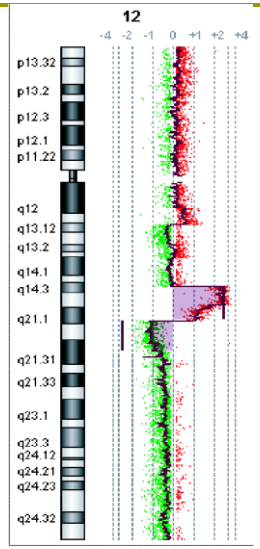
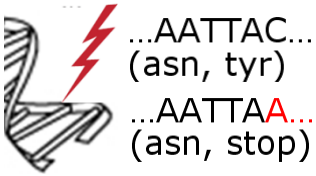
How Individual Variation and Treatment Strategies Affect Cancer Progression and Death?

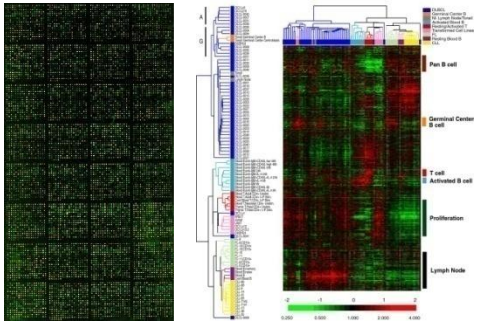
Sampsa Hautaniemi, DTech
Academy Research Fellow
Institute of Biomedicine
Genome-Scale Biology Research Program
Centre of Excellence in Cancer Genetics
Faculty of Medicine
University of Helsinki

Complex Diseases Require Data From Several Levels

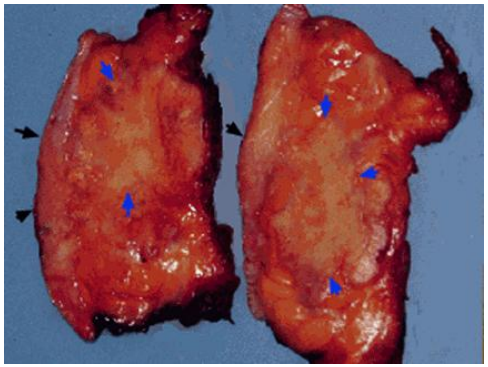


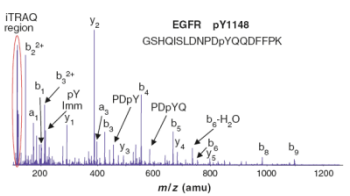


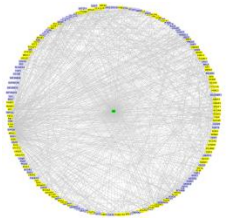
Genetics



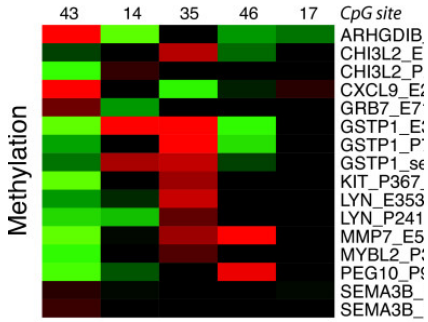
Transcriptome



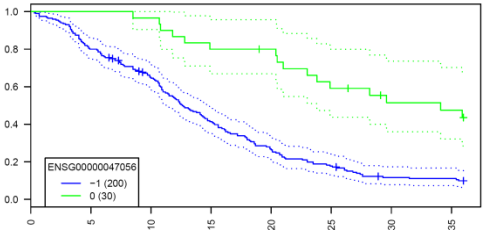




Proteomics



Epigenetics



Clinical

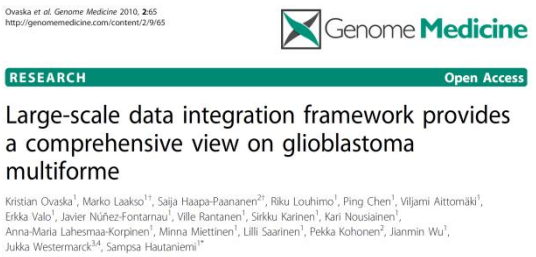
100 samples lead to ~200 million data points.

The Role of Bioinformatics in Biomedical Research

- Storing the data and computing power are the first (but relatively small) hurdles.
- Analysis of large-scale, heterogeneous data is much more challenging than individual genomics or proteomics data analysis.
  - It is a different matter to analyze a couple of tens of samples than hundreds or thousands samples.
- There is a need for computational infrastructure.
  - Writing an analysis program fast without proper infrastructure will lead to delays and errors in larger projects.

Anduril

- Anduril is a computational framework to integrate large-scale and heterogeneous **data**, **knowledge** in bio-databases and analysis **tools**.
- The main design principles are:
  - Modular pipeline analysis approach
  - Scalable
  - Open source, thorough documentation
    - http://csbi.ltdk.helsinki.fi/anduril
- Method written in any programming language executable from the command prompt can be included.
- Produces automatically the result PDF and website containing the results.



Glioblastoma Multiforme

- Glioblastoma multiforme (GBM) is one of the deadliest cancers.
- The Cancer Genome Atlas (TCGA) has published data from >500 GBM patients:
  - comparative genomic hybridization arrays
  - single nucleotide polymorphism arrays
  - exon and gene expression arrays
  - microRNA arrays
  - methylation arrays
  - clinical data
- Which genes or genetic regions have survival effect?

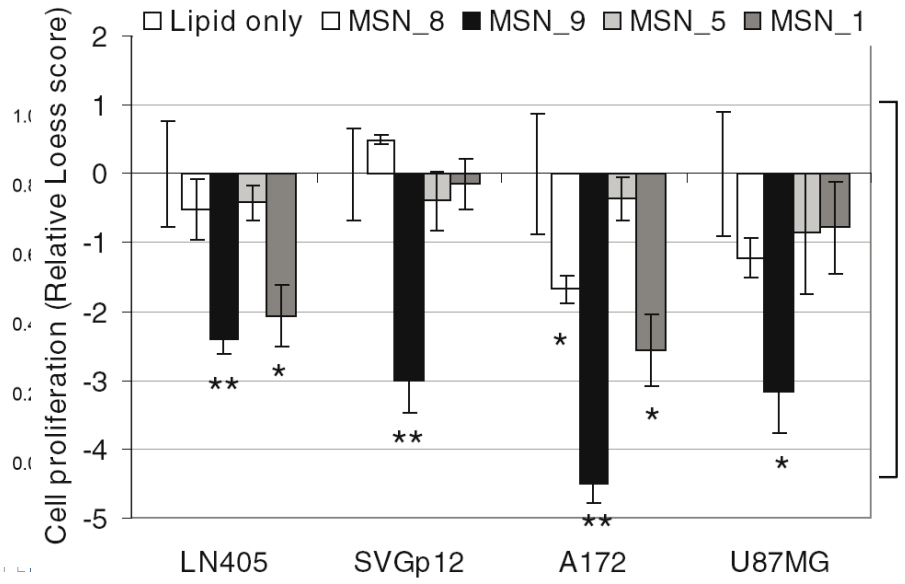
Genome Medicine
RESEARCH Open Access
Large-scale data integration framework provides a comprehensive view on glioblastoma multiforme

GBM Results in Anduril Website

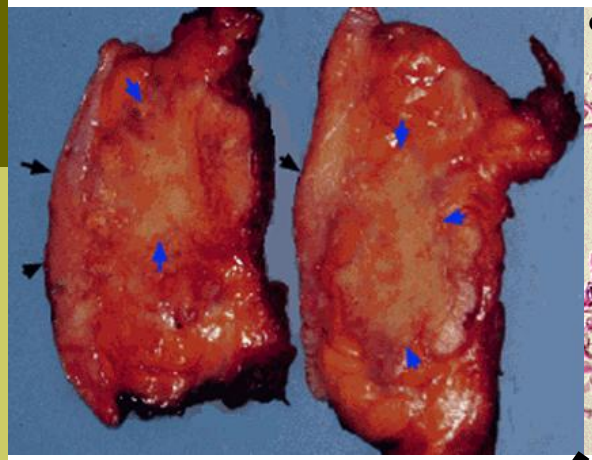
		MedianExonExpression		TranscriptExpression				CGH				
GeneName	GeneExpression	FoldChange	Survival	Min	Max	Survival	SNPSurvival	Gain	Loss	ExpressionIntegration	Methylation	DNABand
ANKRD26	0.639	0.609	-	0.382	0.992	1.45e-6	-	0.0104	0.292	0.0810	-	10p12.1
FAM171A1	0.235	0.437	0.000342	0.280	0.517	1.66e-6	-	0.0104	0.276	0.0120	-	10p13
ADAM22	0.753	0.454	0.000145	0.154	2.32	4.56e-6	-	0.0833	0.00521	-	-	7q21.12
ZNF236	0.814	0.723	-	0.298	0.766	1.23e-5	-	0.00521	0.0104	-	-	18q23
SCRIB	1.41	1.21	-	0.377	3.09	1.85e-5	-	0.0104	0.0156	0.00600	-	8q24.3
NDRG3	0.305	0.486	0.000142	0.119	0.388	2.66e-5	-	0.00521	-	-	-	20q11.23
MSN	5.25	3.55	0.000160	3.42	3.66	2.77e-5	-	-	-	-	-	Xq12
ZRANB1	-	0.449	0.00314	0.361	0.420	3.24e-5	-	-	0.333	0.263	-	10q26.13
NMT2	0.686	0.568	-	0.213	1.10	3.26e-5	-	0.0104	0.276	0.148	0.240	10p13
WAC	0.484	0.553	-	0.252	3.54	3.42e-5	-	0.0104	0.292	0.00	-	10p12.1
TCEAL2	0.156	0.224	0.000112	0.215	0.286	4.94e-5	-	-	-	-	-	Xq22.1
HS3ST3B1	1.09	2.18	0.0228	2.41	3.16	5.04e-5	-	-	0.0156	-	-	17p12

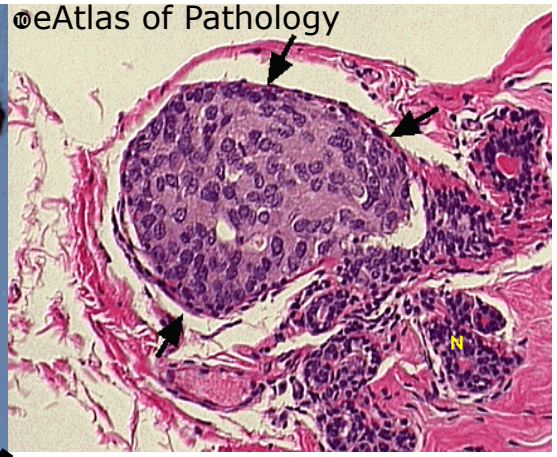
Gene: MSN

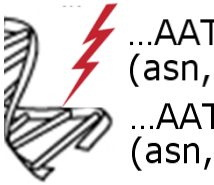
GeneName	MSN
GeneID	ENSG00000147065
GeneExpression	5.25
ExprPValue	-
MedianExonExpression:FoldChange	3.55
MedianExonExpression:PValue	2.43e-10
TranscriptExpression:Min	3.42
TranscriptExpression:Max	3.66
TranscriptExpression:Survival	2.77e-5
SNPSurvival	-
CGH:Gain	-
CGH:Loss	-
CGH:ExpressionIntegration	-
Sequenced	yes
Methylation	-
DNABand	Xq12
Protein Interactions	P26038
GeneDesc	moesin [Source:HGNC Symbol;Acc:7373]
Aliases	4478,ENSG00000147065,MSN,P26038
MedianExonExpression:Survival	0.000160
KEGG pathway	Leukocyte transendothelial migration Regulation of actin cytoskeleton
Image	Kaplan-Meier plot for gene survival: ENSG00000147065 (p-value: 1.595617039756866-4)

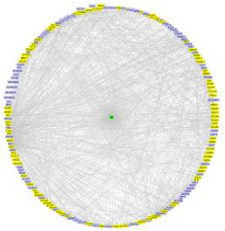


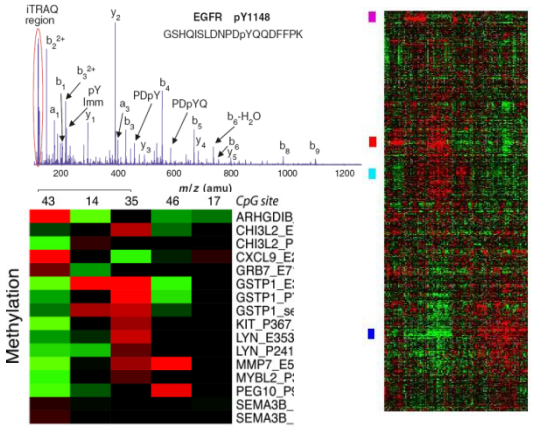
Personalized Treatment

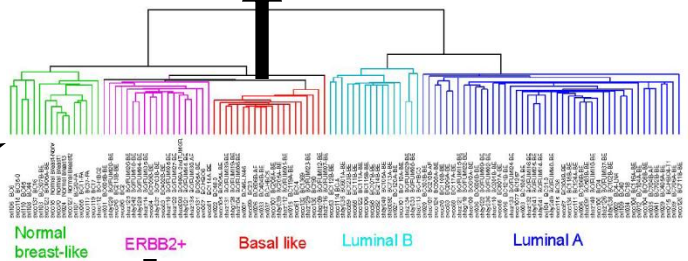


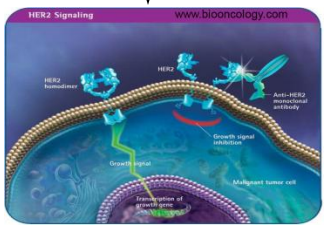


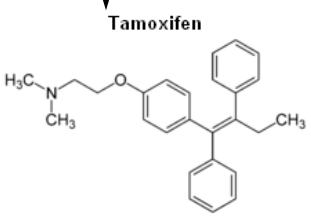






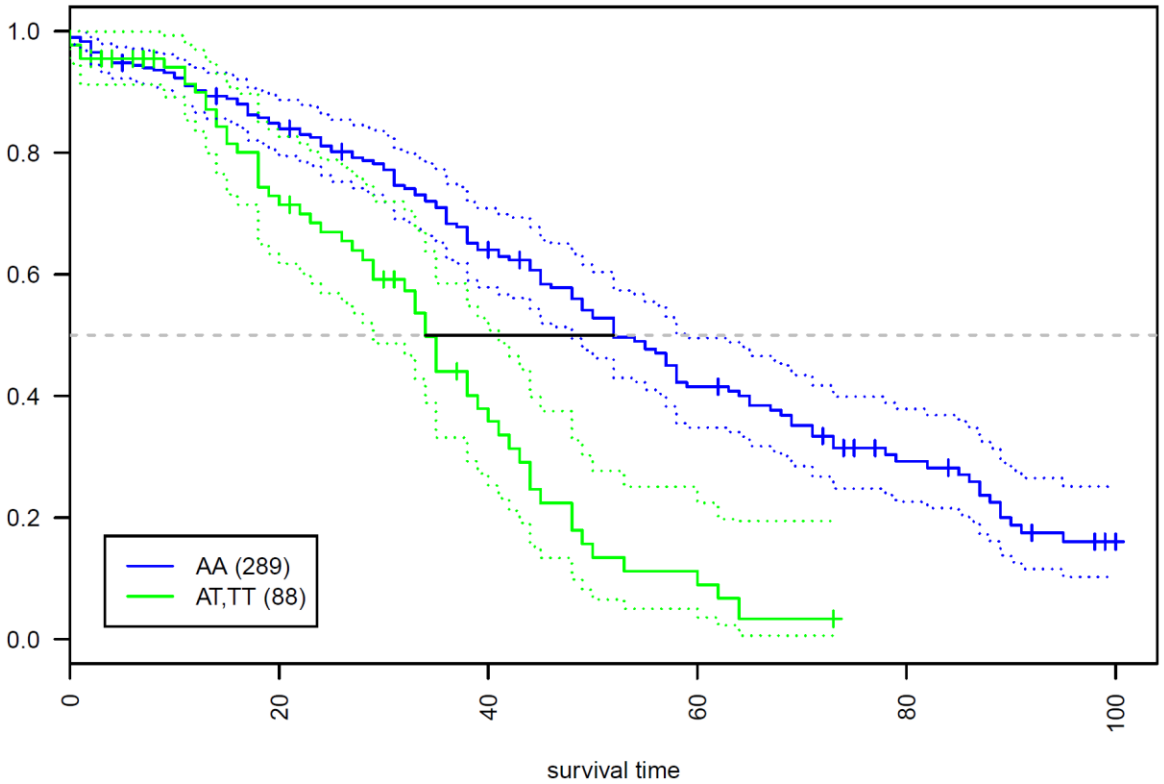






Genetics Play a Key Role in Complex Diseases

- Even a small variation in DNA may have severe effects to protein function, cell phenotypes and survival.

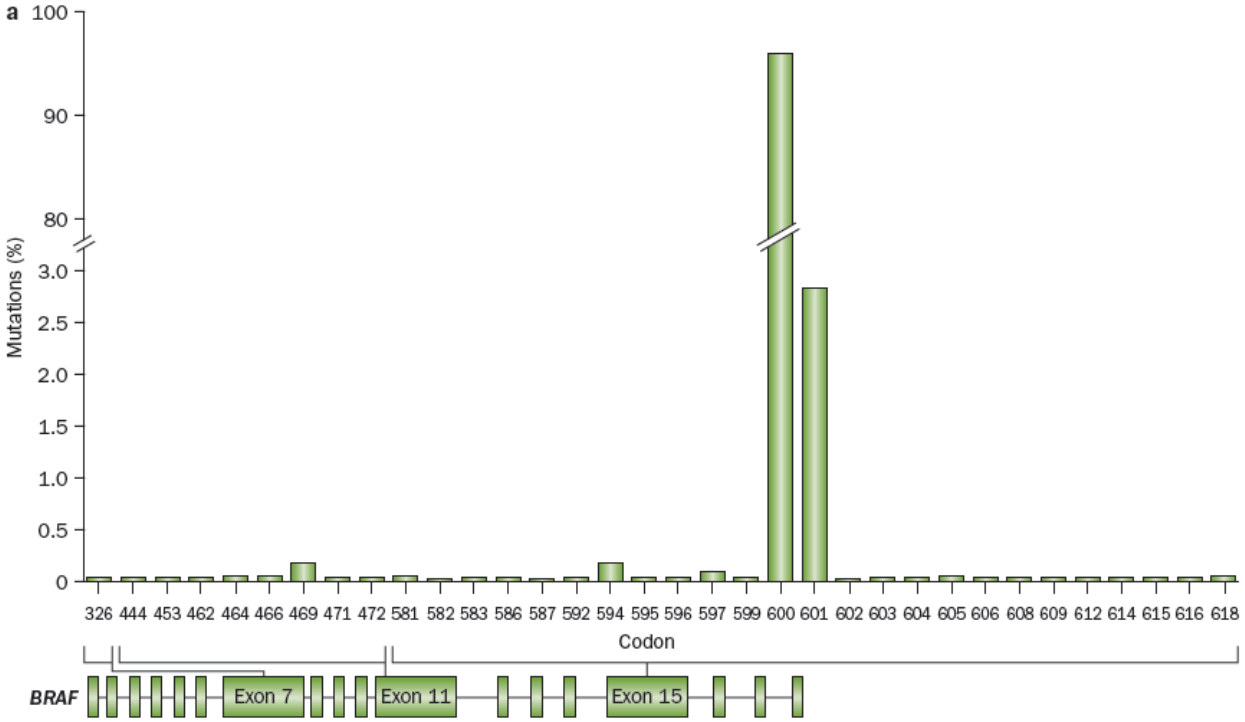


Gene: CLDN10

GeneName	CLDN10
GeneID	ENSG00000134873
GeneExpression: foldchange	0.262
GeneExpression: p-value	1.00
GeneExpression: survival	-
SNPSurvival: p-value	6.04e-8
DNABand	13q32.1
Protein Interactions	P78369
GeneDesc	claudin 10 [Source:HGNC Symbol;Acc:2033]
Aliases	9071,CLDN10,CPETRL3,ENSG00000134873,OSP-L,P78369
TranscriptExpression:Min	0.259
TranscriptExpression:Max	0.280
TranscriptExpression:Survival	0.967
CGH Gain: freq	0.00806
CGH Gain: exp-cor	-
CGH Loss: freq	0.0181
CGH Loss: exp-cor	-
CGHSurvival: p-value	-
Methylation: foldchange	0.982
Methylation: p-value	0.211
Methylation: survival	0.374
CNAmet: score	-
Image	Kaplan-Meier plot for SNP survival: rs12853585 (p-value: 1.91128E-7)
Image	Kaplan-Meier plot for SNP survival: rs12868940 (p-value: 6.04283E-8)
Image	Kaplan-Meier plot for SNP survival: rs4344600 (p-value: 1.30967E-6)
Image	Kaplan-Meier plot for SNP survival: rs4375536 (p-value: 3.75785E-6)

The Location of Mutations Matters

- Mutations are not equally distributed along a gene.
- Below *BRAF*-gene's somatic mutations per codon.



Targeted therapies: how personal should we go?

Miriam Martini, Loredana Vecchione, Salvatore Siena, Sabine Tejpar and Alberto Bardelli

NATURE REVIEWS | CLINICAL ONCOLOGY

VOLUME 9 | FEBRUARY 2012

Use of BRAF Inhibitor In Melanoma

Vemurafenib and *BRAF* mutations

Approximately 40–60% of cutaneous melanomas carry mutations in the *BRAF* gene and the corresponding protein displays increased kinase activity that results in constitutive activation of downstream signaling pathways.[62] *BRAF* mutations are mainly located in the kinase domain, with a single substitution of glutamic acid for valine at codon 600 (V600E) accounting for 80% of all mutations; other, less frequent, activating mutations also occur.[62] Vemurafenib was developed to inhibit the mutated B-Raf protein,[63] and has shown marked antitumor effects on melanoma cell lines carrying the *BRAF* V600E allele but not in cells with wild-type *BRAF*.[64–66] In a phase III randomized clinical trial (BRIM-3), single-agent vemurafenib produced improved rates of overall and progression-free survival in patients with metastatic melanoma, as compared with dacarbazine, the standard treatment comparator.[67] More

The *BRAF* V600E allele is present not only in melanomas but also in other tumor types, including CRC tumors where they are found in approximately 5–10% of cases.[7] It is noteworthy that the presence of the V600E *BRAF* mutation in CRC is apparently not predictive of response to B-Raf inhibitors. For example, most patients with metastatic CRC carrying the *BRAF* V600E allele do not respond to vemurafenib and those that respond do so to a much lesser extent than has been observed in patients with melanoma.[68] The reasons for this discrepancy are not clear; one possibility is that in CRC the

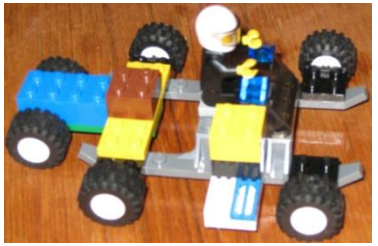
Targeted therapies: how personal should we go?

Miriam Martini, Loredana Vecchione, Salvatore Siena, Sabine Tejpar and Alberto Bardelli

NATURE REVIEWS | CLINICAL ONCOLOGY
VOLUME 9 | FEBRUARY 2012

Genome Medicine: Big Numbers and Promises

- In genomics the number are big.
  - $3x10^9$ nucleotides
  - 20,000-25,000 genes
  - ~100,000 proteins
- These are just the building blocks.
  - Quite a bit to do in categorizing these...
- Real topics still unresolved:
  - Dynamics
  - Context at the pathway level
  - Interactions
  - Impact of cell decisions



Summary

- Characterization of a complex disease first requires identifying the key variables.
- We have tools to measure inner life of cells.
  - Flood of data.
  - Demand for data management and analysis tools.
  - Demand for novel experimental designs and hypotheses.
- Personalized medicine is taking first steps.