagnostics with cloud computing in the future

1 diagnostic sample

Cloud knowledge source & computation

Reference database

Molecular pathology

Morphology, molecular & clinical data

Learning Adaptive Real-time

Diagnosis

Treatment

Recommendations

Accuracy estimate
From translational genomics towards individualized management and treatment of patients

Cancer genomics & next-gen sequencing

Bioinformatics for biomarkers and diagnostics

Molecular pathology, imaging & informatics

Towards individualized medicine
Facilitate individualized treatment of leukemia patients with clues from molecular profiling and drug sensitivity testing
Integrated biobanking, molecular profiling & drug sensitivity testing of adult acute leukemias

Kimmo Porkka and Finnish hematologists

Hematology Clinic

Sample Processing

Biobanking  Exome Seq  RNA Seq  Proteomics  Glycomics  Drug Testing

Data processing, analysis & mining

Impact on tx
Cancer pharmacopeia-wide drug sensitivity testing ex-vivo for individualized therapy

Detailed dose-response curves for all oncology drugs and many emerging cancer compounds for individual patient samples

What’s in the 240 collection?

- All conventional chemotherapeutics
- Tyrosine kinase-type inhibitors
  - Abl, Src, EGFR, FGFR, VEGFR, JAK, IGF1R, PDGFR, Met, ALK Kit, Flt3....
- S/T-type inhibitors
  - Aurora, PLK1, MEK, TTK, PDK1, Akt, Wee1, PKCs, Cdns, Chk1
- HDACi:s
- HSP90 inhibitors
- Bcl-2 inhibitors
- PI3K inhibitors
- mTOR inhibitors
- Survivin inhibitor
- Hh inhibitors
- γ-secretase inhibitors
- Farnesyltransferase inhibitor
- p53 activators
- PARP inhibitors
Implementing cancer drug sensitivity and resistance testing results & molecular profiling

Mika Kontro, Elonen Erkki, Hanna Koskela, Emma Andersson, Satu Mustjoki, Kimmo Porkka

Hematology Research Unit, Department of Medicine, Division of Hematology, Helsinki University Central Hospital, Helsinki, Finland

Caroline Heckman, Evgeny Kulesskiy, Tea Pemovska, Maxim Bespalov, Samuli Eldfors, Henrikki Almusa, Pekka Ellonen, Riikka Karjalainen, Disha Malani, Muntasir Mamun Majumder, John Patrick Mpindi, Astrid Murumägi, Jonathan Knowles, Maija Wolf, Laura Turunen, Janna Saarela, Krister Wennerberg

Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland
George Sledge, the ASCO president said that cancer doctors are entering an era of “genomic chaos,” a phrase that describes both the genetic instability & difficulties that incorporating rapidly advancing genetic technology into cancer care will bring.

The way clinical trials are run will need to change dramatically.

New kinds of electronic health records will need to be created to collect data, inform doctors instantly of new results.

Every oncologist will need to be “a clinical cancer biologist.”

Clinical trials in cancer should soon involve genetic testing for all patients, moving to using next-gen DNA-sequencing technology as quickly as possible.

Things will get “very, very complicated.”

“None of this will be easy,” Sledge said, “but all of it is necessary.”
Facilitating translation by implementing individualized molecular medicine in cancer treatment
Building a bridge from discovery to medicine