Potential and challenges of a human cytome project

G. VALET [1], A. TÁRNOK [2]

[1] Max-Planck-Institut für Biochemie, Martinsried, Germany
[2] Pediatric Cardiology, Heart Center Leipzig GmbH, University Hospital Leipzig, Germany

ABSTRACT: ***Background*** **- The elucidation of the molecular pathways from the 20-40.000 genes of the sequenced human genome via investigation of genetic networks and molecular pathways up to the cellular and organismal phenotypes is highly complex and time consuming.**
Strategy and goals **- The proposed upside-down research strategy of a human cytome project accesses the expressed molecular cell phenotypes by differential screening, for example of diseased versus healthy, or undifferentiated versus differentiated cells to obtain information on disease or differentiation related molecular hotspots at the single cell level. The genome serves as inventory of the biomolecular capacities of organisms while the mechanisms of genome realisation are initially entirely bypassed. Detected molecular hotspots are further investigated by backward directed systems biology, including molecular pathway modelling to elucidate disease related molecular pathways. New drug targets may be identified to specifically influence such pathways. Differential screening provides, in addition, individualized disease course predictions for everyday medicine, in form of "predictive medicine by cytomics." The early recognition of future disease complications enables an immediate application of preventive therapies. This is likely to lower disease related irreversible tissue destruction and adverse drug reactions and will allow to individually optimize patient therapy.**
Outlook **- Immediate medical use, facilitated access to the detection of new drug targets, increased research speed and the stimulation for advanced technological developments represent major driving forces for the efforts to establish a human cytome project. (J Biol Regul Homeost Agents 2004; 18: 87-91)**

KEY WORDS: *Cytome, Cytomics, Predictive medicine, Molecular hotspots, Differential screening*

Received:
Revised:
Accepted:

INTRODUCTION

The genome sequence provides information on the biomolecular capacity of organisms but presently does not explain the observed structural and functional multilevel biocomplexity of cells and cell networks (cytomes). An enormous number of hypotheses and experimental work will be required to experimentally explore and to mathematically model the many interrelations of highly redundant molecular pathway systems. This is especially true when starting at the genome level with genes as elementary information units through genetic networks and protein pathways to finally understand cellular and organismal phenotypes (1).

Problems of multilevel biocomplexity

The observed multilevel biocomplexity concerns for example the distinction of 3D protein structures from amino acid sequences as well as the search for new molecular drug targets. The prediction of 3D protein structures from known amino acid sequences containing the 20 most common amino acids (2) is still not exactly possible after more than 30 years of intensive research in this area, despite the substantial progress in computing potential and software development. The uncovering of the combinatorial complexity of metabolic pathways from 20-40.000 gene products at the cellular level in highly heterogeneous cell populations will be much more demanding. Concerning the pharmaceutical industry, the substantially increased investments for the detection of new drug targets during the last 10 years have resulted in less clinically applicable new candidate substances than in the preceding 10 year period (3,4). This development can be seen as an economic transcript of the very high existing biocomplexity.

Upside-down research strategy as an alternative

This experience raises the demand for more efficient research strategies. It seems particularly important to concentrate, in a systematic approach, on the collection of molecular information from single cells. Single cells represent elementary building units of "cytomes" such as tissues, cell systems, organs and organisms. Diseases emerge as a consequence of molecular changes at the cellular level. Multiparameter flow and image cytometry, high content (HCS) and high throughput (HTP) screening allow the simultaneous collection of a multitude of

© Wichtig Editore, 2004

molecular cell phenotype information from single cells. The genome sequence information can be used in this context as an inventory of the biomolecular capacity of organisms (Fig. 1). This permits initially to entirely bypass the molecular pathway complexity of genome realization. Interpretation problems arising from averaged results of analyzed cell and tissue homogenates are avoided by collecting a maximum of information on the characteristic cellular heterogeneity of cytomes at the single cell level. Cytomics as the analysis of molecular single cell phenotypes in combination with exhaustive bioinformatic knowledge extraction are an essential feature of this approach.

Differential screening like diseased versus healthy or differentiated against undifferentiated cytomes reveals disease or differentiation associated molecular hotspots in the molecular cell phenotypes as they result from genotypic and exposure influences.

The single cell, upside-down oriented research strategy uses molecular cell phenotype differentials to elucidate disease or differentiation related pathways backwards by molecular reverse engineering. Affected molecular pathways will in this way become accessible to systems biology and molecular pathway modeling. Speed and intellectual economy represent major advantages of the single cell based research strategy. The exhaustive bioinformatic information and knowledge extraction seems manageable at this stage of technological development.

Multiparametric single cell characterization

Flow cytometers constitute routine single cell analysis equipment in hospitals while modern image analysis instrumentation especially in fluorescence imaging is still on its way to clinical routine. Individual cells are measured once in a flow cytometer with the options to image cells in flow (5) or to preparatively separate them according to predetermined parameter profiles by a cell sorter module. Image analysis systems (6, 7), in contrast, can relocate cells which is essential for repeated subsequent cell stainings. Cells may be initially stained for cell functions like intracellular pH, transmembrane potentials or Ca2+ levels. They are then fixed to remove the functional stains and restained for specific extra- or intracellular constituents such as antigens, lipids or carbohydrates. After the second destaining, specific nucleic acids may be stained. Multispectral imaging (5, 8) as well as serial optical or histological sections permit 3D-reconstructions of multiparametric molecular morphologies of cell membrane, nucleus, organelle and cytoplasmic compartments including the parametrization of 3D-shapes. Such data are useful for the standardized analysis of proximity and interaction patterns of intracellular structures like nucleus and organelles as well as of different cell types within the tissue architecture (for more details see reference 9).

Information and knowledge extraction

Traditional visual and quantitative evaluations of two or three dimensional cytometric histograms, like in flow cytometry, collect only a very limited amount of the available information and one is never certain whether the really relevant information has been captured. Experience has also shown that quality controlled consensus strategies for multiparameter data evaluation are not easy to develop and there is little pre-existing interpretation knowledge on very complex multiparameter data spaces. Essential information may therefore be lost due to the lack of awareness.

The use of automated, self adjusting evaluation strategies for the systematic collection of the entire information content of all measured cells for the subsequent knowledge extraction is therefore important. Biophysical parameters like light scatter or cell volume signals as well as the presence of DNA or certain antigens like CD45 on leukocytes can be used as gating parameters to assure information collection from more than 95% of the measured cells. Besides relative and absolute cell frequencies, it is essential to automatically evaluate fluorescence intensities of cell populations as well as the relative packing densities of the specifically labeled biomolecules in form of relative mean surface density for cell membrane components and relative concentrations for molecules in the cell interior. The calculation of averaged cellular parameter ratios and coefficients of variations of the parameter value distributions represent essential complements for the determination of the heterogeneity of cell populations. Information extraction generates sometimes several thousand database columns per cell sample.

Knowledge extraction from this information by mathematical or statistical methods may require assumptions on the mathematical distribution of parameter values or generate problems with occasionally missing values. Algorithmic data sieving (10, http://www.biochem.mpg.de/valet/classif1.html) as a mathematical assumption free alternative provides highly discriminatory predictive or diagnostic data patterns from thousands of data columns. The analysis is suitable for parallel computing but also for the processing of multiparametric data from image analysis.

Relational data classification

Multiparametric flow cytometers or microscopes represent complex instrumentation. As a consequence, two instruments built with the same parts will not provide identical results on a given sample. This is due to existing tolerances in the multitude of electronical and optical components of such instruments. Fluorescence and light scatter signals are measured on relative scales and cell population oriented histogram gating procedures remain to some extent arbitrary. The majority of these accuracy errors are cancelled by differential screening. The relational

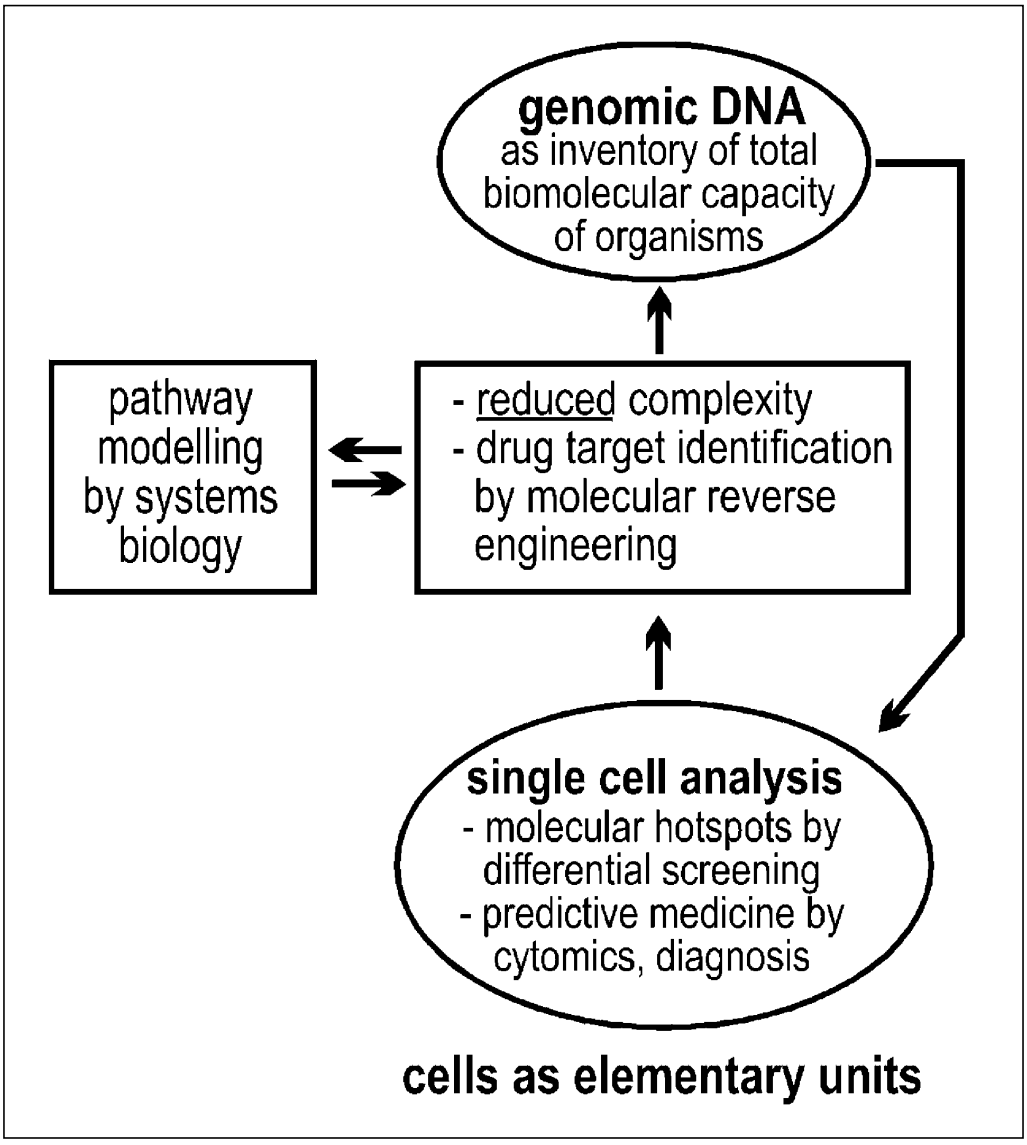

Fig. 1 - Research concept for a human cytome project.

expression of parameter values as fraction of the respective parameter means of a reference group conserves their relative scalar positions including the coefficients of variation as indicators of the dispersion of parameter values. The results are interlaboratory standardized in this way permitting data merging between laboratories. The composition of reference groups is open to prior consensus and results from reference groups of different laboratories can be checked against each other for classification identity. Identity is given when various reference groups are indistinguishable from each other upon relational data pattern classification (11).

A relational system for the objective molecular description of diseases and elementary cellular states like differentiation, maturation, divisions at the cellular level can be established as a cell type, disease or cell differentiation oriented molecular framework. Different cell types will be in a standardized relation to each other in a kind of periodic system of cells.

Predictive medicine by cytomics

This approach is of immediate interest for everyday medicine. Collected single cell and other patient information can be relationally analyzed for future disease course such as; improvement, deterioration or stationary state, complications or adverse drug reactions (ADR) depending on the envisaged therapies. The immediate application of preventive therapies in cases of foreseeable deterioration may avoid irreversible tissue loss, being potentially associated with defective healing processes as cause of lifelong health deterioration or infirmity.

A framework of disease related molecular alterations can be established in the form of discriminatory data patterns at different levels of the genome realization for example at the mRNA, proteome or molecular cell phenotype levels. Parameter numbers in the order of up to 10^5 or more can currently be processed for the elaboration of discriminatory data

patterns, each typically containing between 10 to 30 parameters (12, 13). The essential features of this research concept (9, 10, 14) are the use of molecular data patterns instead of metabolic pathways, the cell to patient approach and the standardized description of cell and disease states in a relational molecular framework.

Individualized disease course predictions according to the predictive medicine by cytomics concept (12, 13) are different from patient group oriented prediction of prognosis (15, 16) like for example Kaplan-Meier survival or therapy response curves. Such curves are of recognized importance for the development of new therapies by multicenter trials. Prognostic results are, however, of little value for the individual patient and the treating physician. As a result of non individualized therapy schemes, a substantial number of patients has actually no benefit from therapies and may even suffer from ADRs.

The predictive medicine by cytomics concept has the potential to overcome this problem by an individually optimized patient therapy including the use of therapeutic lead-times for preventive therapies with the goal of minimizing irreversible tissue losses for example. The goal may also be to suppress disease declaration as in the case of potential screenings for the degree of sensitization for asthma in at risk families.

Potential and challenges

In this situation, a human cytome project can simultaneously advance the biomedical sciences in two important directions. The individualized disease course predictions and the pretherapeutic identification of high risk patients are of immediate use in everyday medicine. This is likely to improve the general efficiency of health care. At the same time, the initial bypassing of the complex mechanisms of genome realization by differential molecular cell phenotype screening for medical purposes, represents a self focusing mechanism for the identification of disease related molecular hotspots. New drug targets may be detectable through the subsequent reverse engineering strategy by functional genome analysis and systems biology.

The focused effort of a human cytome project will profit from substantial quantities of already existing cytometric single cell data as well as clinical chemistry and other patient data being available in many clinical institutions. New studies do not require sampling beyond established ethical criteria since sufficiently informative single cell analysis can be performed within the 1.000-100.000 cell range amounting to microliters or mg requirements as available in typical blood or biopsy specimens.

Major challenges of a human cytome project concern the development of specially adapted cytometric instrumentation with regard to automated sample preparation, staining, measurement and information extraction. Miniaturization of flow cell, light sources and photon capture (17) are further important goals in view of the large scale uses of this technology, especially for flow cytometry as point of care or even home version technology for the early detection of complications in risk patients. Such instrumentation will require the development of sensitive multispectral nanoparticle (18) or other labeling reagents to provide suitable reagent kits for predictive and diagnostic demands in general medicine.

There is also a significant need for bioinformatic software development concerning fast and efficient multiparameter data analysis and knowledge extraction. This is especially true for image analysis during repeated aquisition of molecular information including high throughput systems. It will also be important to describe molecular interrelations within cells simultaneously with the architectural interreactions of cells within 3D reconstructed tissue areas and to parametrize complex 3D contours of cell and oganelle shapes. Furthermore, the development of new strategies of fast knowledge extraction from highly multiparametric data for example for on-line knowledge extraction in high-throughput systems has to be advanced and structures for the permanent storage of relational cell classification systems for various diseases and cell states have to be generated.

Altogether, a human cytome project constitutes a substantial challenge for biomedical and bioinformatic scientists, clinicians and innovative technological developments and has the potential to provide essential new leads to potential molecular drug targets.

Reprint requests to:
Prof. Dr. Günter K. Valet
Max-Planck-Institut für Biochemie
Am Klopferspitz 18
D-82152 Martinsried, Germany
valet@biochem.mpg.de

References

1. Collins FS, Green ED, Guttmacher AE, Guyer MS. A vision for the future of genomics research. Nature 2003; 422: 835-47.
2. Aloy P, Stark A, Hadley C, Russell RR. Predictions without templates: new folds, secondary structure, and contacts in CASP5. Proteins 2003; 53: 436-56.
3. Kermani F. The future of biopharmaceutical research and development. In: Cooper E, ed. Business Briefing: Future drug discovery 2002, London: World Markets Research Institute, 2002; 16-8. (http://www.bbriefings.com)
4. Burrill GS. Fewer drugs approved, more money spent: Where's the beef ? Drug Discovery World 2004; 5: 9-11. (http://www.ddw-online.com/contents.asp)
5. George TC, Hall BE, Zimmermann CA, et al. Distinguishing modes of cell death using the ImageStream multispectral imaging flow cytometer. Cytometry 2004; 59A: 237-45.
6. Gerstner AOH, Trumpfheller C, Racz P, Osmancik P, Tenner-Racz K, Tarnok A. Quantitative histology by multicolor slide-based cytometry. Cytometry 2004; 59A: 210-9.
7. Ecker RC, Steiner GE. Microscopy-based multicolor tissue cytometry at the single cell level. Cytometry 2004; 59A: 182-90.
8. Ecker RC, de Martin R, Steiner GE, Schmid JA. Application of spectral imaging microscopy in cytomics and fluorescence energy transfer (FRET) analysis. Cytometry 2004; 59A: 179-82.
9. Valet GK, Leary J, Tárnok A. Cytomics - New technologies: Towards a human cytome project. Cytometry 2004; 59A: 167-71.
10. Valet G. Predictive medicine by cytomics: Potential and challenges. J Biol Regul Homeost Agents 2002; 16:164-7.
11. Valet G, Valet M, Tschöpe D, et al. White cell and thrombocyte disorders: Standardized, self-learning flow cytometric list mode data classification with the CLASSIF1 program system. Ann NY Acad Sci 1993; 677: 233-51.
12. Valet G, Hoeffkes HG. Data pattern analysis for the individualised pretherapeutic identification of high risk diffuse large B-cell lymphoma (DLBCL) patients by cytomics. Cytometry 2004; 59A: 232-36.
13. Valet G, Repp R, Link H, Ehninger G, Gramatzki M and SHG-AML study group. Pretherapeutic identification of high-risk acute myeloid leukemia (AML) patients from immunophenotypic, cytogenetic, and clinical parameters. Cytometry 2003; 53B: 4-10.
14. Valet GK, Tárnok A. Cytomics in predictive medicine. Cytometry 2003; 53B: 1-3.
15. Rosenwald A, Wright G, Chan WC, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. NEJM 2002; 346: 1937-47.
16. Repp R, Schaekel U, Helm G, et al. Immunophenotyping as an independent factor for risk stratification in AML. Cytometry 2003; 53B: 11-9.
17. Palková Z, Váchová L, Valer M, Preckel T. Single cell analysis of yeast, mammalian cells and fungal spores with a microfluidic pressure driven chip-based system. Cytometry 2004; 59A: 246-53.
18. Parak WJ, Gerion D, Pellegrino T, et al. Biological applications of colloidal nanocrystals. Nanotechnology 2003; 14: 15-27.