

Development and perspectives of scientific services offered by genomic biological resource centres

Florian Wagner, Karsten R. Heidtke, Bernd Drescher and Uwe Radelof

Abstract

A number of fundamental technical developments like the evolution of oligonucleotide microarrays, new sequencing technologies and gene synthesis have considerably changed the character of genomic biological resource centres in recent years. While genomic biological resource centres traditionally served mainly as providers of sparsely characterized cDNA clones and clone sets, there is nowadays a clear tendency towards well-characterized, high-quality clones. In addition, major new service units like microarray services have developed, which are completely independent of clone collections, reflecting the co-evolution of data generation and technology development. The new technologies require an increasingly higher degree of specialization, data integration and quality standards. Altogether, these developments result in spin-offs of highly specialized biotech companies, some of which will take a prominent position in translational medicine.

Keywords: services; resource centre; microarrays; synthetic biology; translational medicine; functional genomics

BACKGROUND

During recent decades, there has been an ever-increasing demand for well-characterized materials within the molecular life sciences. Starting with the sharing of, e.g. cell cultures and plasmids, this development exponentially increased with the onset of the Human Genome Project (HGP) in 1990. An early example for a standardized distribution system was the reference library system (RLS), developed by Zehetner and Lehrach [1] at the Imperial Cancer Research Fund in London. The central principles of the RLS were the free distribution of clone sets and of macroarrays (DNA/clones spotted onto nylon membranes) derived thereof, as well as the establishment of a central relational data base system, the Reference Library DataBase (RLDB), for storing all results generated by the use of the reference clones and libraries. The RLDB allowed all users to effectively gain an overview of existing results, to combine the results of their individual experiments in a common database,

and to analyse these data much more efficiently. It also helped to avoid unintended duplication of experiments and therefore double work.

The HGP was accompanied by the foundation of governmentally funded genomic biological resource centres (GBRCs) in a number of countries, e.g. the MRC Geneservice in UK, the RIKEN Genomic Science Center in Japan, and the RZPD German Resource Center for Genome Research in Germany with the mission to provide the international research community with high-quality, well-annotated research materials, especially clones and clone sets. The general challenges met by GBRCs, like quality management, open access, funding and sustainability, have been discussed elsewhere [2]. The basic necessity for a free access to data and materials within the scientific community has equally been pointed out [3], and the majority of scientific journals nowadays insist on compliance with this principle before a manuscript is accepted for publication.

Corresponding author. Uwe Radelof, ATLAS Biolabs GmbH, Friedrichstraße 147, 10117 Berlin, Germany. Tel: +49 (0)30 31 989 660; Fax: +49 (0)30 700 1431 226; E-mail: radelof@atlas-biolabs.de

Florian Wagner is the former head of microarray service unit at ATLAS Biolabs (former head of microarray service unit at RZPD). **Karsten R. Heidtke** is the head of informatics/bioinformatics unit at ATLAS Biolabs (former computer scientist in the department of bioinformatics at RZPD).

Bernd Drescher is the co-founder and CIO of ATLAS Biolabs (former head of the department of bioinformatics at RZPD).

Uwe Radelof is the co-founder and CEO of ATLAS Biolabs (former head of department of production/services at RZPD).

In our review, we want to focus on recent developments and perspectives of scientific services provided by GBRCs, putting a spotlight onto RZPD German Resource Center for Genome Research. RZPD was founded in 1995 as the central infrastructure unit of the German Human Genome Project (DHGP), and was an approved charitable entity since 2000. Its mission was to aid the advancement of genome research by providing public access to a variety of cutting-edge biotechnological services, biological resources and related data without restrictions that arise from intellectual property rights. As one of the most recognized GBRCs, RZPD also became a distributor of the I.M.A.G.E. [4] and the MGC [5] clone collections, the ORFeome clone collection and other important resources. RZPD closed down its operative business on 31 July 2007 based on a shareholders' resolution [Shareholders of RZPD: Max Planck Society (50%), German Cancer Research Center (25%), Max Delbrueck Center for Molecular Medicine (25%)].

FROM POORLY CHARACTERIZED MATERIALS TOWARDS KNOWLEDGE-BASED RESEARCH AND DIAGNOSTIC TOOLS

Within the molecular life sciences, there has been a clear movement from poorly characterized materials towards knowledge-based research and diagnostic tools. More and more data are produced at different levels (e.g. genome, transcriptome, proteome), these data are integrated, and new research tools are generated on this basis. A self-catalysing co-evolution of data/information generation and technology development is speeding up more and more.

A prominent example for this development is the road from analysing transcripts via macroarrays to multifunctional microarrays. Macroarrays—clones or cDNAs spotted in a regular pattern onto a nylon membrane—definitely represented a tremendous technical advancement: for the first time, the parallel investigation of tens of thousands of genes was possible, a prerequisite for the investigation and subsequent understanding of complex genetic networks. However, macroarray experiments suffer from several constraints, e.g. relatively high sample consumption and a low degree of automation. cDNA microarrays, with glass in a standard slide format as supporting material [6], resolved these

problems partially. This format enabled the precise production and processing of many microarrays in parallel with much lower sample volumes. Another tremendous advancement was the *in situ*-synthesis of oligonucleotide arrays [7], a technology combining combinatorial chemistry and photolithography. In contrast to cDNA microarrays dependent on the quality of the clone sets used for production of PCR products spotted onto the array, defined sequences are generated by *in situ*-synthesis, and any arbitrary sequence can be produced. An even higher level of flexibility was reached by the next generation of oligonucleotide array synthesis technology, with digital micromirrors instead of fixed photolithographic masks [8]. This technology allows for the instant programming and production of new arrays, which is very advantageous for research projects requiring a high degree of flexibility. Of course, sequence information is essential to produce oligonucleotide arrays. The sequences of hundreds of prokaryotic and eukaryotic genomes are nowadays available, and whole-chromosome as well as whole-genome arrays are produced, which allow, e.g. for the unbiased mapping of transcriptional activity or transcription factor binding sites.

The complexity of microarray experiments requires a high degree of standardization, and information about each experimental step has to be provided in order to allow for reproducibility. As one of the high-throughput users of this technology, RZPD was actively involved in the definition of 'Minimum information about a microarray experiment' (MIAME) standards for microarray experiments [9].

The technological development outlined above was well reflected in RZPD's portfolio in the array segment. In 2001, RZPD integrated DNA microarray services into its portfolio, which included the production of cDNA microarrays (custom sets as well as indication-specific arrays) derived from RZPD's clone sets as well as services on commercially available oligonucleotide arrays. At that time, Affymetrix' *in situ*-synthesized oligonucleotide arrays already represented the 'gold standard' of array technology, and RZPD became one of the first Affymetrix Authorized Service Providers in Europe. Three years later, it integrated the digital micromirror-based NimbleGen technology into its portfolio.

Starting with 'classical' 3'-based expression arrays in 2001, a number of other applications were added

since then, namely SNP genotyping, copy number analysis, comparative genomic hybridization (CGH), ChIP on chip (chip analysis following chromatin immunoprecipitation) and exon array analysis. In parallel to this portfolio extension, the amount and the complexity of data produced by array experiments grew considerably. A typical macroarray experiment with 55 000 clones spotted onto a nylon membrane generated about 1 MB of raw data, whereas a single Affymetrix expression tiling array with 6.4 millions different oligonucleotides produces 150 MB of raw data (cel-file). For some well-established applications, like 3'-based expression analysis, a number of proven software tools—both commercial and academic—are available. For other, more recent applications like ChIP-chip, data analysis is still at a very early stage, and there is a high demand from customers for data analysis provided by service providers. Consequently, the connections between laboratory and bioinformatics tasks have become very close.

RZPD offered various standard modules for microarray expression data analysis, starting with a primary analysis using Affymetrix' GeneChip Operating System (GCOS). GCOS yields some basic QC parameters (noise, background, spike-in controls), as well as probe set expression values and corresponding *P*-values. Differential expression results (up-/down-regulation, \log_2 -ratio, *P*-value) for pair-wise chip comparisons are also provided by GCOS; however, there is a lack concerning experiment grouping and secondary analyses like clustering or principal component analysis. To this purpose, RZPD created an automated analysis pipeline using well-established and tested statistical methods (*t*-test, ANOVA) as well as generating comprehensive annotations for significantly regulated genes. CEL-files from Affymetrix experiments could be uploaded, or gene/clone tables in an appropriate format (expression values per column). Results were delivered as a pdf-report with graphical representations of the data, i.e. heatmaps, volcano plots, hierarchical clustering (Figure 1) and principal component analysis.

Up-to-date annotations from Entrez, Ensembl, UniGene, Gene Ontology, etc. were delivered in separate MS ExcelTM sheets with dynamic links to the corresponding databases (Figure 2).

The annotation tool was also available as a stand-alone module to annotate probe set lists imported from GCOS in order to replace the out-dated

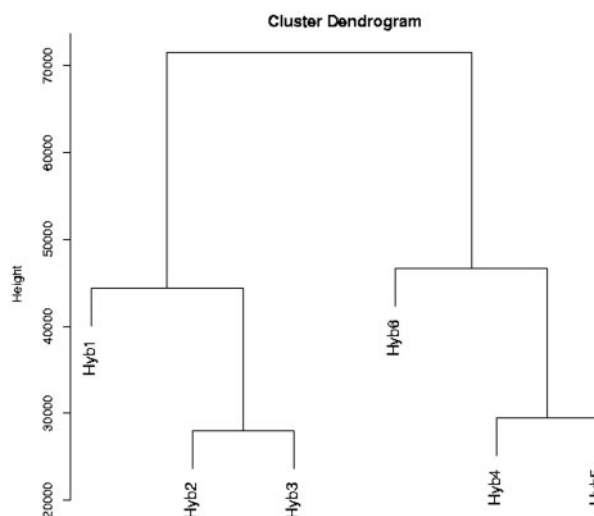


Figure 1: Hierarchical clustering of Affymetrix expression data sets. Hierarchical clustering based on signal intensities shows the degree of similarity between different data sets. In this figure, hybridizations 1–3 and 4–6, respectively, belong to the same experimental group. The height of the *y*-axis at the branching points is a measure of similarity; *y*-axis units are arbitrary.

GCOS annotations. It was based on the bi-weekly updated MASI database (see subsequently).

FROM FUNCTIONAL GENOMICS TO PROTEOMICS

A characteristic feature in molecular life sciences is the remarkable development of research and technologies in the proteome area. As an analogue to HUGO, the 'Human Proteome Organization' (HUPO) was founded in 2001 in order to identify the most important problems in proteomics, to start appropriate initiatives to address these issues, and to support the development of new technologies. Several large-scale proteomics projects have been launched since then, e.g. the Plasma Proteome Project [10], or the Human Antibody Initiative [11].

This development was reflected at RZPD by a coordinated palette of products and services for protein- and proteomics-related research, starting with various expression clones, via the expression and purification of proteins, to the point of functional studies. More than 34 000 full open reading frame (ORF) shuttle clones and nearly 430 000 full ORF expression clones thereof derived were available. RZPD's protein expression service included the choice of the protein of interest from the respective clone collection, including the full ORF

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	AffyId	Stat Pairs	Stat Pairs	Signal	Detection	Detection p-value	Link to RZPD	Uniprot Id	Unigene Id	Ensembl	Entrez Gene	Gene Symbol	Gene Name
2	217733_s_at	11	11	68755	P	0,000244	Products		Hs.496574	ENSG000001094510	9168	TFM5B10	thymosin beta 10
3	200748_s_at	11	11	66712	P	0,000244	Products		Hs.524810	ENSG00000167996	2485	FTH1	ferritin, heavy polypeptide 1
4	212391_x_at	11	11	62613	P	0,000244	Products		Hs.567702 (#188)				
5	201666_at	11	11	62500	P	0,000244	Products		Hs.522632	ENSG00000102265	7076	TIMP1	TIMP metalloproteinase inhibitor 1
6	204892_x_at	11	11	62137	P	0,000244	Products		Hs.522463	ENSG00000156508	1915	EEF1A1	eukaryotic translation elongation factor 1 alpha
7	226131_s_at	11	11	61897	P	0,000244	Products		Hs.397609	ENSG00000106199	6217	RPS16	ribosomal protein S16
8	212185_x_at	11	11	61484	P	0,000244	Products		Hs.418241	ENSG00000126148	4502	MT2A	metallothionein 2A
9	210646_x_at	11	11	61418	P	0,000244	Products	P40429	Hs.546356	ENSG00000142541	23521	RPL13A	ribosomal protein L13a
10	202403_s_at	11	11	61269	P	0,000244	Products		Hs.468142	ENSG00000164632	1278	COL1A2	collagen, type I, alpha 2
11	200817_x_at	11	11	59210	P	0,000244	Products		Hs.538100	ENSG00000124614	6204	RPS10	ribosomal protein S10
12	201254_x_at	11	11	58995	P	0,000732	Products		Hs.408073	ENSG00000137154	6194	RPS6	ribosomal protein S6
13	211542_x_at	11	11	58857	P	0,000244	Products						
14	201257_x_at	11	11	57935	P	0,000244	Products		Hs.356572	ENSG00000145425	6188	RPS3A	ribosomal protein S3A
15	201429_s_at	11	11	57819	P	0,000244	Products		Hs.433701	ENSG00000197756	6168	RPL37A	ribosomal protein L37a
16	213414_s_at	11	11	57815	P	0,000244	Products		Hs.438423	ENSG00000106372	6223	RPS19	ribosomal protein S19
17	208825_x_at	11	11	57740	P	0,000244	Products	P62750	Hs.419483	ENSG00000198242	6147	RPL23A	ribosomal protein L23a
18	AFFX-hum_e	69	69	57475	P	0,000000	Products						
19	208834_x_at	11	11	57354	P	0,000244	Products						
20	213614_x_at	11	11	57158	P	0,000244	Products		Hs.520703	ENSG00000156508	1915	EEF1A1	eukaryotic translation elongation factor 1 alpha
21	200826_at	11	11	56856	P	0,000244	Products		Hs.567333	ENSG00000186468	6228	RPS23	ribosomal protein S23
22	201105_at	11	11	56824	P	0,000244	Products		Hs.445351	ENSG00000100097	3956	LGALS1	lectin, galactoside-binding, soluble, 1 (galectin-1)
23	200963_x_at	11	11	56592	P	0,000244	Products		Hs.469473	ENSG000001071082	6160	RPL31	ribosomal protein L31
24	203034_s_at	11	11	56463	P	0,000244	Products		Hs.523463	ENSG00000166441	6157	RPL27A	ribosomal protein L27a
25	216231_s_at	11	11	56298	P	0,000244	Products		Hs.534255	ENSG00000166710	567	B2M	beta-2-microglobulin
26	212661_x_at	11	11	56214	P	0,000244	Products		Hs.517207	ENSG00000198618	388617		peptidylprolyl isomerase A-like
27	200763_s_at	11	11	56213	P	0,000244	Products		Hs.356502	ENSG00000137818	6176	RPLP1	ribosomal protein, large, P1
28	211956_s_at	11	11	56174	P	0,000244	Products		Hs.150580	ENSG00000173812	10209	EIF1	eukaryotic translation initiation factor 1
29	AFFX-r2-P1	11	11	55902	P	0,000244	Products						
30	213084_x_at	11	11	55846	P	0,000244	Products		Hs.419483	ENSG00000198242	6147	RPL23A	ribosomal protein L23a
31	224187_x_at	11	11	55679	P	0,000244	Products		Hs.180414	ENSG00000108871	3312	HSPAR	heat shock 70kDa protein 8
32	212284_x_at	11	11	55264	P	0,000244	Products		Hs.374596	ENSG00000164116	2982	GLCY1A3	guanylate cyclase 1, soluble, alpha 3
33	200099_s_at	11	11	55275	P	0,000244	Products				54462	KIAA1128	KIAA1128
34	200095_x_at	11	11	55264	P	0,000244	Products		Hs.406620	ENSG00000124614	6204	RPS10	ribosomal protein S10
35	211943_x_at	11	11	55232	P	0,000244	Products		Hs.374596	ENSG00000164116	2982	GLCY1A3	guanylate cyclase 1, soluble, alpha 3
36	212790_x_at	11	11	55174	P	0,000244	Products		Hs.546356	ENSG00000142541	23521	RPL13A	ribosomal protein L13a
37	1553538_s_at	11	11	55125	P	0,000244	Products						
38	212788_x_at	11	11	55067	P	0,000244	Products		Hs.433670	ENSG00000087086	2512	FTL	ferritin, light polypeptide
39	200633_at	11	11	54856	P	0,000244	Products		Hs.356190	ENSG00000170815	7314	UBB	ubiquitin B
40	200869_at	11	11	54838	P	0,000244	Products		Hs.337766	ENSG00000106640	6142	RPL18A	ribosomal protein L18a
41	206687_x_at	11	11	54812	P	0,000244	Products	P11142	Hs.180414	ENSG00000108871	3312	HSPAR	heat shock 70kDa protein 8
42	229563_s_at	11	11	54324	P	0,000244	Products		Hs.546389	ENSG00000198755	4736	RPL10A	ribosomal protein L10a

Figure 2: Annotation of Affymetrix expression data. Affymetrix expression data are provided with a comprehensive, up-to-date annotation from the following data sources: Ensembl, Entrez Gene, Unigene, Uniprot, GeneOntology (data not shown), and biological material available at RZPD. The gene symbol and gene name are also given.

expression clones. A total of 5–10 mg of protein were expressed in an *E. coli* or a Baculovirus expression system and purified up to 90% by FPLC. Three protein array services were available, namely serum screening and antibody epitope mapping using RZPD's protein arrays, and protein expression profiling using Becton–Dickinson (BD) antibody arrays. RZPD's protein arrays were derived from several human tissues including fetal brain, testis and T-lymphocytes. Each protein array consisted of up to 27 648 *E. coli*-expressed proteins, printed in duplicate on a PVDF membrane. The serum screening service on these arrays was designed to screen for auto-antibodies in patient plasma or serum, which was of special interest for clinicians and researchers dealing with autoimmune diseases like diabetes type I or multiple sclerosis. An automated yeast two-hybrid (Y2H) interaction screening service was also available. In contrast to most traditional Y2H methods, a fluorogenic reporter was used that could be accurately quantified [12].

SYNTHETIC BIOLOGY

Synthetic biology emerged as a new research area when the detection and application of restriction

endonucleases nearly 40 years ago [13, 14] enabled the directed construction of DNA sequences that do not occur naturally. Since then, this field has taken a rapid development, and today synthetic biology is defined as '(i) the design and construction of new biological parts, devices and systems and (ii) the re-design of existing, natural biological systems for useful purposes' (definition taken from <http://syntheticbiology.org>). Several groups are aiming at the construction of complete organisms like the bacteriophage T7, in order to test our current genetic knowledge [15]. Within this context, the new large-scale, non-clone-based sequencing methods [16, 17] represent a key technology, as they provide valuable information about naturally occurring genomes.

From a practical point of view, gene synthesis is currently the most important application of synthetic biology, and a number of companies compete in this field. For RZPD, too, the generation of synthetic clones was of increasing importance. Although RZPD offered the most comprehensive clone collection worldwide with 35 millions of clones and more than 1000 cDNA libraries, this collection was far from being complete, even for the most important model organisms. This gap could

be filled with synthetic cDNA clones generated via gene synthesis, which guaranteed 100% sequence identity to a database sequence and allowed for sequence optimization to maximize protein yields in a certain expression system. On account of this, RZPD integrated synthetic clones into its portfolio, which allowed to offer a virtually infinite clone collection. Synthetic clones could be identified and ordered by RZPD's online search tool GenomeCubeTM and were delivered within 4–6 weeks. Gene synthesis also represents an ideal complement to next-generation sequencing. The classical way from the clone to the sequence is simply reverted: the sequence information is generated first—without classical cloning, directly starting from DNA,—and then the clone needed, e.g. for protein expression, is synthesized with 100% compliance to the specified requirements.

BIOINFORMATICS AT RZPD

Structure of RZPD's primary database

During the last years, the structure of RZPD's primary database was completely re-organized in order to meet the requirements of gene-centred queries for biological material, and to provide RZPD's customers with comprehensive and up-to-date annotations for these materials. With respect to different details and information, each major product group was stored in a product database of its own, e.g. clones, clone sets and libraries, tissue arrays, etc., where each product had one or more references to genes. These references were the basis to connect the product data with all kinds of annotations, e.g. sequence data, functional annotation, disease-related information, etc. The annotation was retrieved from different sources. On one hand, the submitters of biological material provided basic information with their submission. On the other hand, RZPD's bioinformatics group spent an enormous effort to improve the level of annotation with data from public resources. Data from Uniprot, Unigene, OMIM, Ensembl, Entrez, GeneOntology, KEGG, BioCarta and many more had to be interlinked with each other, the product databases, and the RZPD services. The resource for up-to-date data annotations was MASI@RZPD. MASI (developed by Insilico Software GmbH), 'Meta Annotated Sequence Investigation', is a compilation of public biological databases into a locally available relational database management system (RDBMS).

The increasing spectrum of information provided covers all relevant sources of genetic, proteomic and metabolomic data. Therefore, particular biological entities cited at different sources became comparable and thus provided more comprehensive information. MASI was updated on a bi-weekly basis from the public resources.


Product search and ordering with the GenomeCubeTM

The GenomeCubeTM was an intuitive-to-use interface for gene-based material retrieval. It was named after the three 'dimensions' required to identify the most appropriate resources for the gene(s) of interest: gene, species and clone/product type. Instant access was given to validated materials, e.g. sequence-verified cDNA clones, full length clones, full ORF clones, siRNA resources, etc. by entering a variety of search terms, e.g. gene symbols and descriptions, GenBank Accession numbers, Unigene Ids, Ensembl Ids and Affymetrix Ids. The result of a GenomeCubeTM query was a list of appropriate products (Figure 3).

Comprehensive information could be obtained by a single mouse click, e.g. gene information, functional annotation details, disease-related information, sequence data and many more. Nevertheless, the search for appropriate biological material was not always straightforward. Therefore, to find biological material that was related to each other by the same genetic context, a new concept named 'GeneContext' was introduced into the GenomeCubeTM. Genes were represented in (i) Positional Gene Sets, (ii) Motif Gene Sets, (iii) Curated Gene Sets, (iv) Computational Gene Sets provided by the Broad Institute laboratories (http://www.broad.mit.edu/gsea/msigdb/msigdb_index.html), (v) the controlled vocabulary used for indexing articles for MEDLINE/PubMed (MeSH Terms) and (vi) Disease Concepts provided by the Autoimmune Disease Database (<http://www.sbi.uni-rostock.de/aidb/home.php>).

CENTRAL RESEARCH INFRASTRUCTURE FOR MOLECULAR PATHOLOGY (CRIP)

Biobanks have recently been developed under the auspices of public or private initiatives as suitable tools for biomedical research purposes. Therefore, at RZPD a central research infrastructure for molecular

Search 

for

Page 1 / 1

Search Result :

Your Entry	No	Gene	Description	Organism	Hit	Product Type	Product	Select	Price
MAOA	1	MAOA Gene Info Gene Context All Products	Monoamine oxidase A	Homo sapiens	RZPD ORF Expression Clone		EX-RZPD0839E07157 (w/o stopcodon)	<input type="button" value="Go"/>	625.00 Euro
							EX-RZPD0839E07158 (with stopcodon)	<input type="button" value="Go"/>	625.00 Euro
							EX-IOH5847 (with stopcodon)	<input type="button" value="Go"/>	625.00 Euro
							OmicsLink ORF Expression Clone EX-A0833 (with stopcodon)	<input type="button" value="Go"/>	634.00 Euro
							Full ORF Shuttle Clone		
							RZPD0839E07157D - Gateway (w/o stopcodon)	<input type="checkbox"/>	500.00 Euro
							RZPD0839E07158D - Gateway (with stopcodon)	<input type="checkbox"/>	500.00 Euro
							IOH5847 - Gateway (with stopcodon)	<input type="checkbox"/>	500.00 Euro
							Full Length cDNA Clone		
							IRATp970D09103D	<input type="checkbox"/>	89.00 Euro
							• EST sequence verified	<input type="checkbox"/>	124.00 Euro
							esiWay Silencing Resource		
							RZPDp3000E0835D	<input type="checkbox"/>	95.00 Euro
EST Clone (sequence-verif.)									
HU3_p983A0433D	<input type="checkbox"/>	63.00 Euro							
Genomic Clone									
RZPDB737B0929D(-10kb:complete, gene:partial)	<input type="checkbox"/>	63.00 Euro							
2	Maoa Gene Info All Products	Monoamine oxidase A	Mus musculus	Genomic Clone			RPCIB731B08186Q(-10kb:complete, gene:complete)	<input type="checkbox"/>	39.00 Euro

Figure 3: RZPD's GenomeCube™. The GenomeCube™ offered direct access to a multitude of validated clone materials corresponding to the search term provided. Here, 'MAOA' (monoamine oxidase A) was entered, for which a variety of different clones from several organisms was available.

pathology (CRIP) was established, which can be used as a portal to different tissue banks.

CRIP (<https://crip.rzpd.de>; from July 2007: www.crip.fraunhofer.de) displays anonymized data on human tissue samples (diseased and normal) and corresponding biospecimens available for research projects in its database partners' institutes. CRIP covers all disease areas and provides information on sample preparation (fresh frozen, paraffin embedded, etc.) according to standard operating procedures, clinical and follow up data and patient's written informed consent. Partner institutes are academic Institutes of Pathology.

SUMMARY AND PERSPECTIVES

A number of fundamental technical developments within the molecular life sciences induced a distinct

change in RZPD's portfolio in the last years. There was a clear tendency away from poorly characterized cDNA clones towards high-quality, well-annotated clones (e.g. full ORF and expression clones) to be used in functional genome research. Synthetic clones, which can be synthesized according to researchers requirements, started to replace natural clones, a process which is speeding up also catalysed by next-generation sequencing technologies. In the microarray segment, cDNA clone collections were no longer on offer. cDNA arrays were replaced by *in situ*-synthesized oligonucleotide arrays from commercial manufacturers, which cover more organisms and allow for a much broader portfolio of applications (e.g. whole genome expression analysis, differential splicing analysis, ChIP-Chip, re-sequencing, etc.). Moreover, the quality of *in situ*-synthesized oligonucleotide arrays is clearly superior

to those of home-made cDNA arrays [18], and prices of oligonucleotide arrays have dropped by several hundred percent in the last 5 years. Services for proteomics research were established, like protein expression and purification, protein array generation and related screening services, as well as a Y2H screening service. The amount and complexity of data generated by RZPD's services grew considerably, leading to an increased demand for data storage, analysis, integration and information gathering. This demand was met by the offering of various, modular data analysis tools and services, which represented a combination of in-house developments and commercial software. Here, it becomes obvious that today in addition to physical and infrastructural resources the term 'resources' includes software tools and the knowledge about their reasonable application.

In summary, RZPD's original character as a non-for-profit provider of clones and clone sets changed considerably. Although, especially high-quality clones still formed an important segment of its portfolio, other service units like microarray services became completely independent from the clone segment. To a certain degree, this was also true for protein-related services, which were not necessarily dependent on the availability of (expression) clones within the same entity.

Another tendency, which has an effect especially in the microarray segment, is the ongoing development of DNA chips for clinical applications, especially for molecular diagnostic purposes. Prominent examples are Roche's AmpliChip[®] CyP450 or Agendia's MammaPrint[®] breast cancer diagnostic test. These microarrays are not designed to serve basic research, but to answer diagnostic questions. Hence, they are used preferably by clinicians, and not by academic research groups. A company which wants to provide services in the molecular diagnostics segment will have a different client base and has to cope with other challenges like higher quality standards and a very close interaction with physicians.

Altogether, these developments supported tendencies towards privatization of independent and profitable service units like the microarray service unit or a protein-based services unit, a trend not unique to RZPD, as exemplified by the foundation of GeneService Ltd as a spin-off from the former MRC GeneService in 2004. Of course, the basic principles of GBRCs like free access or the

sustainable archiving of research materials cannot be executed in a private entity for some obvious reasons [2]. The era of classical, non-for-profit GBRCs seems to come to its end, and highly specialized biotech companies focusing on clearly defined business segments spin out in order to take their place. RZPD's successor companies will focus on technologies needing profound expert knowledge as well as expensive and highly specialized equipment, as for the vast majority of academic research groups it does not seem to be reasonable for economic and scientific reasons to establish these technologies in their own laboratory. The microarray segment definitely falls within this category, a field that is already very well acknowledged within the scientific community, but nevertheless is still rapidly evolving and holding an enormous potential for future applications, like, e.g. gene expression analyses from formalin-fixed, paraffin-embedded samples, or the development of DNA chips for clinical applications. Also, large-scale sequencing projects using ultra-high throughput platforms like Roche-454 or Illumina-Solexa are pre-disposed for outsourcing to companies providing the necessary equipment as well as bioinformatics expertise.

Key Points

- The original character of genomic biological resource centres as providers of clones and clone sets has changed considerably by the integration of fundamentally new technologies like oligonucleotide microarrays or gene synthesis, and by the buildup of proteomics-related services.
- Within the clone segment, there is a clear tendency towards well-annotated, high quality clones, which might be mainly of synthetic origin in the near future.
- Major service units like microarray services are completely independent from clone collections, and they require a high degree of specialization and data integration.
- The era of classical, non-for-profit GBRCs seems to come to its end, and highly specialized biotech companies focusing on clearly defined business segments spin out in order to take their place.

Acknowledgement

The authors wish to thank Christina Schröder for reading parts of the article.

References

1. Zehetner G, Lehrach H. The reference library system—sharing biological material and experimental data. *Nature* 1994;**367**:489–91.

2. Weaver T, Maurer J, Hayashizaki Y. Sharing genomes: an integrated approach to funding, managing and distributing genomic clone resources. *Nat Rev Genet* 2004;**5**:861–66.
3. Cech TR. Sharing publication-related data and materials: responsibilities of authorship in the life sciences. Available at www.nap.edu/books/0309088593/html (27 September 2007, date last accessed).
4. Lennon G, Auffray C, Polymeropoulos M, et al. The I.M.A.G.E. consortium: an integrated molecular analysis of genomes and their expression. *Genomics* 1996;**33**:151–52.
5. Strausberg RL, Feingold EA, Klausner RD, et al. The mammalian gene collection. *Science* 1999;**286**:455–57.
6. Schena M, Shalon D, Davis RW, et al. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995;**270**:467–70.
7. Fodor SP, Read JL, Pirrung MC, et al. Light-directed, spatially addressable parallel chemical synthesis. *Science* 1991;**251**:767–73.
8. Singh-Gasson S, Green RD, Yue Y, et al. Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nat Biotechnol* 1999;**17**:974–78.
9. Brazma A, Hingamp P, Quackenbush J, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 2001;**29**:365–71.
10. Omenn GS, States DJ, Adamski M, et al. Overview of the HUPO plasma proteome project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* 2005;**5**:3226–45.
11. Nilsson P, Paavilainen L, Larsson K, et al. Towards a human proteome atlas: high-throughput generation of mono-specific antibodies for tissue profiling. *Mol Cell Proteomics* 2005;**5**:4327–37.
12. Albers M, Kranz H, Kober I, et al. Automated yeast two-hybrid screening for nuclear receptor-interacting proteins. *Mol Cell Proteomics* 2005;**4**:205–13.
13. Arber W, Wauters-Willems D. Host specificity of DNA produced by *Escherichia coli*. XII. The two restriction and modification systems of strain 15T. *Mol Gen Genet* 1970;**3**:203–17.
14. Danna K, Nathans D. Specific cleavage of simian virus 40 DNA by restriction endonuclease of *Hemophilus influenzae*. *Proc Natl Acad Sci U S A* 1971;**68**:2913–17.
15. Chan LY, Kosuri S, Endy D. Refactoring bacteriophage T7. *Mol Syst Biol* 2005;**1**, 2005.0018, Epub 2005 Sep 13.
16. Margulies M, Egholm M, Altman WE. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005;**437**:376–80.
17. Bennett ST, Barnes C, Cox A. Toward the 1,000 dollars human genome. *Pharmacogenomics* 2005;**6**:373–82.
18. Shi L, Reid LH, Jones WD. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* 2006;**24**:1151–61.