Techniques and Applications

A database generator for human brain imaging

Per Roland, Gert Svensson, Tony Lindeberg, Tore Risch, Peter Baumann, Andreas Dehmel, Jesper Frederiksson, Hjörleifer Halldorson, Lars Forsberg, Jeremy Young and Karl Zilles

Sharing scientific data containing complex information requires new concepts and new technology. NEUROGENERATOR is a database generator for the neuroimaging community. A database generator is a database that generates new databases. The scientists submit raw PET and fMRI data to NEUROGENERATOR, which then processes the data in a uniform way to create databases of homogenous data suitable for data sharing, met-analysis and modelling the human brain at the systems level. These databases are then distributed to the scientists.

Imaging of the functions of the human brain is one of the most rapidly growing fields in neuroscience. The main objective is to map the location of cerebral functions. Mapping of such functions with PET or fMRI results in changes in blood flow, metabolism and the fMRI-BOLD signal at multiple distinct sites in the grey matter of the human brain. Each site reflects the activation, that is, the increased neuronal and synaptic activity of a large population of neurons subserving the function being performed by the brain.

In the past decade, the interest of scientists has moved to characterize the functional contribution of each large, active neuronal population and the interactions between different neuronal populations, as opposed to just being satisfied with the picture of the multiple active neuronal populations subserving the function as a whole. Such characterization is considerably harder. Any small part of the primate cerebral cortex has reciprocal connections to 10–20 other brain regions[1]. This means that in principle a small part of the cortex could make various functional contributions for which a common denominator might be hard to find. It would take many experiments to find such a common denominator for a single brain region in addition to characterizing the interactions with other brain regions. One possibility is

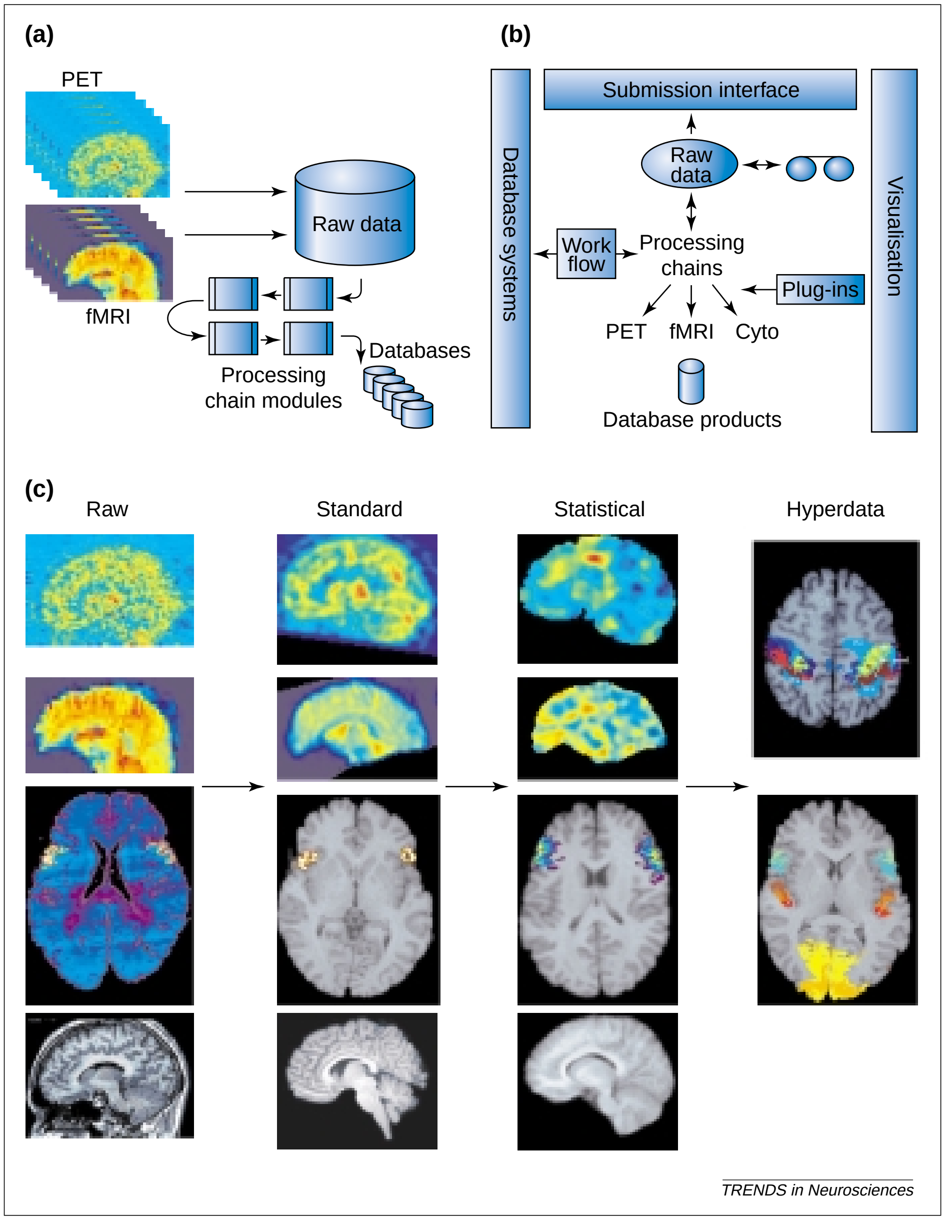

Fig. 1. The basic architecture of NEUROGENERATOR. (a) The scientists submit their raw PET and fMRI-data and get databases back with database management systems, query-, visualization- and modelling- tools. Together with the raw functional image data, cyto-architectural data and anatomical data of the brains are entered together with the data describing the detailed experimental conditions. (b) The database management systems and special software are controlling the databases, image analysis processing modules and database production. The tape robots store and retrieve the raw data to be processed by the software of the processing chain to yield a new database with uniform data and query systems. When better software becomes available the corresponding software module is replaced. (c) NEUROGENERATOR contains and produces many databases at each level of processing. From the top: PET, fMRI, cytoarchitectural, and anatomical (MR) data. These are available in the raw format, transformed into a standard anatomical format (of several possible), into statistical images, and by the modelling and query tools into hyperdata [here examples of a functional connectivity map (top, see details Fig. 2) and a cytoarchitectural probability map (bottom)].

  PII: S0166-2236(00)01924-X

to compare published reports about the activations associated with the function with similar functions. The published reports, however, contain only a fraction of the underlying data presented in tables, descriptions and two-dimensional (2-D)-figures which give only a gist of the 3-D results. To remedy this, several laboratories are starting initiatives of comparing data across laboratories and establishing databases with the purpose of arriving at more-general conclusions about the functional contributions of different parts of the human brain[2].

Data in-homogeneity and data sharing

The analysis of fMRI and PET data is complex and involves several steps, such as image segmentation, warping into a standard anatomical format, noise reduction, statistical modelling and hypotheses testing. A large variety of image analysis and statistical methods are used. These methods change rapidly, and their diversity hampers scientific comparison, meta-research, and induces variance in the location, extent, number and amplitude of the PET and fMRI activations. Consequently it will be difficult, if not impossible, to arrive at scientifically valid conclusions by comparing 3-D statistical images of different nature, for example with different spatial resolution, underlying segmentation or anatomical format.

Most neuroimaging neuroscientists will be willing to share published data in a format that is close to that in which their data was published. But this type of sharing or database is hampered not only by its limited scientific value but also by different image formats, different hardware platforms and lack of compatible software. Therefore, sharing of raw data is much better. Submitting raw PET and fMRI data to databases has, however, created a debate among neuroscientists concerning (1) the scientist's willingness and opinions in sharing their raw data; (2) the issue of distributing the databases back to the scientists themselves; (3) the issue of collecting data not just for the sake of sharing them and comparing data across laboratories but also to use this data to build computational models of the human brain at the systems level; and (4) the lack of standards in fMRI and PET data and neuroimage data processing and statistics in general[2–8].

Initiatives are developing in the form of distributed 'Napster-like' databases that enable scientists to download data, including raw data, from another laboratory[7]. To be successful, such distributed databases should contain precise descriptions for each study of how the data were created in the form of meta-data accompanying the images. This, in turn, entails rather advanced search and query systems. However, problems remain with image formats, availability of converters, etc. although these could be accessible at the website of the distributed database.

The database generator

We are creating a database generator, called NEUROGENERATOR. A database generator is a database that can generate other databases (Fig. 1). NEUROGENERATOR is based on researchers voluntarily submitting their raw untreated PET and fMRI data along with detailed documentation of the experimental conditions (corresponding to the methods section in a submitted article) to a central storage at a super-computer centre. Because there is little variation in the quality of state-of-the-art scanners, it is possible to define a minimum quality standard for the submitted data. At the super-computer centre, the raw data are processed with state-of-the-art image segmentation software, noise reduction software, software which transform images into an optimal standard brain format, and statistical modelling software, which will enable hypotheses testing, estimation of variance, correlations, etc. Thus, it will be possible to analyse experiments from different laboratories using identical software (Fig. 1). As the brain image analysis methods are rapidly improving, each step in the image processing can be replaced once better software is available. For each particular software chain, the raw data will be processed into a homogenous database product. This database product contains the statistical images in a single anatomical format and database software to retrieve, query and model the statistical images. Therefore, all statistical images will be directly comparable and can be queried and modelled with advanced database query tools. As one or several links in the processing chain are interchanged with presumably better software, new homogenous database products are produced, which are available to the scientific community, either to download into the scientist's host computer or to be accessed through the Internet.

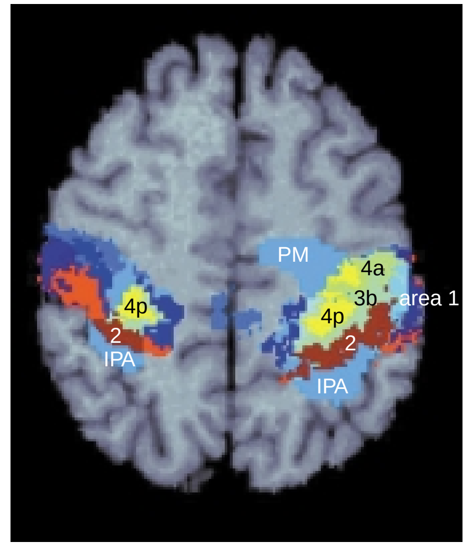

Fig. 2. Functional connectivity of somatosensory areas. The primary motor areas 4a,4p and somatosensory areas 3a,3b,1 and 2 were delimited statistically in nine post-mortem brains based on quantitative cytoarchitecture and receptor densities[8]. Voxels in brain space were then allocated to the area having the highest probability of representing the voxels (= a cytoarchitectural probability map). Here are shown areas which have statistically significant correlations ($P < 0.01$ omnibus) with the relative regional cerebral blood flow of the hand section of the left area 1. Based on raw data from six independent studies of 40 subjects totally at rest, that is, not moving, relaxed, with eyes closed, and with darkness in front of the mind's eye, thinking of nothing. The blood flow is expected to show correlation for areas that are anatomically interconnected. The blood flow in area 1 was correlated with that of motor areas 4a and 4p and the dorsal premotor area (PM) and somatosensory areas 3a, and 3b in the same hemisphere and area 2, 4p and the cortex lining the anterior part of the intraparietal sulcus (IPA) in both hemispheres. This is in line with the known connectivity in macaque monkeys[10]. Please note that the selection of test space (hand section of cytoarchitectural area 1, light blue) and the extent of the significant correlations are selected/determined in an unbiased way[9,11].

Contents and further modelling facilities

Apart from raw data and meta-data from PET and fMRI experiments, the NEUROGENERATOR databases also contain data of microstructural parcellations of the cerebral cortex, thalamic nuclei and other sub-cortical structures in the human brain (Fig. 1). The microstructural data consist of the cytoarchitecture, that is, the packing density and the laminar distribution of neurons and of myelinated fibres, and measurements of the regional and laminar distributions of neurotransmitter receptors. In this way, it is possible to create models of the functional–structural organization of the human telencephalon, and to correlate patterns of multiple, activated cortical populations with objectively and statistically defined cortical areas, or to use these objectively defined microstructural cortical areas as volumes of interest to study, for example, the functional connectivity of the cerebral cortex in man (Fig. 2). At the same time NEUROGENERATOR is also intended to be a resource for computational neuroscientists and a project to model data from the human brain at the systems level. The database generator is built on the database mediator AMOS II, integrating descriptive experimental data and processed images (http://www.dis.uu.se/~udbl/amos). The multidimensional images are handled by the database system RasDaMan (http://www.rasdaman.com).

In general, databases containing raw image data are tools to examine the reproducibility and consistency of functional activations of the human brain. By its design NEUROGENERATOR will also be a tool for testing the efficiency, accuracy and sensitivity of image analysis and statistical software, because such software can be directly compared in large homogenous data sets. Meta-research using of raw databases avoids the common pitfalls of other meta-research, such as publication bias, a failure to operate on all pertinent data, failures to make a priori exclusions based on objective criteria, and first of all data in-homogeneity. The submitted raw data remain the property of the submitting laboratory and the data can only be used for publication with the acceptance of the submitting laboratory.

The concept of database generators is new. See more details about NEUROGENERATOR at http://www.neurogenerator.org. NEUROGENERATOR is an EU-project for the advancement of Neuroinformatics (QLG 1999 00677).

References

1 Felleman, D.J. and Van Essen, D.C. (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47
2 Chicurel, M. (2000) Databasing the brain. *Nature* 406, 822–825
3 Editorial (2000) Whose scans are they, anyway? *Nature* 406, 443
4 Aldhous, P. (2000) Prospect of data sharing gives brain mappers a headache. *Nature* 406, 445–446
5 Editorial (2000) A debate over fMRI data sharing. *Nat. Neurosci.* 3, 845–846
6 Koslow, S. (2000) Should the neuroscience community make a paradigm shift to sharing primary data? *Nat. Neurosci.* 3, 863–865
7 Marshall, E. (2000) Downloading the human brain with security. *Science* 289, 2250
8 OHBM (2001) Neuroimaging Databases. *Science* 292, 1–4
9 Geyer, S. *et al.* (1999) Areas 3a, 3b and 1 of human primary somatosensory cortex. *Neuroimage* 10, 63–83
10 Burton, H. and Fabri, M. J. (1995) Ipsilateral intracortical connections of physiologically defined cutaneous representations in areas 3b and 1 of macaque monkeys: projections in the vicinity of the central sulcus. *Comp. Neurol.* 355, 508–538
11 Lindeberg, T. *et al.* (1999) Analysis of brain activation patterns using a 3-D scale-space primal sketch. *Hum. Brain Mapp.* 7, 166–194

Per Roland
e-mail: Per.Roland@neuro.ki.se
Jeremy Young
e-mail: jeremy.young@neuro.ki.se
Karolinska Institute, Sweden.

Tony Lindeberg
e-mail: tony@nada.kth.se
Gert Svensson
e-mail: gert@pdc.kth.se
Jesper Frederiksson
e-mail: jesperf@nada.kth.se
Hjörleifur Halldorson
e-mail: hjalle@nada.kth.se
Lars Forsberg
e-mail: lfo@pdc.kth.se
Royal Institute of Technology, Sweden.

Tore Risch
e-mail: Tore.Risch@dis.uu.se
Uppsala University, Sweden.

Peter Baumann
e-mail: baumann@forwiss.tu-muenchen.de
Active Knowledge, Germany.

Andreas Dehmel
e-mail: dehmel@forwill.tu-muenchen.de
Bavarian Research Center for Knowledge-Based Systems, Germany.

Karl Zilles
e-mail: zilles@hirn.uni-dusseldorf.de
Heinrich.Heine-University, Düsseldorf, Germany.