

SMP: Cell

Project: THE GERMAN cDNA CONSORTIUM: MAINTENANCE OF ORF AND EXPRESSION CONSTRUCTS

Stephanie Bechtel – Deutsches Krebsforschungszentrum (DKFZ), Heidelberg – s.becht@dkfz.de

Introduction

The SMP-Cell follows a systematic strategy towards the identification and functional validation of novel genes with the aim to validate proteins of clinical relevance and to select targets for diagnostics and therapy. After gene modeling, protein coding regions are PCR amplified and cloned using the Gateway® cloning technology (1, 2). This allows for protein expression in the cellular functional gene analysis projects of SMP-Cell and in collaborating projects. Prior to their functional exploitation, all entry clones are sequence verified as one most essential step of quality control, only validated ORF resources are processed further into GFP-tagged expression vectors and are subsequently used in the functional genomics pipeline within SMP-cell (3) and by collaborating SMPs and KGs.

The generation of the ORF resources (Gateway entry clones) and the subsequent functional analysis of the encoded proteins are performed by different partners within SMP cell. The central collection and maintenance of ORF clone resources ensures their permanent availability to the partners within SMP-cell and to collaboration SMP and KGs, thus assuring a successful networking within NGFN-2.

Project Status

The cloning and sequencing of the ORFs is distributed within the German cDNA Consortium (4) between several partners of SMP-Cell, some doing both cloning and sequencing (AGOWA GmbH Berlin, DKFZ Heidelberg, Qiagen GmbH Hilden, and the BMFZ of the University Düsseldorf), while

others either clone (RZPD Heidelberg) or sequence (GBF Braunschweig and Medigenomix Munich) only. All the respective partners have long proven capabilities in the large-scale application of the necessary technologies.

The DKFZ centrally collects, arrays and maintains the ORF cloning templates that are identified by systematic modeling of full-length genes and ORFs, and distributes them to the partners who then perform the cloning of ORFs (Fig. 1). Several parameters required standardization due to the decentralized generation of the ORF resources and subsequent exploitation in cell-based experiments (Fig. 1): i) All entry clones and other material are named following a common nomenclature which allows for the tracking of any data that is collected in cellular functional gene analysis back to the ORF and clone they derived from; ii) the clone-IDs are automatically generated to avoid possible errors from manual typing. This is done with help of the LIFEdb LIMS system (5). This system ensures that all material receives unique identifiers and thus circumvents inconsistencies; iii) The quality-controlled entry clones are centrally collected and maintained in the repository of the DKFZ where they are replicated and stored in 96-well plates and on IsoCode® cards (Schleicher & Schuell) (Fig. 2). Stock keeping is mirrored *in silico*, again through the LIFEdb database system (5). This system currently contains 3,500 entry clones from 1,400 different ORFs, most are sequenced to completion. The entry-clone resource is provided to the partners within SMP-cell and to collaborating SMPs and KGs. In addition, the clones are commercialized through the RZPD.

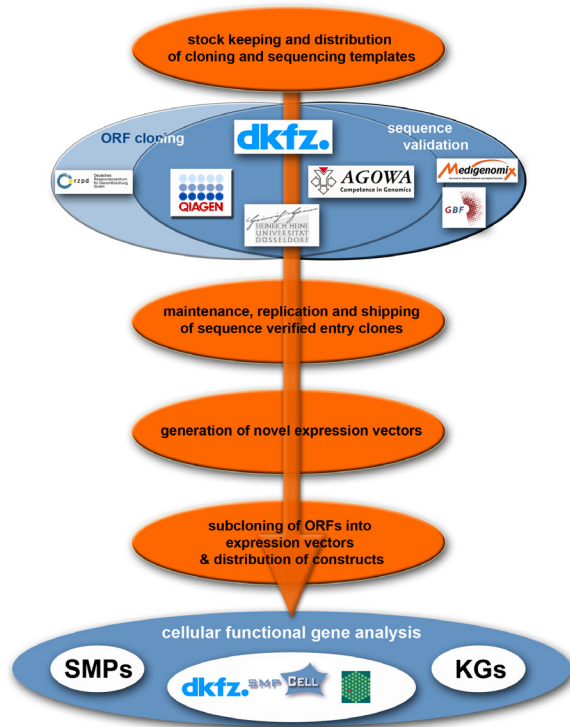


Fig 1: Schematic presentation of responsibility assignment towards the generation, maintenance and usage of quality-controlled clone resources in SMP-cell. Tasks in red are carried out in this project exclusively at the DKFZ.

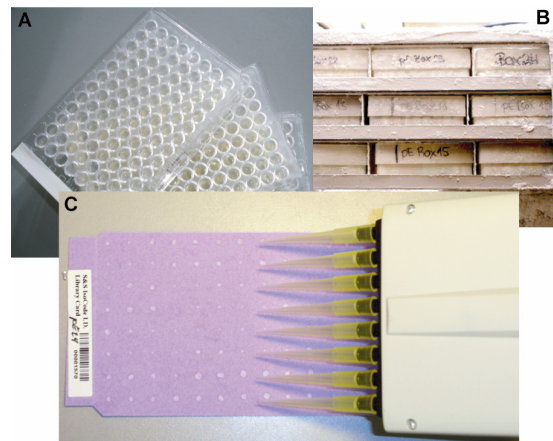


Fig 2: Clone storage in SMP-Cell: (A) Glycerol stocks are kept in 96-well plates, (B) and (C) bacterial cultures are spotted on IsoCode® cards.

ORFs are centrally subcloned at the DKFZ from entry clones into a variety of expression vectors (Fig. 3) for utilization within SMP-Cell and beyond. N- and C-terminally tagged YFP expression vectors are targeted by default as these constructs are utilized in subsequent screening of encoded proteins in cellular functional gene analysis projects at the DKFZ and the EMBL. The respective expression vectors are either generated in this project (1), or obtained from commercial sources. The vectors are distributed to scientists worldwide, maps and sequences are published at <http://www.smp-cell.org/groups.asp?siteID=48>.

SOPs have been established to standardize the cloning process. These are disseminated via the SMP-Cell webpage (<http://www.smp-cell.org>). Expression clones are quality controlled by restriction digest for the presence of ORFs of the expected size. Over 6,500 different expression constructs have thus far been generated, mostly in mammalian expression vectors as fusions with derivatives of GFP. As for entry clones the generation of unique expression clone IDs that provide information on the respective entry clone and expression vector are automated. All expression clones are stored as shown in Fig. 2, plate and filter coordinates are kept in the LIFEdb database for the tracking of individual clones.

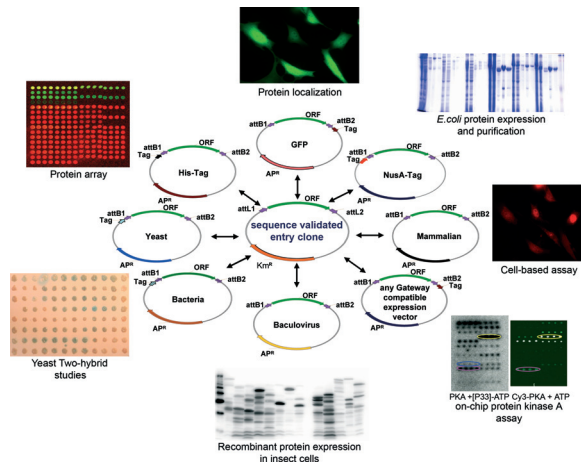


Fig 3: schematic view of the shuttling of ORFs from sequence validated entry clones into a variety of expression vectors. The resulting expression clones allow for a range of expression systems and applications in functional genomics and proteomics.

Maintenance of expression clones is also done centrally at the DKFZ, clones are distributed to collaborators within SMP-cell and from other SMPs or KGs (Fig. 1). This project thus generates and keeps a quality controlled and standardized resource that is made widely available.

Outlook

The continuously extended physical resource of sequence-verified entry clones that is generated by the joint effort of the German cDNA Consortium in NGFN-2 is centrally stored, maintained and distributed by a repository at the DKFZ. Any new ORFs are centrally subcloned into expression vectors to extend the resource of expression clones for use in the functional analysis pipeline within SMP cell. In collaborative efforts with other SMPs and with KGs a growing number of genes and ORFs is analyzed, where this project again provides the resources for immediate functional exploitation. Once candidate proteins have been identified in cellular functional gene analysis, the detailed functional characterization requires further expression constructs to be generated in order to allow for the expression of the proteins e.g. under control of another promoter, without a protein tag, or in an different expression system. In this line further expression vectors will be developed according to the desired application, and the respective ORFs are subcloned into these vectors.

Lit.: 1. Simpson JC *et al.* Systematic subcellular localization of novel proteins identified by large scale cDNA sequencing. *EMBO Rep.* 2000 1(3):287-92. 2. Hartley JL *et al.* DNA cloning using in vitro site-specific recombination. *Genome Res.* 2000 Mar;10(11):1788-95. 3. Wiemann S *et al.* From ORFeome to biology: a functional genomics pipeline. *Genome Res.* 2004 14(10b):2136-44. 4. Wiemann S *et al.* Toward a Catalog of Human Genes and Proteins: Sequencing and Analysis of 500 Novel Complete Protein Coding Human cDNAs. *Genome Res.* 2001 Mar;11(3):422-35. 5. Bannasch D *et al.* LIFEdb: A database for functional genomics experiments integrating information from external sources, and serving as a sample tracking system. *Nucleic Acids Res.* 2004 32(1):D505-8.