www.manchester.ac.uk

CIGMR www.medicine.manchester.ac.uk/cigmr

www.dna-network.ac.uk

www.biobanks.eu

Biobanking infrastructure

Policy development and practice in research infrastructures for human genome bioresources in UK and EU

Martin Yuille

University of Manchester

Contents

- Definitions
- Why is biobanking important?
- Types of biobank
- An example
- The growth of biobanking
- A development strategy
- Challenges in biobanking

7

Definitions

- Infrastructure
  - The basic facilities, services and installations needed for the functioning of a community or society
    - Transport and communications systems
    - Water and power lines
    - Institutions (schools, prisons, universities etc.)

8

Definitions

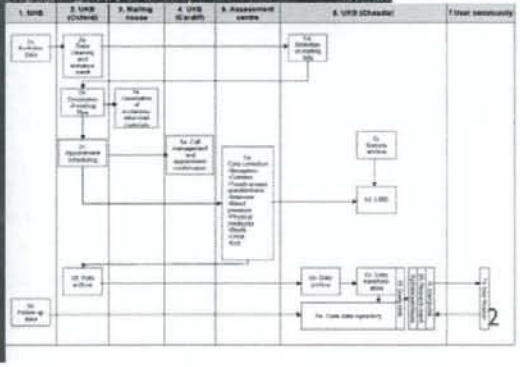
- Research infrastructure
  - The special facilities, services and installations needed to implement experimental designs e.g. Telescopes, Satellites, Accelerators, Resource networks

Manchester University's Jodrell Bank Radiotelescope

Definitions

- Human biobank
  - A research infrastructure needed to implement experimental designs in molecular epidemiology and pathology
    - For example: liquid handling robots; data networks
  - This infrastructure is large – relative to historical facilities for biobanking
- Human biobanking
  - Activities undertaken by a (human) biobank
    - For example: accrual, management and distribution of samples and data

10



1. View of UK BioBank
2. Workflow for accrual at UK Biobank
3. View of UK DNA Banking Network

11

Contents

- Definitions
- Why is biobanking important?
- Types of biobank
- An example: a 'fair access' biobank
- The growth of biobanking
- Biobanking policies
- Challenges in biobanking

12

Why is biobanking important?

- Scientific justification
- Social justification
- Economic justification

13

Why is biobanking important?

- Scientific justification
  - Biobanks are research infrastructures that underpin virtually all biomedical experimental research aimed at improving human health
  - The need for larger and larger sets of samples and data grows directly as researchers ask more and more detailed questions: large datasets are required to detect small effects
  - Biobanking is essential to
    - Ensure adequate and consistent quality of the resources
    - Reduce unit costs
    - Deliver resources in a timely way
  - Biobanking changes the way research is conducted
    - It reduces competition for resources
    - It increases intellectual competition

14

The need for big collections

Nature (online) 27 May 2007

ARTICLES

Genome-wide association study identifies novel breast cancer susceptibility loci

Douglas F. Easton[1], Karen A. Pooley[1], Alison M. Dunning[2], Paul D. P. Pharoah[2], Deborah Thompson[1], Dennis G. Ballinger[3], Jeffery P. Struewing[4], Jonathan Morrison[5], Helen Field[5], Robert Luben[6], Nicholas Wareham[6], Shahana Ahmed[6], Catherine S. Healey[2], Richard Bowman[7], the SEARCH collaborators[8], Kerstin B. Meyer[1], Christopher A. Haiman[9], Laurence K. Kolonel[9], Brian E. Henderson[9], Loic Le Marchand[9], Paul Brennan[10], Suleeporn Sangrajrang[11], Valerie Gaborieau[10], Fabrice Odefrey[10], Chen-Yang Shen[12], Pei-Ei Wu[12], Hui-Chun Wang[13], Diana Eccles[14], G. Gareth Evans[15], Julian Peto[16], Olivia Fletcher[16], Nichola Johnson[16], Sheila Seal[17], Michael R. Stratton[17,18], Nazneen Rahman[17], Georgia Chenevix-Trench[19], Stig E. Bojesen[20], Børge G. Nordestgaard[20], Christen K. Axelsson[20], Montserrat Garcia-Closas[21], Louise Brinton[21], Stephen Chanock[21], Jolanta Lissowska[22], Beata Peplonska[23], Heli Nevanlinna[24], Rainer Fagerholm[24], Hannaleena Eerola[24], Daehee Kang[25], Keun-Young Yoo[26,27], Dong-Young Noh[28], Sei-Hyun Ahn[29], David J. Hunter[30,31], Susan E. Hankinson[32], David G. Cox[33], Per Hall[34], Sara Wedren[35], Jianjun Liu[36], Yen-Ling Low[36], Natalia Bogdanova[37,38], Peter Schürmann[38], Thilo Dörk[38], Rob A. E. M. Tollenaar[39], Catharina E. Jacobi[39], Peter Devilee[39], Jan G. M. Klijn[40], Alice J. Sigurdson[21], Michele M. Doody[21], Bruce H. Alexander[30], Jinghui Zhang[3], Angela Cox[41], Ian W. Brock[41], Gordon MacPherson[42], Malcolm W. R. Reed[43], Fergus J. Couch[44], Ellen L. Goode[44], Janet E. Olson[44], Hanne Meijers-Heijboer[45,46], Ans van den Ouweland[47], Andre Uitterlinden[48], Fernando Rivadeneira[48], Roger L. Milne[49], Gloria Ribas[49], Anna Gonzalez-Neira[49], Javier Benitez[49], John L. Hopper[50], Margaret McCredie[51], Melissa Southey[52], Graham G. Giles[53], Chris Haiman[54], Christina Justenhoven[55], Hiltrud Brauch[56], Ute Hamann[57], Yon-Dschun Ko[58], Amanda B. Spurdle[59], Jonathan Beesley[59], Xiaoqing Chen[59], kConFab[59], AOCS Management Group[59,60], Arto Mannermaa[61,62], Veli-Matti Kosma[61,62], Vesa Kataja[63,64], Jaana Hartikainen[65,66], Nicholas E. Day[1], David R. Cox[67] & Bruce A. J. Ponder[3,7]

Breast cancer exhibits familial aggregation, consistent with variation in genetic susceptibility to the disease. Known susceptibility genes account for less than 25% of the familial risk of breast cancer, and the residual genetic variance is likely to be due to variants conferring more moderate risks. To identify further susceptibility alleles, we conducted a two-stage genome-wide association study in 4,398 breast cancer cases and 4,316 controls, followed by a third stage in which 30 single nucleotide polymorphisms (SNPs) were tested for confirmation in 21,860 cases and 22,578 controls from 22 studies. We used 227,876 SNPs that were estimated to correlate with 77% of known common SNPs in Europeans at $r^2 > 0.5$. SNPs in five novel independent loci exhibited strong and consistent evidence of association with breast cancer ($P < 10^{-5}$). Four of these contain plausible causative genes (*FGFR2*, *TNRC9*, *MAP3K1* and *LSP1*). In the second stage, 1,792 SNPs were significant at the $P < 0.05$ level, compared with an estimated 1,343 that would be expected by chance, indicating that many additional common susceptibility alleles may be identifiable by this approach.

plausible causative genes...
FGFR2
TNRC9
MAP3K1
LSP1

15

Breast cancer: >17 countries, 50,000 samples

1. Clinical genetics centres in the UK and a national study of bilateral breast cancer.
2. European Prospective Investigation of Cancer
3. Australian Breast Cancer Family Study
4. Breast Cancer Study in Taiwan
5. Copenhagen Breast Cancer Study and General Population Study
6. Gene Environment Interaction and Breast Cancer in Germany
7. Hannover Breast Cancer Study
8. Helsinki Breast Cancer Project + additional familial cases
9. IARC – Thai Breast Cancer Study
10. Kathleen Cuningham Foundation Consortium for Familial Breast Cancer and Australian Ovarian Cancer
11. Australia clinic-based recruitment of familial breast cancer patients (cases)
12. New Zealand; clinic-based recruitment of familial breast cancer patients (cases)
13. Australia; population-based case-control study of ovarian cancer (controls only)
14. Kuopio Breast Cancer Project
15. Leiden University Medical Centre Breast Cancer Study
16. Mayo Clinic Breast Cancer Study
17. Melbourne Collaborative Cohort Study
18. Multi-ethnic cohort (White)
19. Multi-ethnic cohort (Japanese)
20. Nurses Health Study NHS
21. Polish Breast Cancer Study
22. Rotterdam Breast Cancer Study
23. Singapore and Sweden Breast Cancer Study
24. SEARCH
25. Seoul Breast Cancer Project n
26. Sheffield Breast Cancer Study
27. Spanish National Cancer Centre Breast Cancer Study
28. US Radiologic Technologist Study

"The detection of further susceptibility loci will require genome-wide studies with more complete coverage and using *larger numbers of cases and controls*, together with the combination of results across multiple studies." Easton et al.

16

Breast cancer: funders

1 Academy of Finland
2 ASTAR Singapore
3 Australian Cancer Council
4 Australian National Breast Cancer Foundation
5 Australian NHMRC
6 Australian VicHealth
7 British Heart Foundation
8 Cancer Council of New South Wales
9 Cancer Council of Queensland
10 Cancer Council of South Australia
11 Cancer Council of Tasmania
12 Cancer Council of Victoria
13 Cancer Council of Western Australia
14 Cancer Foundation of Western Australia
15 Cancer Research Technology
16 Cancer Research UK
17 Center for Cancer Research
18 Copenhagen County and Herlev University Hospital
19 DK Boserup Fund
20 DK Medical Research Council,
21 FI Cancer Society
22 FI Sigrid Juselius Fund
23 Genome Spain Foundation
24 Hannover Medical School
25 Helsinki University Central Hospital Research Fund
26 Intramural Research Program of the NIH
27 Komen Breast Cancer Foundation
28 Leiden University Medical Centre
29 National Cancer Institute
30 New South Wales Cancer Council,
31 Perlegen Sciences
32 PL Nofer Institute of Occupational Medicine
33 PL Sklodowska-Curie Institute of Oncology and Cancer Center
34 UK Academy of Medical Sciences
35 UK Breakthrough Breast Cancer
36 UK Breast Cancer Susceptibility Collaboration
37 UK Department of Health
38 UK Medical Research Council
39 UK Research into Ageing
40 UK Stroke Association
41 UK University of Cambridge
42 US Army Medical Research and Material Command
43 US Department of Health and Human Services
44 Victorian Health Promotion Foundation

17

International Human Genome Sequencing Consortium

1 Chinese Academy of Sciences, Beijing
2 France Genoscope and CNRS UMR-8030
3 Germany Jena Institute of Molecular Biotechnology
4 Germany Max Planck Institute for Molecular Genetics, Berlin,
5 Germany Research Centre for Biotechnology, Braunschweig,
6 Japan Keio University, Tokyo,
7 Japan RIKEN Genomic Sciences Center, Yokohama
8 UK Wellcome Trust Sanger Institute
9 US Baylor College of Medicine
10 US Cold Spring Harbor Laboratory
11 US Department of Energy
12 US Genome Therapeutics Corporation, Mass
13 US Institute for Systems Biology, Seattle
14 US Stanford University
15 US University of Oklahoma
16 US University of Texas Dallas
17 US Washington University
18 US Whitehead Institute/MIT Center for Genome Research

- ~ 6 samples $3billion

18

Why is biobanking important?

- Social justification
  - Key social and scientific justifications coincide
    - Improving human health
    - Unlike non-biomedical research infrastructures
  - Biobanks can be responsive to societal priorities in biomedical research
    - They manage resources to support research into virtually any disease that society defines as a priority
  - Ethical, legal and societal norms in biomedical research are easier to implement with biobanks than with conventional 'private' collections
  - Biobanks will underpin 'personalised medicine'

19

Why is biobanking important?

- Economic justification
  - Research (in science and humanities) is the motor of the UK knowledge economy

Occupations (%)	'84	'94	'04
Knowledge workers	31	36	41
Personal services etc	25	28	28
Skilled/unskilled etc	44	37	30

    - Ref: Working Futures 2004-2014, Work Foundation
  - Life sciences are big job providers
    - the biggest job provider in North West England
  - The most potent investment for economic development is in infrastructure

20

Contents

- Definitions
- Why is biobanking important?
- Types of biobank
- An example
- The growth of biobanking
- A development strategy
- Challenges in biobanking

21

Types of biobank: historical definitions

- Defined by the investigation that required the infrastructure
  - Disease-specific
  - Exposure-specific
  - Population-based

22

Types of biobank: operational definitions

- Sample type
  - Defined by the nature of the resource
    - Blood
    - Plasma
    - DNA
    - Cell lines
    - Fresh frozen tissue
    - Fixed tissue

23

Types of biobank: policy based-definitions

- Primary and secondary biobanks
  - Primary: I manage my samples + data
  - Secondary: I manage your samples + data
- Access arrangements
  - No access: not a biobank
  - Private access (by invitation only): not a biobank
  - Open access to all samples + all data
  - Open access to summary data
  - Fair access to all samples + data

24

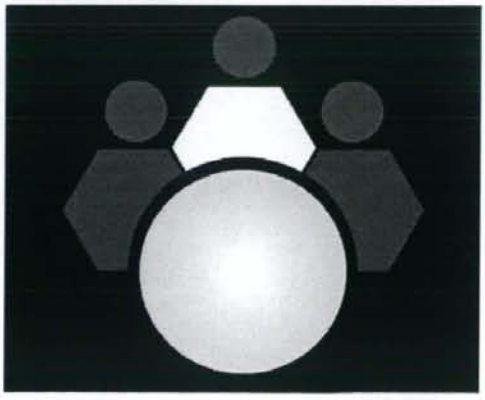
Contents

- Definitions
- Why is biobanking important?
- Types of biobank
- An example
- The growth of biobanking
- A development strategy
- Challenges in biobanking

25

An example

- UK DNA Banking Network (UDBN)
  - A secondary DNA / cell line biobank with fair access



26

UDBN: preparatory phase

- Late 1990s
  - Government allocates funds to MRC specifically for post Human Genome Project translational research
- 2000
  - MRC identifies genetic epidemiology as key
  - MRC issues Call for Proposals for large genetic collections. Grants have special conditions attached.
- 2002
  - MRC funds infrastructure to store and distribute these collections

27

UDBN: construction

DNA

28

MANCHESTER 1824

CIGMR

UDBN

BBMRI

UDBN: construction

DNA
Cells

27

MANCHESTER 1824

CIGMR

UDBN

BBMRI

UDBN: construction

DNA
Cells
IT

Research infrastructure

30

UDBN: construction

- U Manchester
- ECACC*
- Azura

*ECACC = European Collection of Cell Cultures

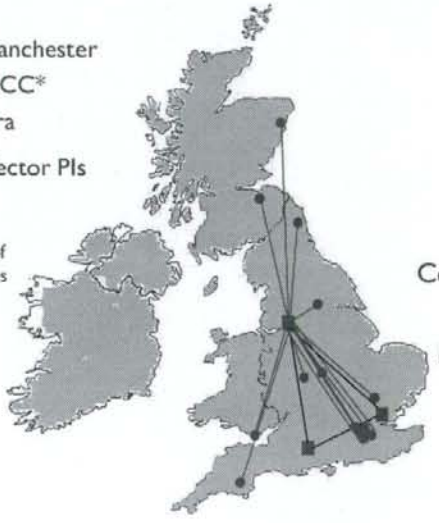
Research infrastructure

31

UDBN: construction

- U Manchester
- ECACC*
- Azura
- Collector PIs

*ECACC = European Collection of Cell Cultures

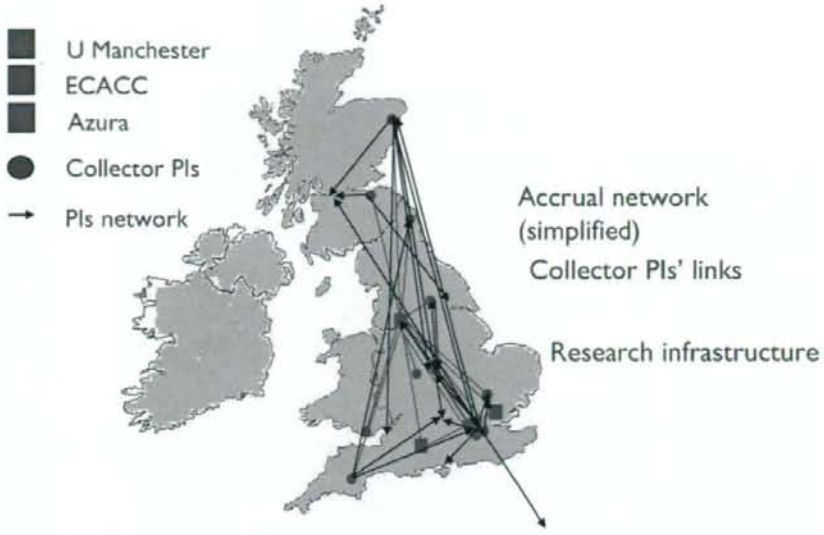


Collector PIs' links

Research infrastructure

32

UDBN: construction



UDBN operations: the basics

- U Manchester, ECACC and Azura are ISO9001:2000 accredited
  - This ensures all operating procedures are standardised and are always followed
    - biannual external inspection
    - must specify quality improvements each time
    - partners must be brought under ISO9001

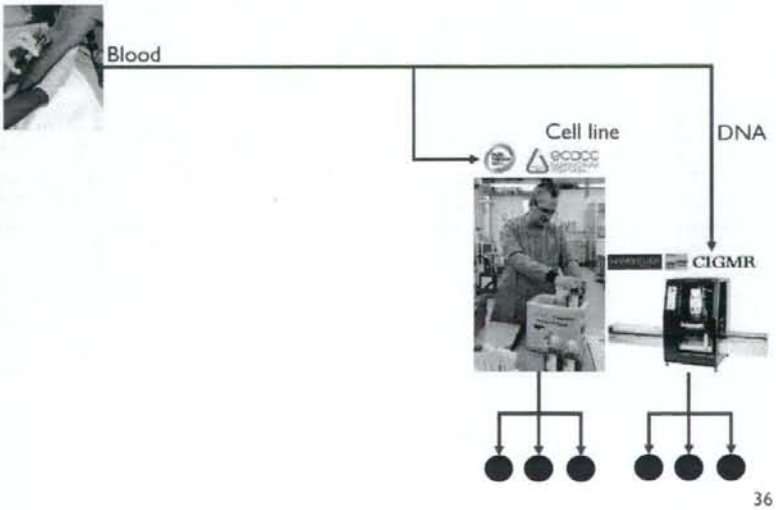
34

UDBN operations: the basics

- U Manchester
  - 12 MRC collections (Acute coronary event; Colorectal cancer; Glomerulonephritis; Parkinson's Disease; Asthma and Eczema; Type 2 Diabetes; Breast Cancer; Hypertension; Multiple Sclerosis; Alzheimer's Disease; Macular Degeneration; Unipolar Depression)
  - 4 additional collections (2 CAD; Motor Neurone Disease; Vesico-Urethral Reflux)
  - new collections (Down's, Osteoarthritis, PKD, Periodontitis, myositis, ALL,...)
  - >26,000 DNA samples
  - 6.10^9 genotypes
- ECACC
  - >15,000 peripheral blood cell samples
  - >7,500 cell lines
- Azura
  - Website, My LabSpace (= Collaboration Management System)
  - Access route for samples and data on all collections

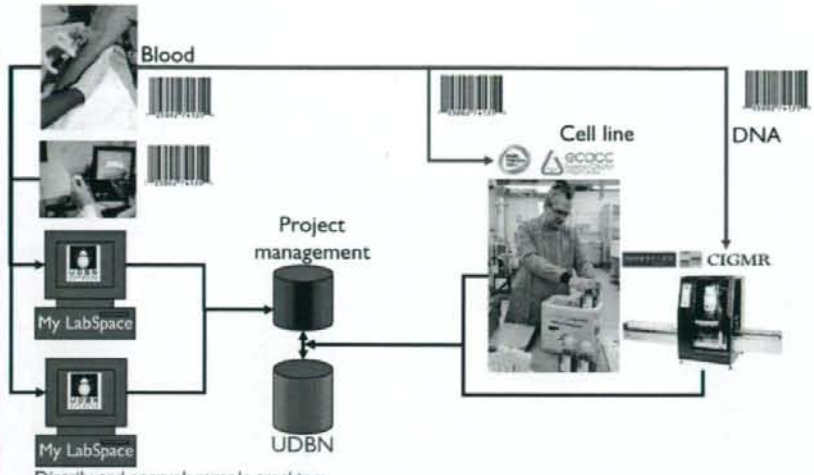
35

Sample management



36

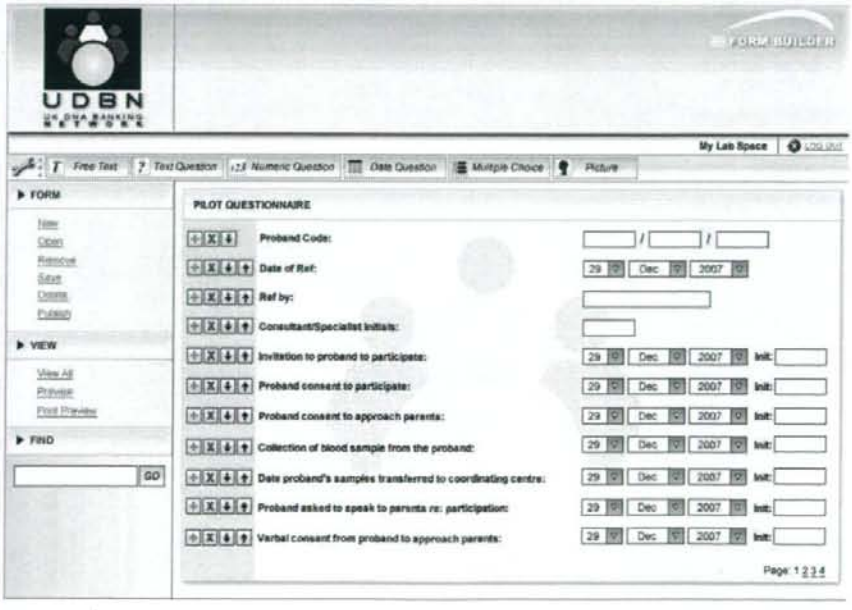
Sample management: accrual



Distributed accrual: sample tracking; consent tracking; phenotype data entry

37

Accrual: inputting data



Sample management: cells

Blood

Cell line

DNA

PBL separation

EBV transformation

CIGMR

LCL expansion

LCL DNA extraction

Replenishment

39

Sample management: DNA

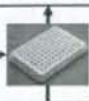
Blood

DNA

CIGMR

DNA extraction

WGA-DNA



Quantitation
Normalisation
Storage (2 sites)
Cherry picking

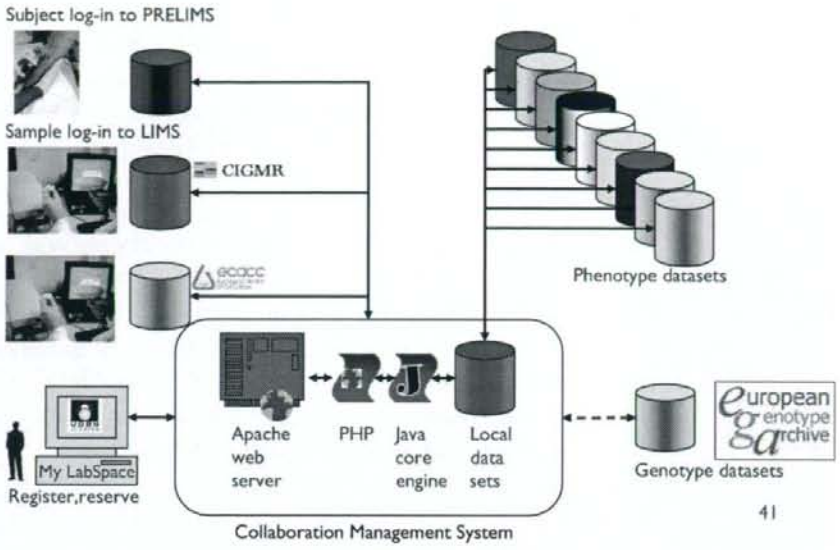
sanger

CNG

illumina

AFFYMETRIX

Data management



41

European Genotyping Archive

- EGA is part of Ensembl
- EGA
  - provides security to allow access rules defined by data depositor
  - is accessible www.ebi.ac.uk/ega/
  - is fully searchable with respect to SNPs, genomic regions, studies (case, population, family) etc
  - supports deposit and retrieval of genotype data in an internationally agreed standard format
  - will support appropriate raw data format in discussion with genotype providers and analysts
  - will exchange data with other appropriate data archive sites
  - will support necessary links to other related biological data resources such as genome browsers, clinical phenotype databases, etc

42

EGA services

- EGA will provide service resources:
  - Complete genotype estimation by imputing missing SNPs
  - Functional markup of genotyped individuals based on known SNP consequences provided by Ensembl
  - A population structure resource with the ability to study characteristics such as admixture
  - Separation of a publicly available list of variants and a restricted access collection of genotypes with potential phenotypic information associated to them
  - Downloads of all data in a single study
  - Queries on specific genome regions, diseases or phenotypes. Data mining based on BioMart.
  - Public data (summaries of information) will be provided through Ensembl and other sources

43

UDBN promotes collaboration

- Web-based process
  1. Registration of new users
  2. User browses summary data on collections
  3. User contacts a collection consortium
  4. Collaboration established
  5. Samples are identified
  6. Approval process
  7. Aliquots despatched
  8. Data returned

44