II INTERNATIONAL SUMMER SCHOOL

Rare disease and orphan drug registries

Day 1
16.09.2014

Omics and links with biobanks and registries

Mattia Calissano, MD, PhD

MRC Centre for Neuromuscular Diseases at Newcastle
Institute of Genetic Medicine
University of Newcastle, UK

Organised by Istituto Superiore di Sanità
Rome (Italy), September 15-19, 2014

Some (very brief) history

500 years ago

Doctors used to be dressed like below and believed in the influence of stars (the origin of the term ‘flu’ for infulenza) on human affairs.

c.1656

Patients described their symptoms according to a chart

c. 1500

The analysis of ‘biomaterials’ was performed by sight (and often by taste) and therapy was mainly bloodletting

500 years later….

DNA (and not proteins) is the genetic material

"Studies on the chemical nature of the substance inducing transformation of pneumococcal types induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III." Oswald T. Avery, Colin M. MacLeod, and Maclyn McCarty, February 1, **1944**, JEM vol. 79 no. 2 137-158

First sequence of DNA…24 bases

"Nucleotide Sequence Analysis of DNA" Ray Wu *Nature* **236**, 198-200 (19 April **1972**)

'Sanger' sequencing (600bp)

"DNA sequencing with chain-terminating inhibitors"
F. Sanger, S. Nicklen, and A. R. Coulson, Proc Natl Acad Sci U S A. Dec **1977**; 74(12): 5463–5467.

Human genome sequenced (9 months, 3 billion dollars)

'The Sequence of the Human Genome' JC Venter et al, Science 16 February **2001**:
Vol. 291 no. 5507 pp. 1304-1351

Who is depicted in the previous images:

Doctors

Diagnostics

Patients

Therapeutics

Our priorities have not changed much

-Omics, Registries and biobanks

-Omics, Registries and Biobanks

-**Omics**: The collective technologies used to explore the identity, function and interaction of the various molecules that contribute to the form, shape and function of cells and of organisms.

-Registry: A database that collects uniform data regarding the details of individuals and the characteristics of a particular disease which affects them. This serves for scientific, clinical, or policy purposes including information on the disease, current research, clinical trials, help for patients and their families etc.

-Biobanks: A collection of biomedical data and materials (blood, serum, plasma, urine etc.) collected from individuals and physically stored in secure premises under standardized conditions.

Rare diseases

The definition of rare diseases differs in every country.

-USA: prevalence of 1:1500 (a total of circa 200.000 people)
-Japan: 1:2500
-EU: 1:2000

Eurordis (European Organization on rare diseases) estimates that circa 6-8% of the population is affected by a rare disease).

‘Omics are divided according to the subcellular network they explore:

-**Genomics**: the study of DNA encompassing structural genomics (the sequencing of genes) and functional genomics (the function of the encoded genes). This field now also includes epigenetic (events altering gene expression without actual changes of the DNA sequence).

-**Transcriptomics**: The qualitative and quantitative analysis of the RNA population expressed in a defined group of cells, tissues, organs or organisms.

-**Proteomics**: The qualitative and quantitative analysis of the proteins expressed in a defined group of cells, tissue, organ or organism.

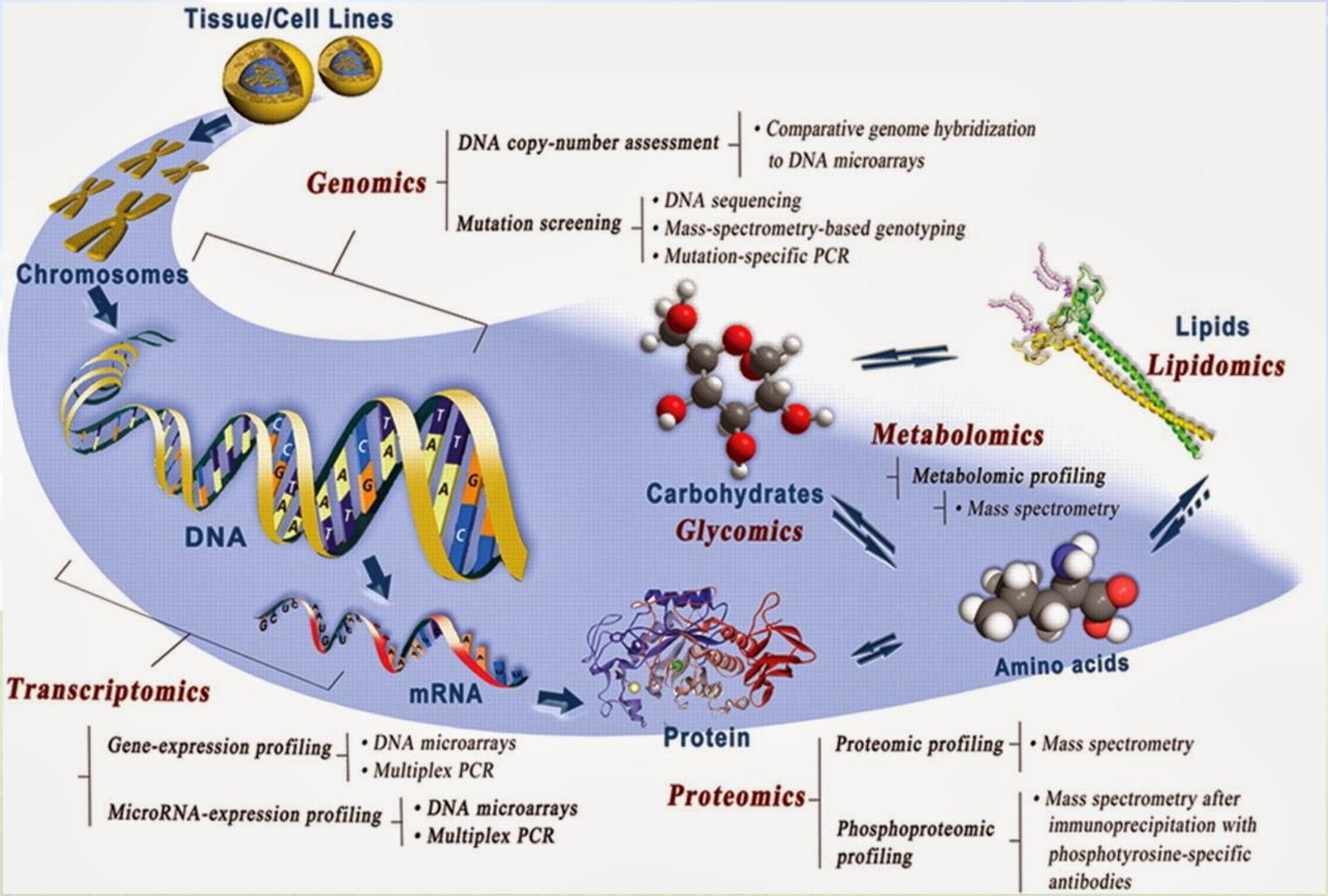
-**Lipidomic**: The qualitative and quantitative analysis of lipids expressed in a defined group of cells, tissue, organ or organism.

Glycomics: The qualitative and quantitative analysis of complex sugars expressed in a defined group of cells, tissue, organ or organism.

-**Peptidomics:** The qualitative and quantitative analysis of the peptides (derived from protein degradation) expressed in a defined group of cells, tissue or organs.

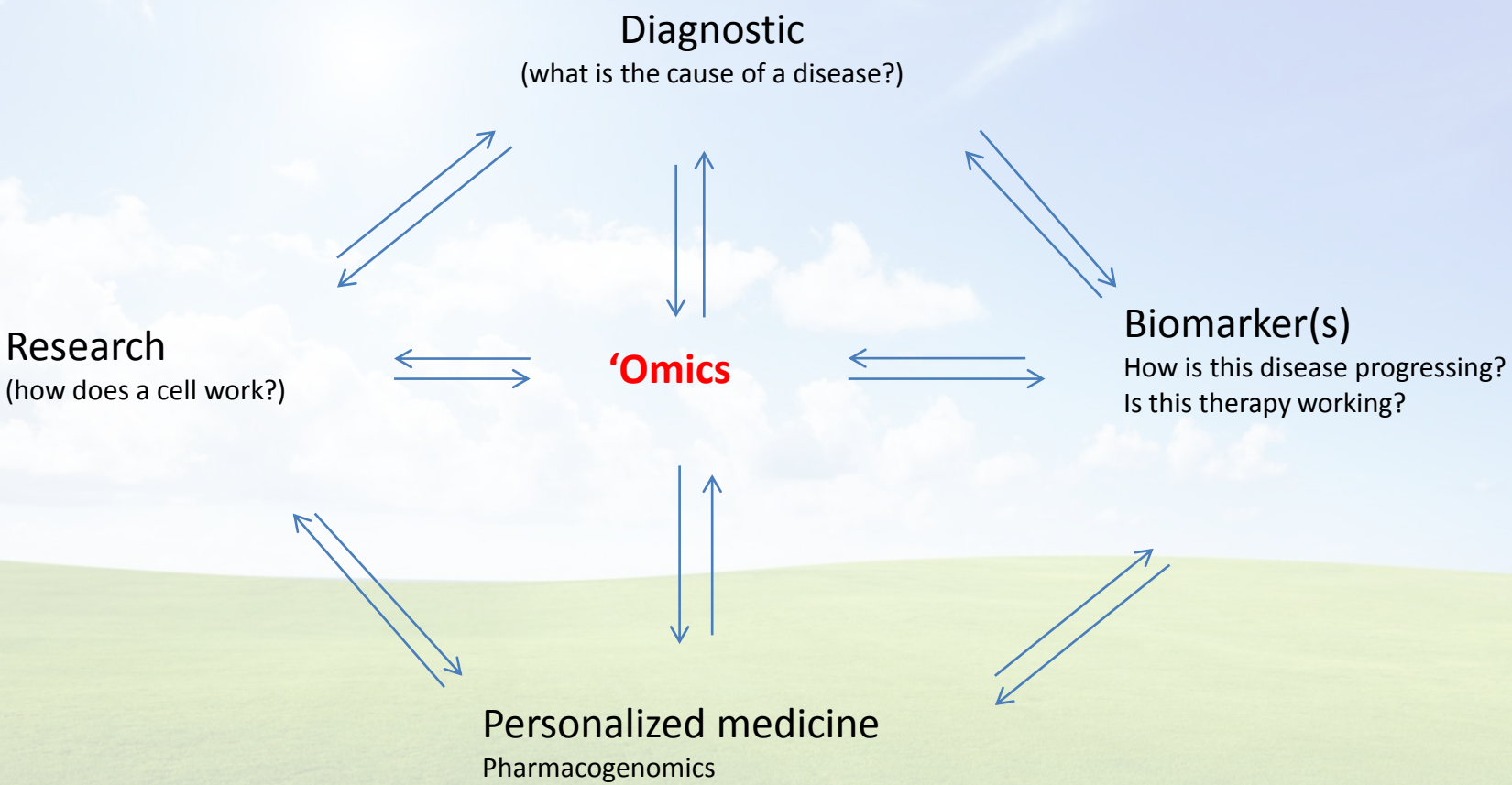
Etc. Etc...

From Genomics to… –Omics

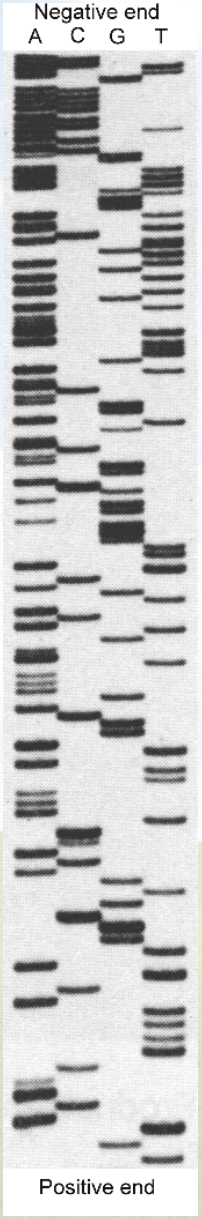


Wu RD et al. JDR 2011; 90:561-572

From -Omics to…

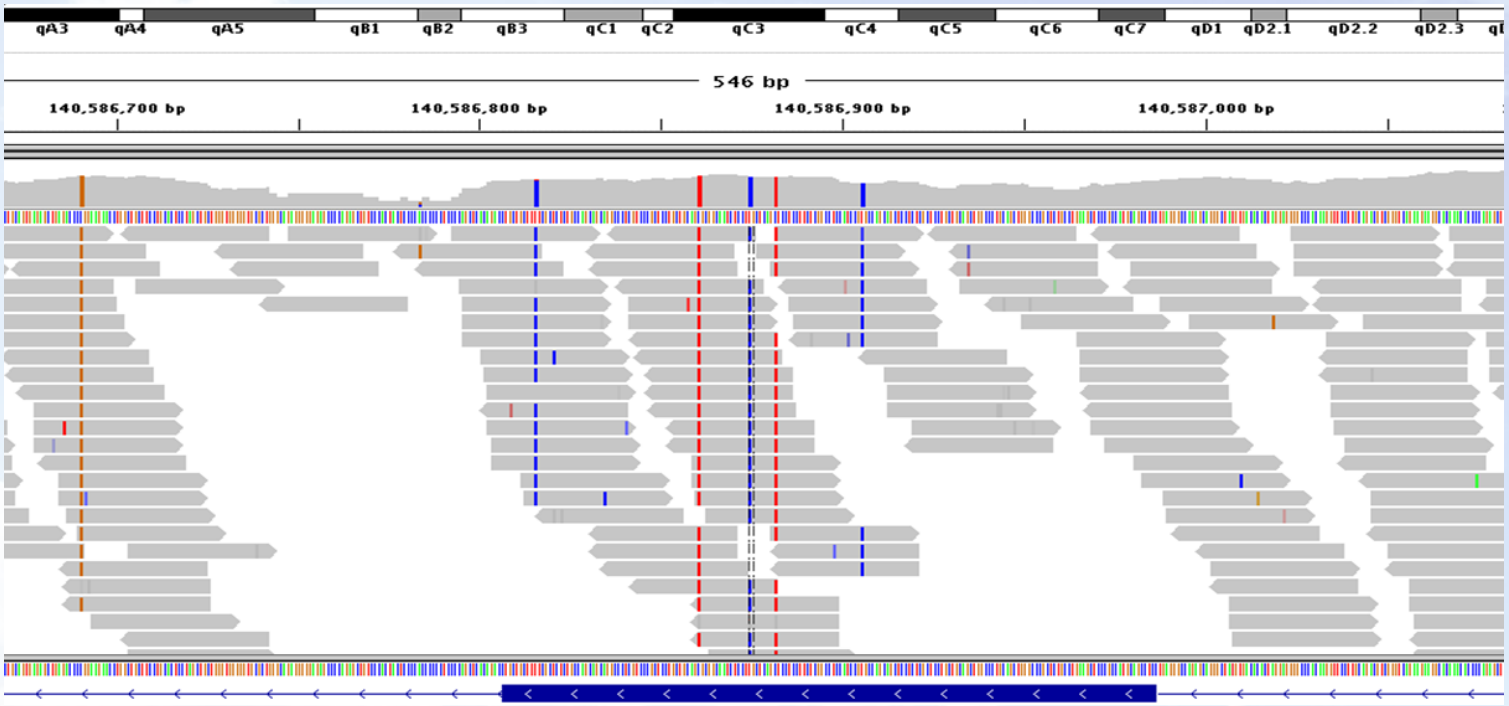


In particular ‘Genomics’ has opened many possibilities as science moved from….



End of 90’s, ‘Sanger sequencing’: 2-3 days for 500-600 bp read.

To…



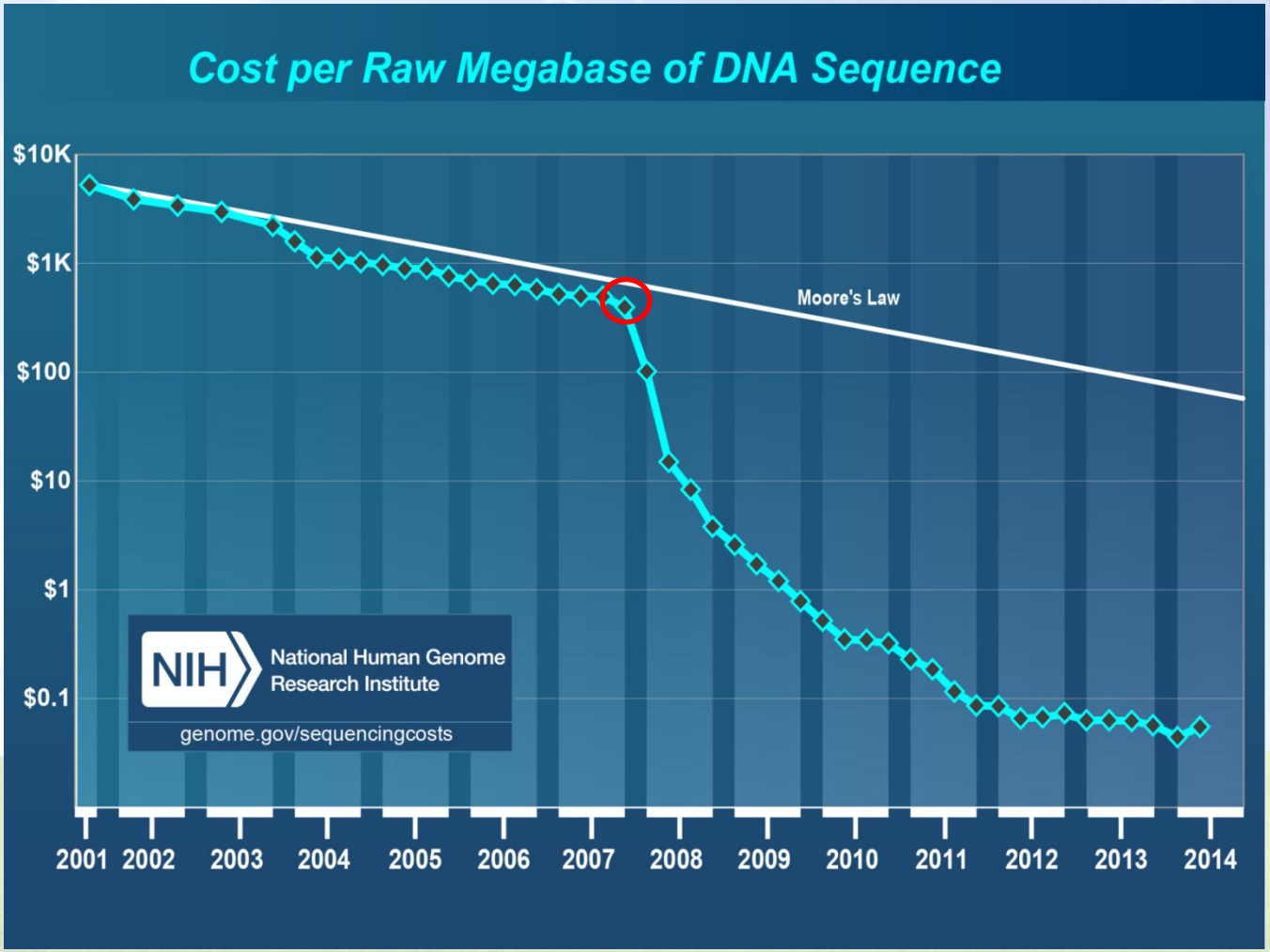
15 years later (circa 2010)… an entire human genome (**3 billion bases**) in **6 hours!**

The dramatic fall in cost for DNA sequencing – working toward the $1,000 genome

Year	1990	2002	2006	2008	2012	2014
Cost	$3 billion	$300 million	$20 million	$2 million	$5,000	$1,000

Source: Carr G

Cost per sequenced megabase



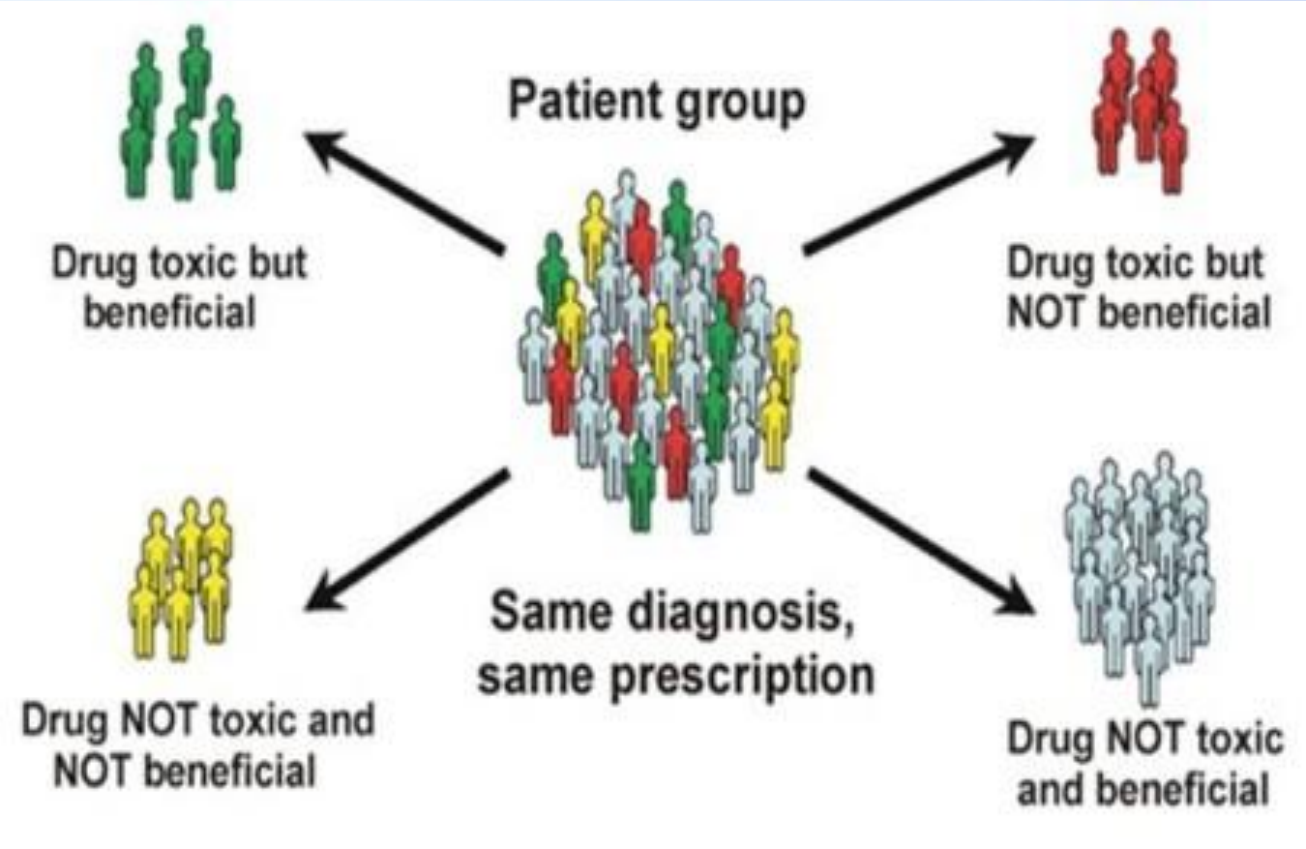
Data from Wetterstrand KA. DNA Sequencing Costs: NHGRI Genome Sequencing Program (GSP)

Transition from Sanger-based (dideoxy chain termination sequencing) to 'second generation' (or 'next-generation') DNA sequencing technologies.

Next Gen. Sequencing costs

	NGS costs (2014)			
	Whole Genome	**Exome**	**Custome capture**	**Amplicon**
Size (Gigabyte)	$100	$2.5	$0.13-1	$0.03-0.13
Preparation	$400	$200	$80	$40
Sequencing	$4300	$400	$12-100	$1-12
Data processing/storage	$350	$200	$50	$25
Clinical analysis	$5000-10000	$2000-6000	$700-2000	$400-900
Total	$10000-15000	$2800-6800	$1000-2000	$500-1000

Pharmacogenomics: personalized treatment with minimum adverse effects.



'What Is Pharmacogenomics', Natalie Khan, 2014

The variability in drug metabolism is due to polymorphisms naturally occurring in the population.

Pharmacogenomics

Gene variants	Disease	Therapeutic agent	Individual response
Her2/Neu receptor	breast cancer	Trastuzumab	polymorphisms in HER2 and cardiac toxicity
Bcr-Abl	chronic myeloid leukaemia	Imatinib mesylate	Various genes affecting response and toxicity
UGT1A1 TA7 variant	colorectal cancer	Irinotecan, Cetuximab, Panitumab	increased toxicity
SNPs in ABCG2 promoter	lung cancer	Erlotinib	skin and gastrointestinal toxicity

List of drugs currently approved by the Food and Drug Administration (FDA) with associated pharmacogenomic information

Clinical specialty	Drugs used	Associated genes
Allergy	Desloratadine and Pseudoephedrine	*CYP2D6*
Analgesics	Celecoxib, Codeine	*CYP2C9, CYP2D6*
	Tramadol and Acetaminophen	*CYP2D6*
Antiarrhythmics	Quinidine	*CYP2D6*
Antifungals	Terbinafine, Voriconazole	*CYP2D6, CYP2C19*
Anti-infectives	Chloroquine, Rifampin, Isoniazid, and Pyrazinamide	*G6PD, NAT1; NAT2*
Antivirals	Abacavir, Boceprevir, Maraviroc Nelfinavir, Peginterferon alfa-2b, Telaprevir	*HLA-B*5701, IL28B, CCR5, CYP2C19, IL28B IL28B*
Cardiovascular	Carvedilol, Clopidogrel, Isosorbide and Hydralazine, Metoprolol, Prasugrel, Pravastatin, Propafenone, Propranolol, Ticagrelor	*CYP2D6, CYP2C19, NAT1; NAT2, CYP2D6 CYP2C19*, Genotype E2/ E2 and Fredrickson Type III dysbetalipoproteinemia, *CYP2D6 CYP2D6, CYP2C19*
Dermatology and Dental	Cevimeline, Dapsone, Fluorouracil, Tretinoin	*CYP2D6, G6PD, DPD PML/RARα*
Gastroenterology	Dexlansoprazole (1)‡, Dexlansoprazole (2), Esomeprazole, Pantoprazole, Rabeprazole Sodium Phenylacetate and Sodium Benzoate, Sodium Phenylbutyrate	*CYP2C19, CYP1A2, CYP2C19, CYP2C19, CYP2C19, UCD (NAGS; CPS; ASS; OTC; ASL; ARG), UCD (NAGS; CPS; ASS; OTC; ASL; ARG)*
Hematology	Lenalidomide, Warfarin (1), Warfarin (2)	5q Chromosome, *CYP2C9, VKORC1*
Metabolic and Endocrinology	Atorvastatin	LDL receptor
Musculoskeletal	Carisoprodol, Mivacurium	*CYP2C9*, Cholinesterase gene
Neurology	Carbamazepine, Dextromethorphan and Quinidine, Galantamine, Tetrabenazine	*HLA-B*1502, CYP2D6, CYP2D6, CYP2D6*
Oncology	Arsenic Trioxide, Brentuximab Vedotin, Busulfan, Capecitabine, Cetuximab (1), Cetuximab (2), Crizotinib, Dasatinib, Erlotinib Fulvestrant, Gefitinib (1), Gefitinib (2), Imatinib (1), Imatinib (2), Imatinib (3), Imatinib (4) Irinotecan, Lapatinib, Mercaptopurine, Nilotinib (1), Nilotinib (2), Panitumumab (1), Panitumumab (2), Rasburicase, Tamoxifen, Thioguanine, Tositumomab, Trastuzumab, Vemurafenib	*PML/RARα*, CD30, Ph Chromosome, DPD *EGFR, KRAS, ALK*, Ph Chromosome, *EGFR* ER receptor, *CYP2D6, EGFR*, C-Kit, Ph Chromosome, *PDGFR, FIP1L1-PDGFRα, UGT1A1, Her2/neu, TPMT*, Ph Chromosome, *UGT1A1, EGFR, KRAS, G6PD*, ER receptor, *TPMT*, CD20 antigen, Her2/neu, *BRAF*
Opthalmology	Timolol	*CYP2D6*
Psychiatry	Aripiprazole, Atomoxetine, Chlordiazepoxide and Amitriptyline, Citalopram (1), Citalopram (2) Clomipramine, Clozapine, Desipramine, Diazepam, Doxepin, Fluoxetine, Fluoxetine and Olanzapine, Fluvoxamine (1), Fluvoxamine (2), Fluvoxamine (3), Iloperidone, Imipramine, Modafinil (1), Modafinil (2), Nefazodone, Nortriptyline, Paroxetine, Perphenazine, Pimozide, Protriptyline, Risperidone, Thioridazine, Trimipramine, Valproic Acid, Venlafaxine	*CYP2D6, CYP2D6, CYP2D6, CYP2C19, CYP2D6, CYP2D6, CYP2D6, CYP2D6, CYP2C19, CYP2D6, CYP2D6, CYP2D6, CYP2C9, CYP2C19, CYP2D6, CYP2D6, CYP2D6, CYP2C19, CYP2D6, CYP2D6, CYP2D6, CYP2D6, CYP2D6, CYP2D6, CYP2D6, CYP2D6, CYP2D6, CYP2D6, CYP2D6, CYP2D6, CYP2D6, UCD (NAGS; CPS; ASS; OTC; ASL; ARG), CYP2D6*
Pulmonary	Tiotropium	*CYP2D6*
Reproductive	Drospirenone and Ethinyl Estradiol Clomiphene, Tolterodine	*CYP2C19*, Rh genotype, *CYP2D6*
Rheumatology	Azathioprine, Flurbiprofen	*TPMT, CYP2C9*

‡Data source - http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm, ‡Numbers in the brackets indicate that the drug is affected by multiple genetic polymorphisms

R Gullapalli et al. J Pathol Inform 2012, 3:40

Implications of reduced NGS costs

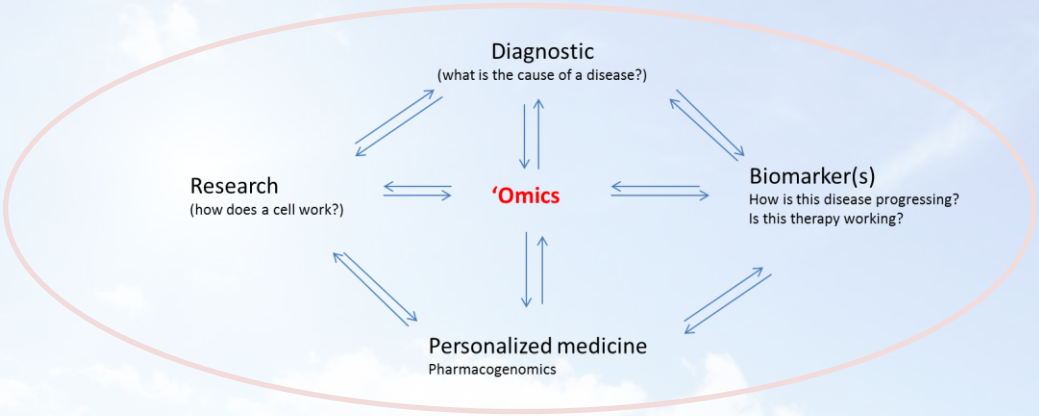
-**Increased** accuracy of diagnosis of genetic diseases.
-**Faster** diagnosis.
-**Increased** statistical analysis of genetic diseases for health policies.
-**Possibility** of pharmacogenomics (reduced adverse events, secondary effects and overall expenses).

But

Before the actual DNA sequence and other 'omics studies can occur, the following are needed:

-Details of the patient and the disease, natural history, family history etc.
-Specific consent from the patient for sampling, storage and analysis of biomaterials

From omics to registries and biobanks and…back



Information about the disease and the patient.

-Who is it?
-What age, sex, type of disease?
-What is the gene(s) involved? Type of mutation?
-Consent for clinical/natural history studies?
-Etc.

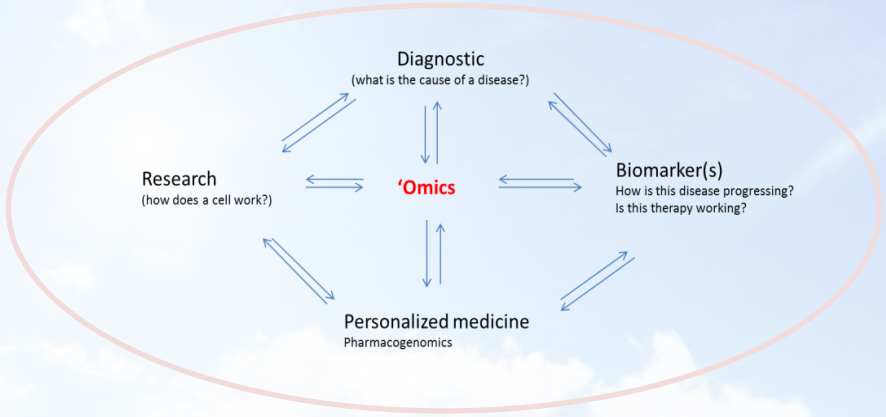
Registries

Storage and availability of biomaterials

-What type of biomaterial: blood, serum, cells, etc.
-How many samples originating from patients with similar charact
are stored?
-What type of controls are stored?
-Etc.

Biobanks

Issues with the current systems



Information about the sample
(What is it, when was it taken, Where does it come from?)

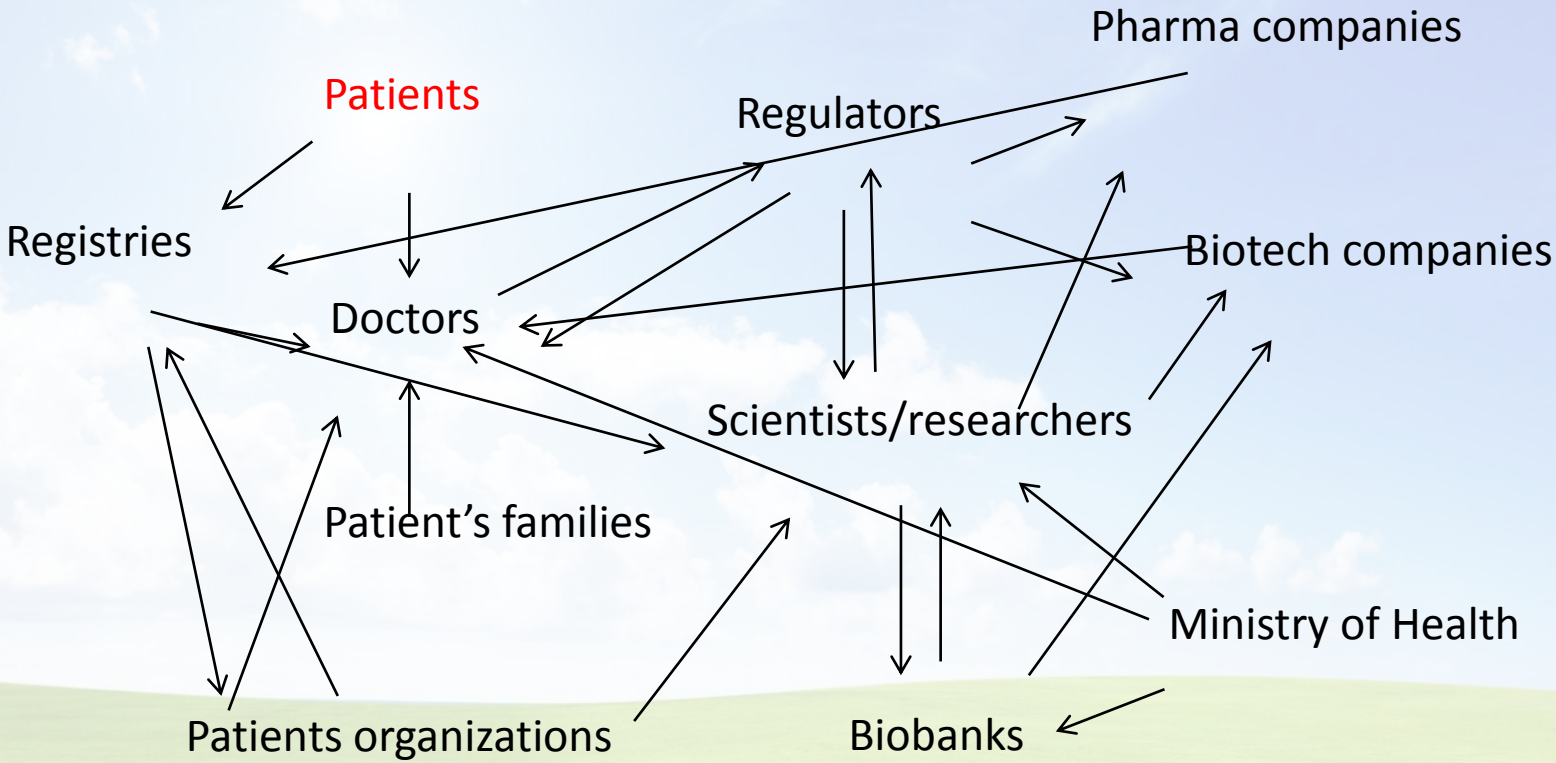
Storage and availability to study the sample
(where is it? How many similar samples are stored? Are then any controls stored?)

Registries ⇄ **Biobanks**

Lack of-/incomplete integration between biobanks and registries

Who will benefit by an increased integration between biobanks and registries?

Who are the stakeholders when it comes to health?



The high number of stakeholders and the deep fragmentation of information regarding patients, their disease, the type(s) of samples stored, their availability etc., has several negative implications for patients and **all** research/diagnostic/clinical trial studies.

National/international registries, biobanks and their functions

Registries

Database of patients with same disease, info on disease, support groups, sharing scientific advances etc.

Time

Identification of patients with similar characteristics for research purposes.

Time

Pharma company investigation for feasibility studies.

Time

Anonymization of data for external use

Biobanks

Database of stored biomaterial, some info regarding the material (e.g. serum, plasma, urine etc.) and the disease (e.g. gene, mutation).

Time

Enquiry by researchers, biotech and pharma companies for 'omics/biomarker studies.

Time

Anonymization of data for external use

Time

Functional study of new genes.

Consent

Consent

Consent

Consent

Partial list of European Biobanks

European Biobanks		
AT	Graz	Biobank Graz
AT	Vienna	MUW Biobank
AT	Vienna	Ludwig Boltzman Tumour Bank
BE	Brussels	Saint-Luc Tumour Bank
BE	Luxembourg	Integrated Biobank of Luxembourg
CH	Bern	Biobank Suisse
CH	Pfaeffikon	Cryo-Save
CH	Bern	Tumor Bank Bern
CH	Geneva	East West Biopharma
CH	Monthey	Swiss Biobank
CY	Limassol	StemCure
DE	Berlin	Central Biomaterial Bank
DE	Berlin	SepNet Central Sample Bank
DE	Munich	BrainNet Europe
DE	Munich	KORA-gen
DE	Regensburg	Danubian Biobank Consortium
EE	Tartu	Estonia Biobank
ES	Barcelona	Biobank of Hospital Clínic - IDIBAPS
ES	Granada	Andalusian Regional Tumour Bank
ES	Madrid	HIV Biobank
ES	Madrid	Spanish National Tumour Bank Network
FI	Helsinki	National Biobank of Finland
FI	Turku	Biobanking and Biomolecular Resources Research
Infrastructure Institute BBMRI		
FR	Amiens	Biobanque de Picardie
FR	Evry	Genethon DNA and Cell Bank
FR	Lyon	Biobank Resource Centre de France
FR	Lyon	European Sarcoma & Tumour Bank
FR	Marseille	Tumour Bank of Provence
FR	Toulouse	Southwest France Tumour Bank
GR	Marousi	Stem-Health Hellas Stem Cell Bank
HU	Budapest	Hungarian Neurological-Psychiatric Biobank
IE	Dublin	Biobank Ireland Trust
IE	Dublin	Dublin Brain Bank
IE	Belfast	Northern Ireland Virtual Tissue Archive
IE	Wexford	Biostor Ireland
IE	Trinity	Trinity Biobank
IT	Florence	da Vinci European Biobank
IT	Genoa	Biological Resource Centre - National Institute for
Cancer Research		
IT	Genoa	Telethon Genetic Biobank Network
IT	Milan	BioRep
IT	Milan	Movement Disorders Biobank
IT	Padua	EuroBioBank
LV	Riga	Latvia Biobank
NL	Groningen	ErasmusMC
NL	Groningen	Lifelines Biobank
NL	Rotterdam	European Human Frozen Tissue Bank
NO	Trondheim	Regional Biobank of Central Norway
NO	Levanger	Hunt Biobank
SE	Lund	Swedish Regional Biobank
SE	Malmo	European Cancer Biobank
SE	Stockholm	Karolisnka Biobank
SE	Stockholm	LifeGene
SE	Stockhom	Swedish National Biobank Program
SE	Umea	Biobank of Northern Sweden
UK	Birmingham	Central England Haemato-Oncology Research Biobank
UK	Cambridge	CamUro-Oncology Biobank
UK	Cambridge	Geneservice
UK	Cardiff	Wales Cancer Bank
UK	Dundee	Tayside Tissue Bank
UK	Edinburgh	Edinburg Brain & Tissue Banks
UK	Glasgow	Biopta
UK	Hertfordshire	onCore UK
UK	Hertfordshire	UK Stem Cell Bank
UK	Leicester	Children's Cancer and Leukaemia Tissue Bank
UK	Liverpool	Cancer Tissue Bank Research Centre
UK	London	Chernobyl Tissue Bank
UK	London	Confederation of Cancer Banks
UK	London	King's College Infectious Diseases BioBank
UK	London	Neuroendocrine Tumors Biobank
UK	London	UK Parkinson's Disease Tissue Bank
UK	London	UK Multiple Schlerosis Tissue Bank
UK	London	Virgin Health Bank
UK	Manchester	UK DNA Banking Network
UK	Midlothian	Roslin Wellcome Trust Tick Cell Biobank
UK	Newcastle Upon Tyne	HDBR - Human Developmental Biology Resource
UK	Oxford	Autism Brain Bank
UK	Plymouth	BioVault
UK	Salisbury	European Collection of Cell Cultures
UK	Southampton	Southampton Tumour Bank
UK	Stockport	UK Biobank

Example of patient registries for neuromuscular diseases.

Disease Name	Type of Registry
Congenital muscular dystrophies (CMD)	International
Congenital myasthenic syndromes (CMS)	International (under construction)
Charcot Marie Tooth disease (CMT)	International
Duchenne/Becker muscular dystrophy (DMD/BMD)	National
Facioscapulohumeral muscular dystrophy (FSHD)	National
Hereditary inclusion body myopathy (HIBM/GNE myopathy)	International
Limb girdle muscular dystrophies (LGMD)	International (several)
Myotonic dystrophy (DM)	National
Myotubular and centronuclear myopathy (MTM and CNM)	International
Spinal muscular atrophy (SMA)	National

Ideal registries and biobanks

-Easily accessible databases for the identification of patients with similar characteristics for clinical trials and other studies.

-Internationally standardized protocols for the sampling and storage of biomaterials.

-Simplification and harmonisation of the consent forms required to perform the above studies.

-Shared ethical principles for good research practices.

-Incentives to share data amongst various centres to promote fast, cost-effective and efficient research.

RD-Connect: a possible solution to the current fragmentation and reduced integration

RD-Connect is a global EU-funded project that links up databases, registries, biobanks and clinical bioinformatics data used in rare disease research into a **central resource** for **researchers worldwide**.

Major objectives:

Contribution to the IRDiRC objectives of delivering 200 new therapies for rare diseases and means to diagnose most rare diseases by the year 2020

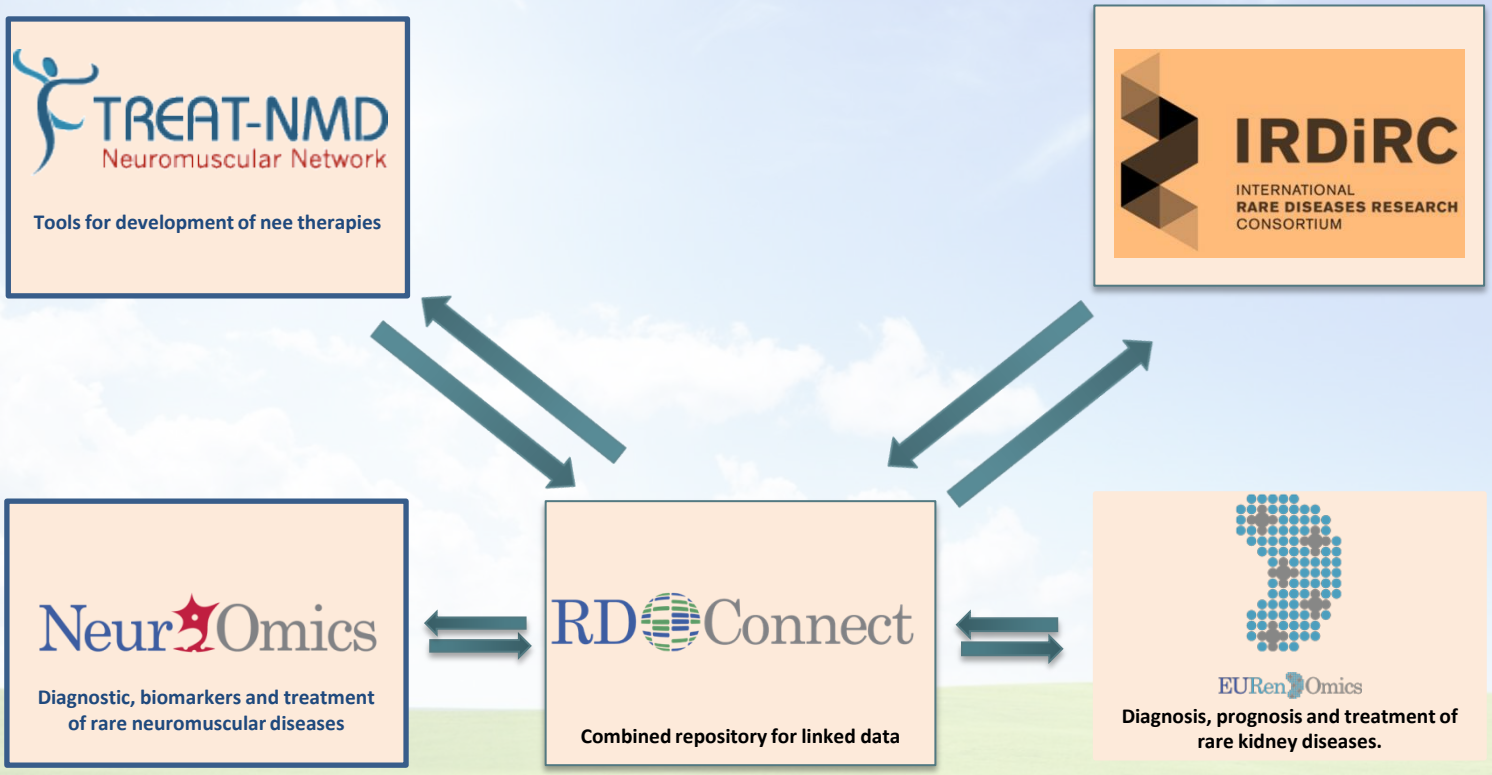
Development of an integrated, quality-assured and comprehensive platform in which complete clinical profiles are combined with -omics data and sample availability for rare disease research, in particular IRDiRC-funded research.

(From the RD-connect website)

RD-Connect additional objectives

- **Patient registries:** developing best practice for registries used for research – establishing interoperability standards, common data elements => feeding into central platform.
- **Biobanks:** developing interoperability standards, common MTAs, searchable online catalogue of sample availability (building on EuroBioBank and BBMRI) => feeding into central platform.
- **Bioinformatics tools:** developing and integrating clinical bioinformatics tools and making them accessible through the central platform and via APIs and web services.
- **Ethical, legal and social issues:** addressing data sharing and informed consent for omics research, proposing a regulatory framework for linking RD medical and personal data, integrating patient perspective.

RD-Connect links and connects two research projects (Neuromics and EURenOmics) and participates in the IRDiRC (International Rare Diseases Research consortium) and TREAT-NMD.



NeurOmics: revolutionise diagnostics and develop new therapies for ten mayor neurodegenerative and neuromuscular diseases through collaborative –omics research.

EURenOmics: develop novel tools to make more accurate diagnoses, predict the disease course and the efficacy of new therapies for rare kidney diseases.

IRDiRC: deliver 200 new therapies and means to diagnose most rare diseases by 2020.

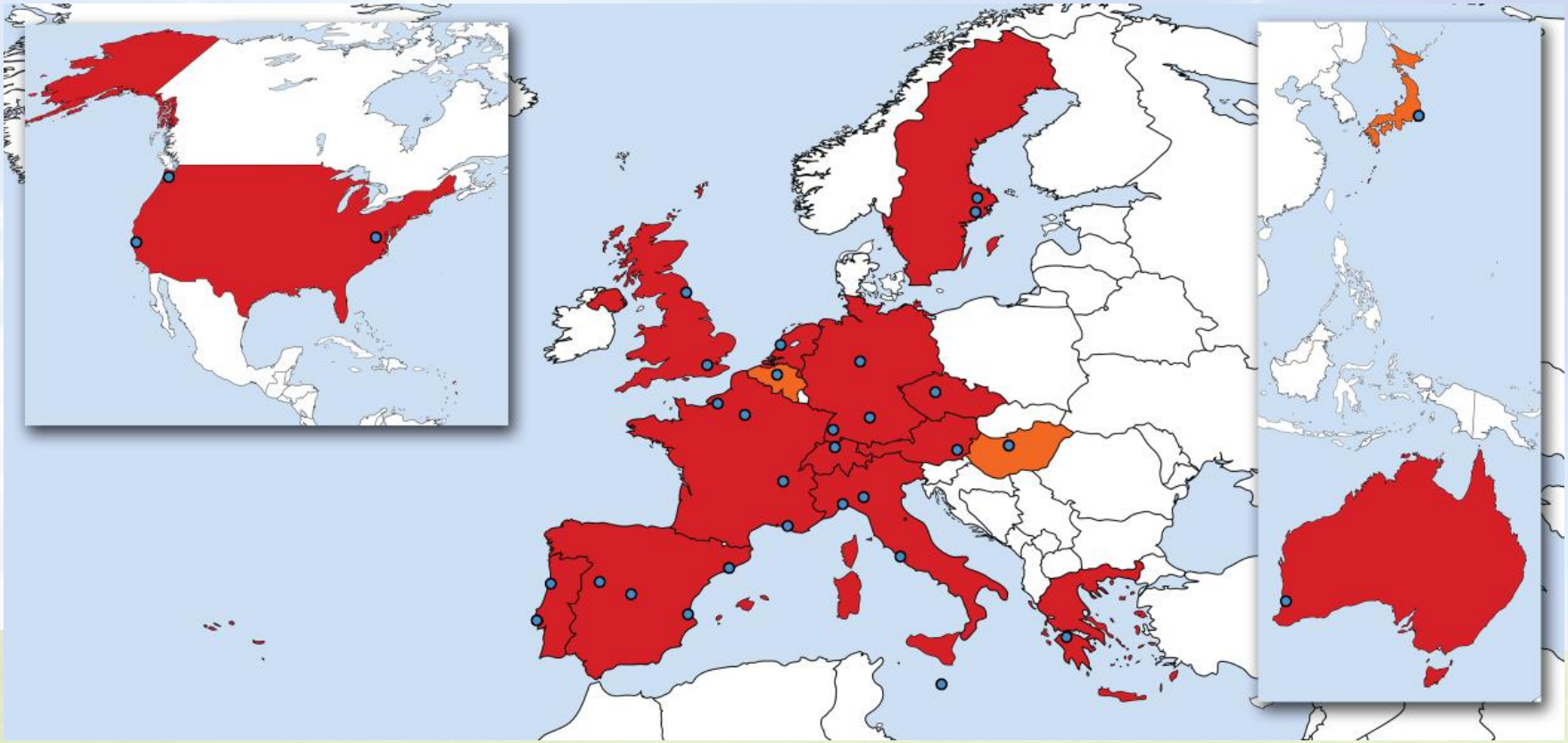
TREAT-NMD: development of tools that industry, clinicians and scientists need to bring novel therapeutic approaches through preclinical development and into the clinic, and on establishing best-practice care for neuromuscular patients worldwide.


European platform
for rare diseases
EPIRARE
TREAT-NMD
Neuromuscular Network
Tools for development of nee therapies
IRDiRC
INTERNATIONAL
RARE DISEASES RESEARCH
CONSORTIUM
NeuroOmics
Diagnostic, biomarkers and treatment
of rare neuromuscular diseases
RD Connect
Combined repository for linked data
EURenOmics
Diagnosis, prognosis and treatment of
rare kidney diseases.
EuroBioBank
European network of biobanks

“8-point plan” for sharing within RD-Connect

1. Project investigator sends sample for omics processing.
2. Raw data is generated and the investigator is given access within the project.
3. Investigator may continue to work towards his goals on that processed data.
4. A copy of the **raw data** is deposited with the European Genome-Phenome Archive (**EGA**) with appropriate restrictions.
5. After consent of the investigator, RD-Connect takes a copy of the raw data from the EGA and runs it through a standardised pipeline – the same pipeline is to be used for data from all IRDiRC projects to ensure cross-compatibility.
6. The processed data after going through the standardised pipeline is held in the RD-Connect Data Coordination Centre (DCC).
7. Access to the data in DCC is granted to individual researchers based on their credentials.
8. Analysis tools will be made available and researchers will have the opportunity to access and analyse their data and cross-compare with all other data from other researchers.

RD-Connect partners



27 full partners, 17 associated partners, 17 countries

Workpackage leaders

WP1: Coordination

Hanns Lochmüller
Newcastle and TREAT-NMD

WP2: Patient registries

Domenica Taruscio
ISS and EPIRARE

WP3: Biobanks

Lucia Monaco
Fondaz. Telethon & EuroBioBank

WP4: Bioinformatics

Christophe Béroud
INSERM Marseille

WP5: Unified platform

Ivo Gut
CNAG Barcelona

WP6 Ethical/legal/social

Mats Hansson
Uppsala

WP7: Impact and innovation

Kate Bushby
Newcastle and EUCERD/ EJARD

From RD-Connect to other practices

The example of RD-Connect can be expanded from rare-diseases to other fields

facilitating

the exchange of information, material and knowledge between various stakeholders

benefiting

patients, health organizations, regulators and the industry.

Further reading:

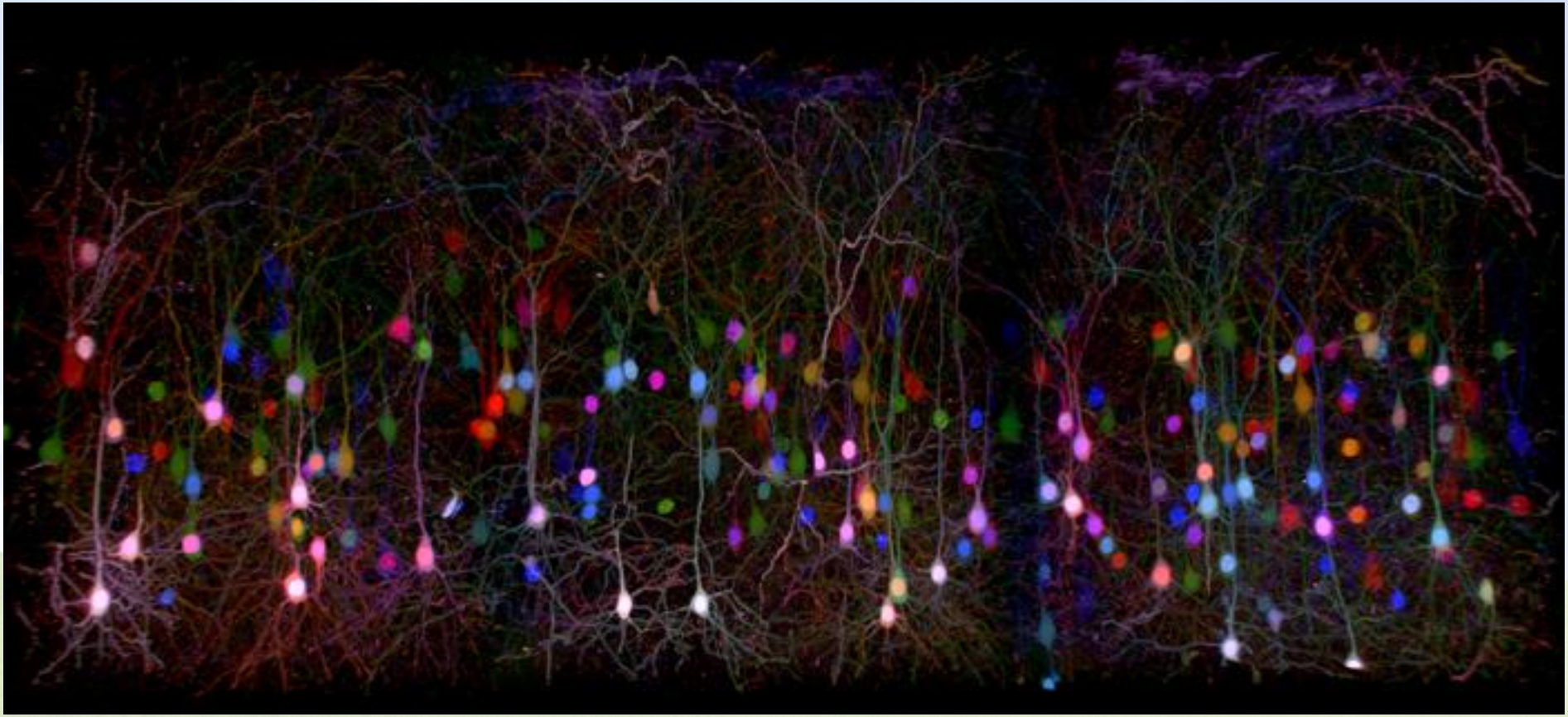
Websites:

http://rd-connect.eu/,
http://www.eurobiobank.org/,
http://rd-neuromics.eu/
http://www.treat-nmd.eu/
http://www.irdirc.org/
http://www.eucerd.eu/
http://www.epirare.eu
http://www.rarebestpractices.eu
http://www.orpha.net/consor/cgi-bin/index.php

Articles:

-Thompson R. et al. “RD-Connect: An Integrated Platform Connecting Databases, Registries, Biobanks and Clinical Bioinformatics for Rare Disease Research”. Journal of General Internal Medicine August 2014, Volume 29, Issue 3 Supplement, pp 780-787

‘Brainbow’: labelling of individual neurons in the mouse brain.



“Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system”. J. Livet et al., Nature, 2007

Thank you!

A special thanks to:

Hanns Lochmuller,
Volker Straub
Karen Rafferty
Libby Woods,
Oksana Pogoryelova
Rachel Thompson
Emma Heslop,
Victoria Hedley