Using & Creating Software & Toolboxes for Research: Cases from Neuroimaging Data Analyses

UNAL 'ZAK' SAKOGLU

COMPUTER ENGINEERING,

UNIVERSITY OF HOUSTON - CLEAR LAKE

HTTP://SCE.UHCL.EDU/SAKOGLU

SAKOGLU@UHCL.EDU

UNIVERSITY OF HOUSTON - CLEAR LAKE

Neuroimaging Research Data
is complex/sophisticated...

❑ Multidimensional: 3D or 4D (and also include surely 2D and 1D)

❑ Saved in special formats: DICOM, ANALYZE, NIFTI, ...

❑ Large in size, and also:

❑ Accompanied with lots of other data (behavioral, genetic, other imaging modalities, ...)

❑ Overall, neuroimaging research data has large number of pieces/files that need to be stored neatly (with easy to reach/query database hierarchy)

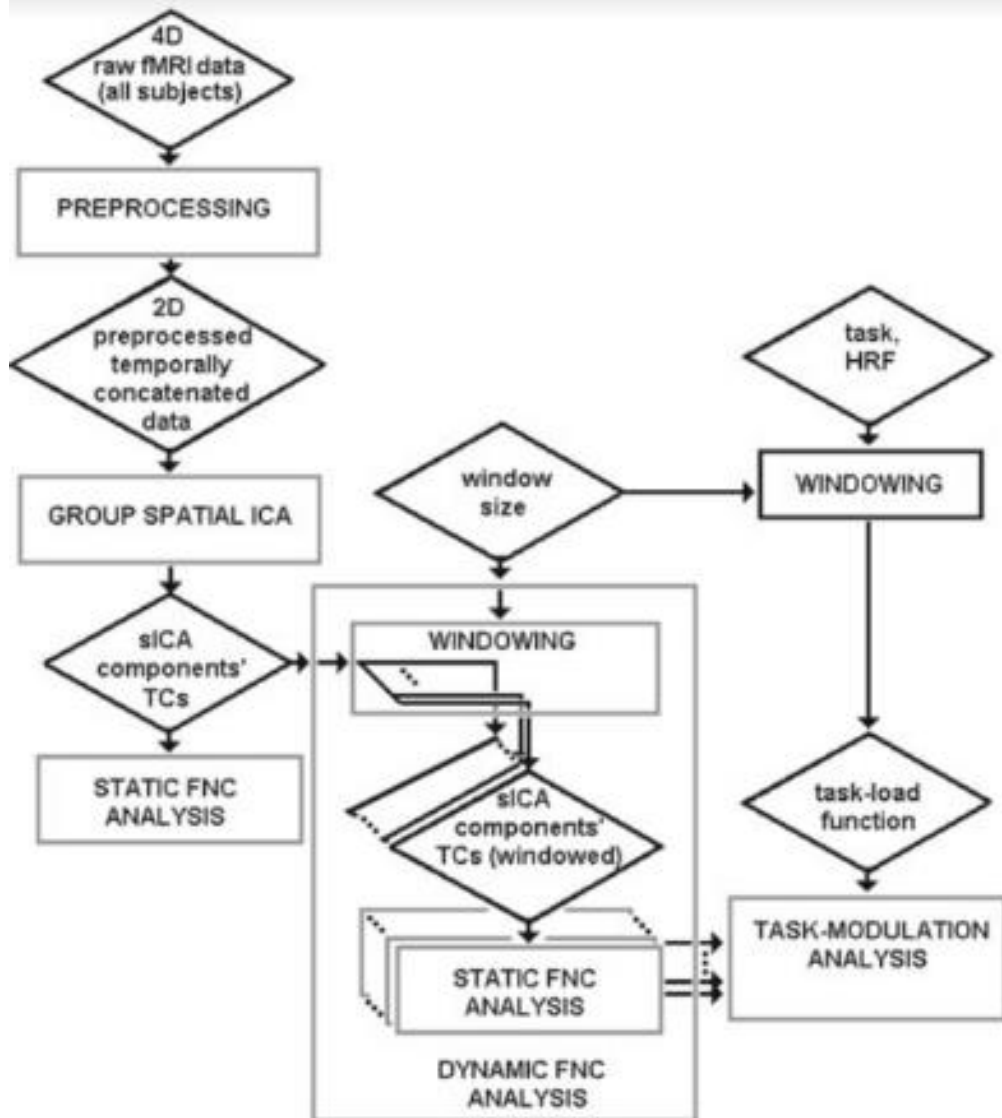❑ Sophisticated analysis pipeline(s) (sequence of methods) used

and the pipelines are not fixed, constantly evolving/ being updated over time as new techniques are developed!
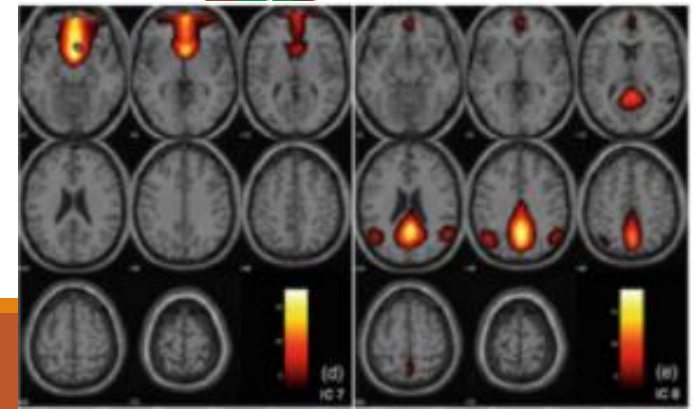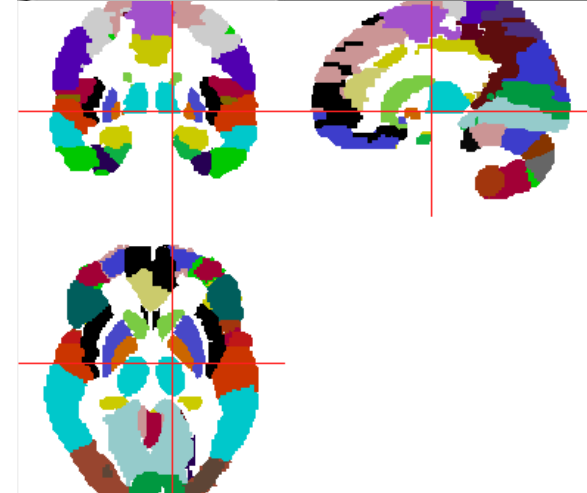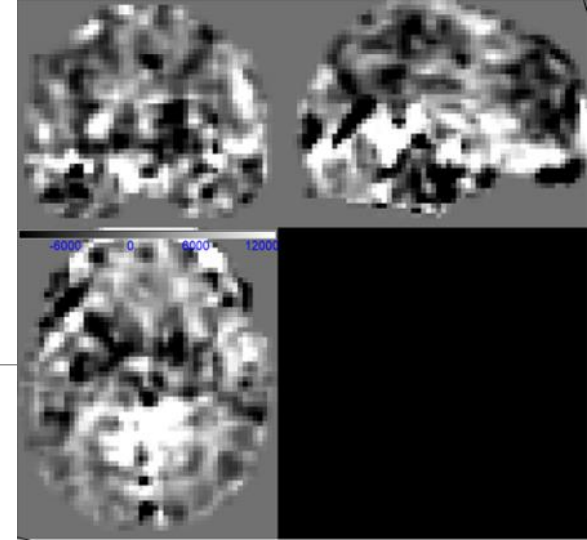
Analyses produce more and more (intermediate) data

❑ There are many processing & analyses steps:

❑ data cleanup, QC/inspection, pre-processing, noise-removal, filtering, slice-time correction, motion correction / co-registration, warping to a common template, reslicing, plus various high-level statistics, and then higher-level meta analyses, group analyses, etc,…

❑ There are various methods for each step (constantly scan recent methods papers)

❑ There are various tools which can do one or some of these steps (and each may be requiring different file formats!!!)

   so, data conversion tools/software may be needed (compatibility)

❑ As a result:

   for each GB of original/raw data → 100s of GBs of intermediate data may be generated!

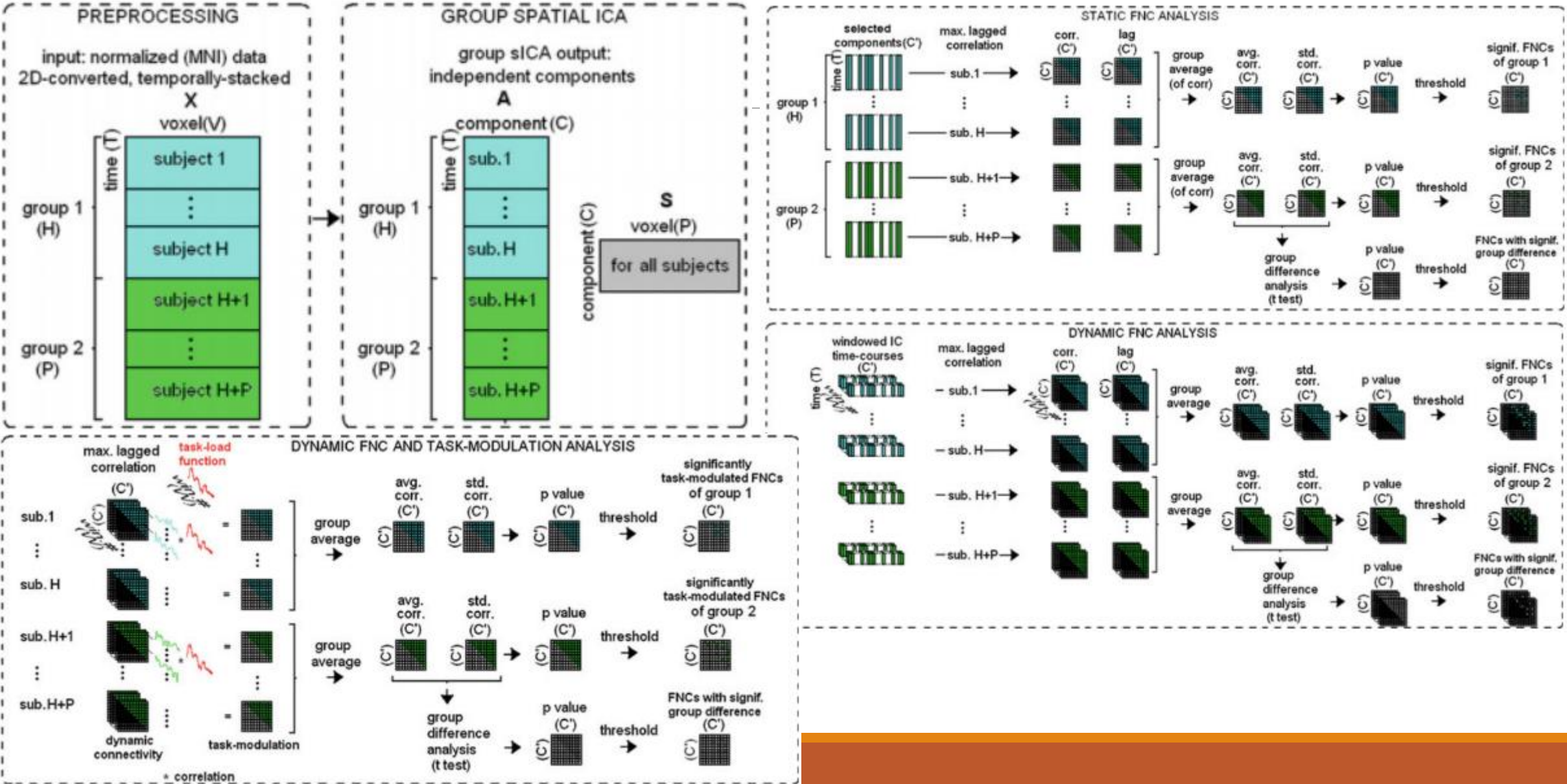# Analyses produce more and more (intermediate) data

RESEARCH ARTICLE

A method for evaluating dynamic functional network connectivity and task-modulation: application to schizophrenia
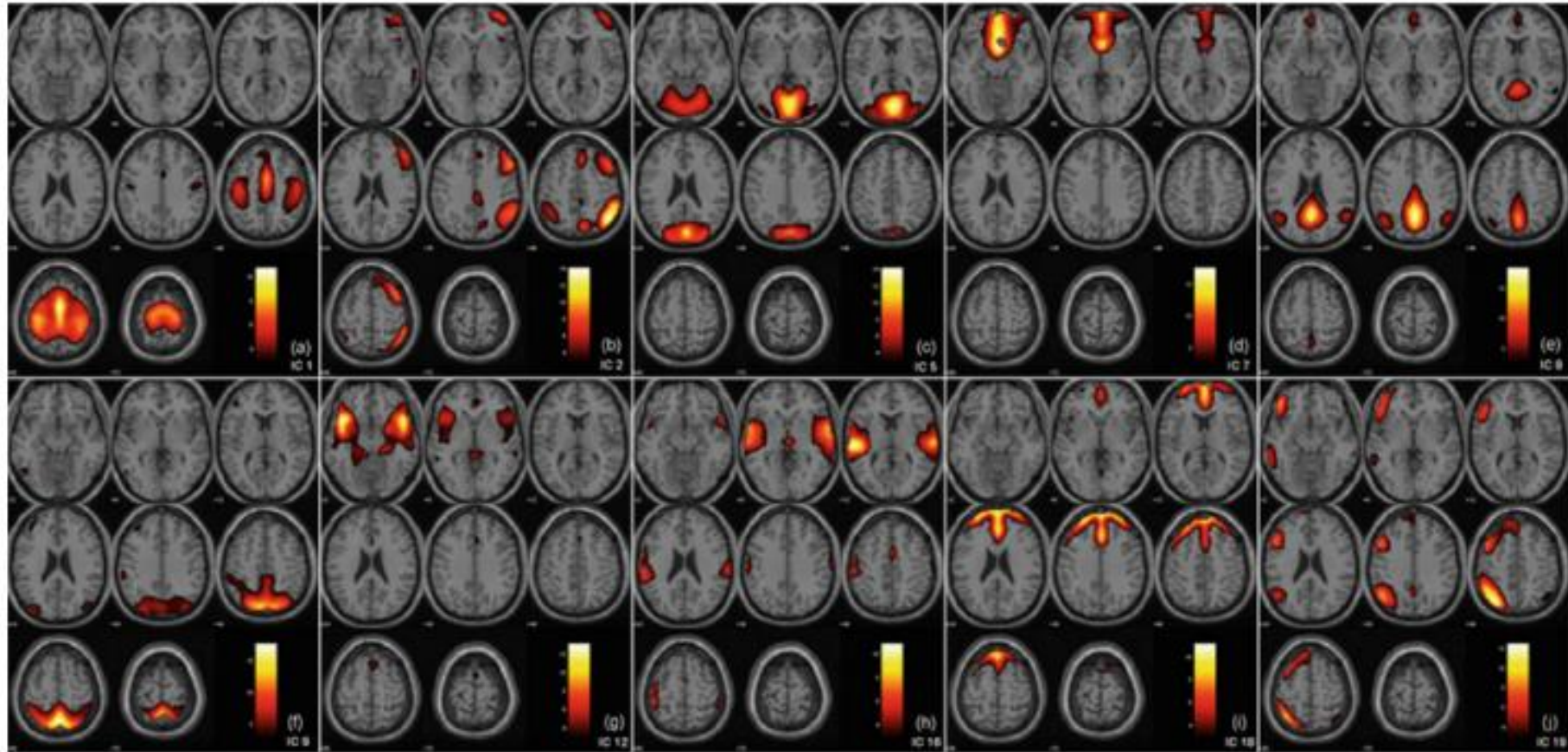
Ünal Sakoğlu · Godfrey D. Pearlson · Kent A. Kiehl ·
Y. Michelle Wang · Andrew M. Michael ·
Vince D. Calhoun

# Analyses produce more and more (intermediate) data

# Analyses produce more and more (intermediate) data

Fig. 9 Ten selected spatial ICA components (networks). **a** IC#1: motor network (M), **b** IC#2: right lateral fronto-parietal network (RLFP), **c** IC#5: medial visual network (mV), **d** IC#7: orbito-frontal network (OF), **e** IC#8: posterior default mode network (pDM), **f** IC#9: parietal network (P), **g** IC#12: anterior temporal network (aT), **h** IC#16: medial temporal network (mT), **i** IC#18: frontal network (F), and **j** IC#19: left lateral fronto-parietal network (LLFP)

Analyses produce more and more (intermediate) data
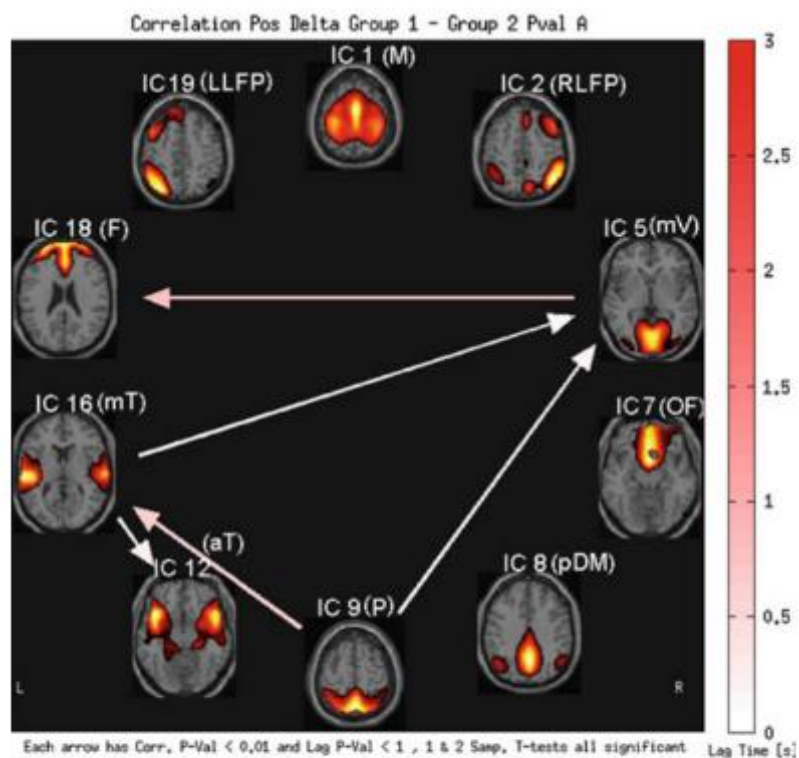
Fig. 10 HC–SP static FNC analysis results. Connections with significant group difference are shown ($P < 0.01$). The lag differences are ignored (lag difference $P < 1$). Bootstrap results with 20 subjects per group, 1,000 randomized trials, resulted in 100% occurrence (of the significant group difference of the same connections)
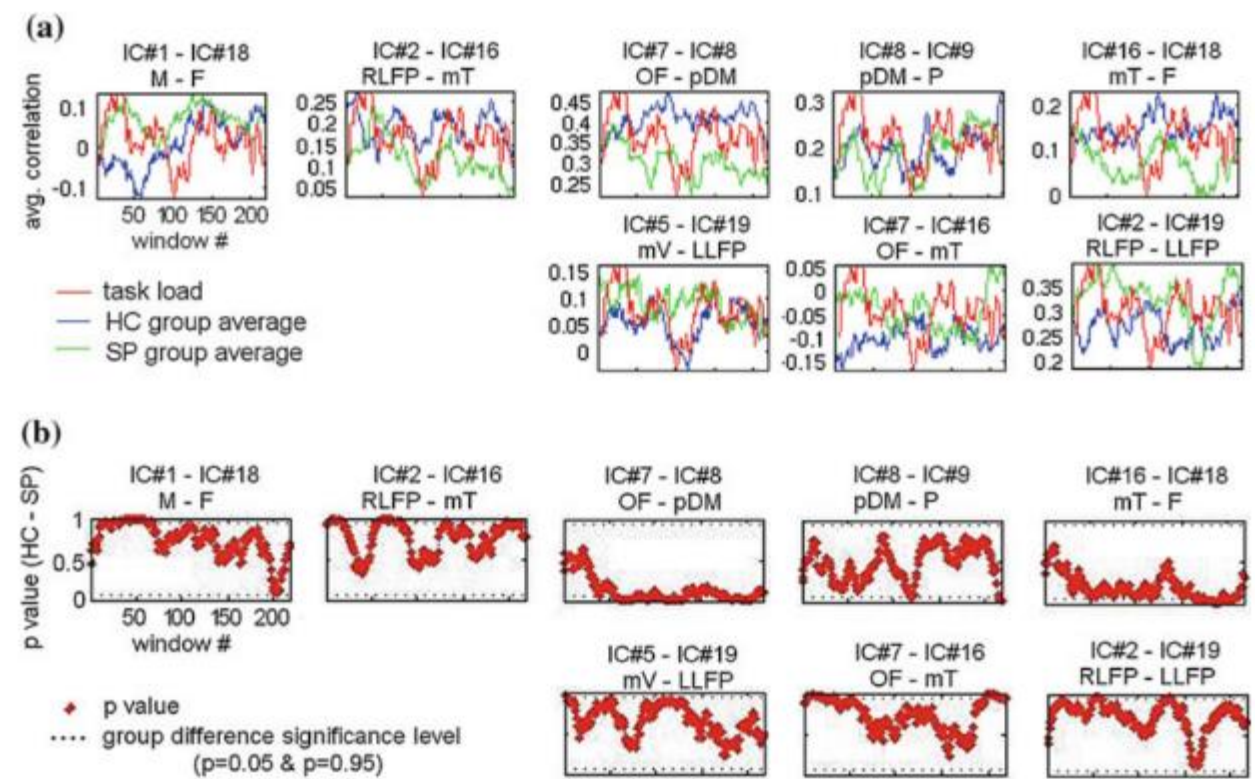
Fig. 11 Dynamic FNC analysis results. **a** Temporal evolution of average correlation coefficient for the two groups (*blue*: HC group average, *green*: SP group average), for selected pairs. Temporal evolution of task-load (*red*) is also presented for comparison. **b** Corresponding dynamic P-values of group significance of the dynamic correlation for these pairs

Analyses produce more and more (intermediate) data

component pairs.

Fig. 12 Group difference of HC–SP time–frequency plot of the a

362
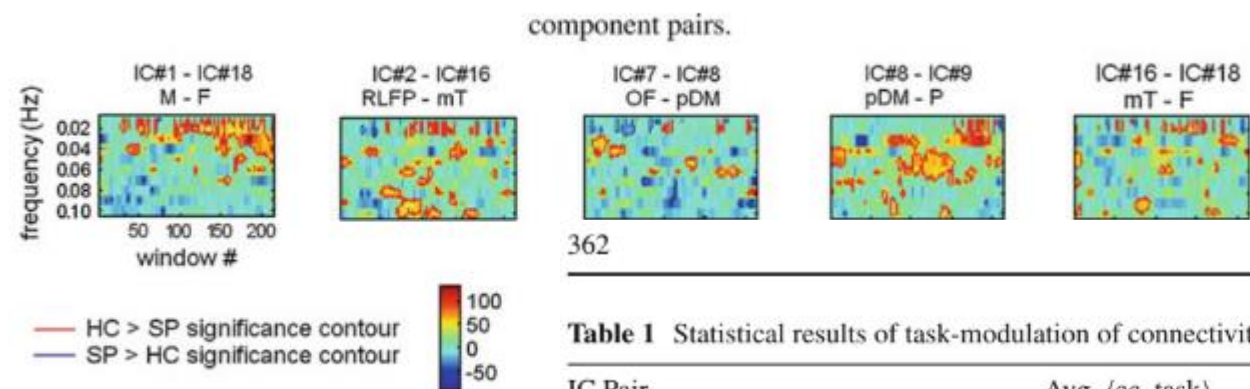
Table 1 Statistical results of task-modulation of connectivity

IC Pair	Avg. ⟨cc, task⟩		HC–SP (T-test)	Bootstrap results (N1 = N2 = 20, Repetition = 1,000)
	HC	SP	P-value	P < 0.05 (%)
IC#1 (M) &IC#18 (F)	0.0604	0.1221	0.0107 (SP>HC)*	69.4
IC#2 (RLFP) &IC#16 (mT)	0.1069	0.0562	0.0223 (HC>SP)*	54.6
IC#2 (RLFP) &IC#19 (LLFP)	0.0974	0.0960	0.4815 (HC>SP)	
IC#5 (mV) &IC#19 (LLFP)	0.1173	0.0786	0.0655 (HC>SP)	
IC#7 (OF) &IC#8 (pDM)	0.0528	0.0951	0.0128 (SP>HC)*	63.5
IC#7 (OF) &IC#16 (mT)	0.0868	0.0564	0.0566 (HC>SP)	
IC#8 (pDM) &IC#9 (P)	0.0456	0.0970	0.0133 (SP>HC)*	69.1
IC#16 (mT) &IC#18 (F)	0.0708	0.0376	0.0195 (HC>SP)*	55.0

R^2 value of regression of dynamic average correlation coefficients with the task load function for all component pairs, for the two groups and P value of the group difference. The component pairs marked with '*' show significant group difference, hence show significant group difference ($P < 0.05$) in specified task modulation of the functional connectivity between the specified component pair. None of them are significant when the Bonferroni correction is applied ($P < 0.002$). When a false discovery rate of $P < 0.05$ was applied, the cut-off value in the original P values was 0.0223. Bootstrapping results, in terms of percentage of the subgroups that provide $P < 0.05$ are also provided. Subgroups consisted of 20 subjects randomly chosen out of 28 subjects per each group and 1,000 random repetitions
* Statistically significant ($P < 0.05$)

So, lots of scripting, coding, batch processing
most of the time in different coding environments, and even with different OS & HW

pv2nifti_mouse.sh - Notepad

File Edit Format View Help

```
#!/bin/bash

progname=`basename $0`
cwd=$PWD

if [ $OSTYPE == 'solaris' ]
    then
    MYGAWK=/app/abt/nosw/bin/gawk
    MYGREP=/usr/bin/grep
    RM=/bin/rm
else
    MYGAWK=/bin/gawk
    MYGREP=/bin/grep
    RM=/bin/rm
fi

# Subroutines
###########################
usage ()
{
    echo "USAGE: $progname input <y/n>"
    echo "    * for input use 'subject' (to convert all image sets) or '2dseq' (to convert just one)"
    echo "    * output filenames are automatically generated in the base directory like directory.subject.re
    echo "    use and y to automatically overwrite existing files and n to automatically skip existing files (
}

getpar ()
{
    ${MYGREP} $2 $1 | awk -F= '{print $2}'
}
```

Ln 690, Col 3

e.g. shell scripting in Unix/Linux for preprocessing…

US_ICATCextract.m - Notepad

File Edit Format View Help

```
cd("C:\UHCL\DODProjectGopi\DFNCofGWNI\ResearchWork\DataAtEmory\DOD_GWI\...
Unal_old_MNIdata\Group_ICA\Anals_July25_dtAROMA_nIC44\GWS2_NCVC_FNC_regressionresults\...
Anals_Oct17_dtAROMA_CtrlsAndGWS2\gica_cmd_scaling_components_files")

clear all
myfilenames=ls
myfilenames2=myfilenames(3:55,:);

%Extract ICA TCs for each i subject
for i=1:size(myfilenames2,1),
    mydata=spm_read_vols(spm_vol(myfilenames2(i,:)));
    mydataICATC_all53subj(:,:,i)=mydata;
end

%Static FC
for i=1:53,
    subjData=mydataICATC_all53subj(:,:,i);
    mycc=corrcoef(subjData), %
    mycc_all53subj(:,:,i)=mycc;
end

%Dynamic FC
for i=1:53,
    subjData=mydataICATC_all53subj(:,:,i);
    for k
        compute k dynamic mycc matrices for each i
        mycc_all53subj_dynamic(:,:,k,i)=kth mycc matrice;
    end
end
```
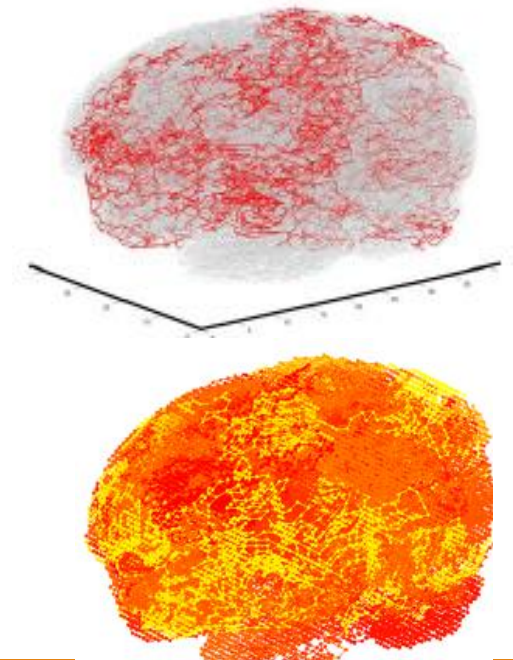
Ln 1, Col 1

Coding in MATLAB, in Windows for analyses
(have written hundreds of functions/subroutines/modules and tens of thousands of lines of code…)

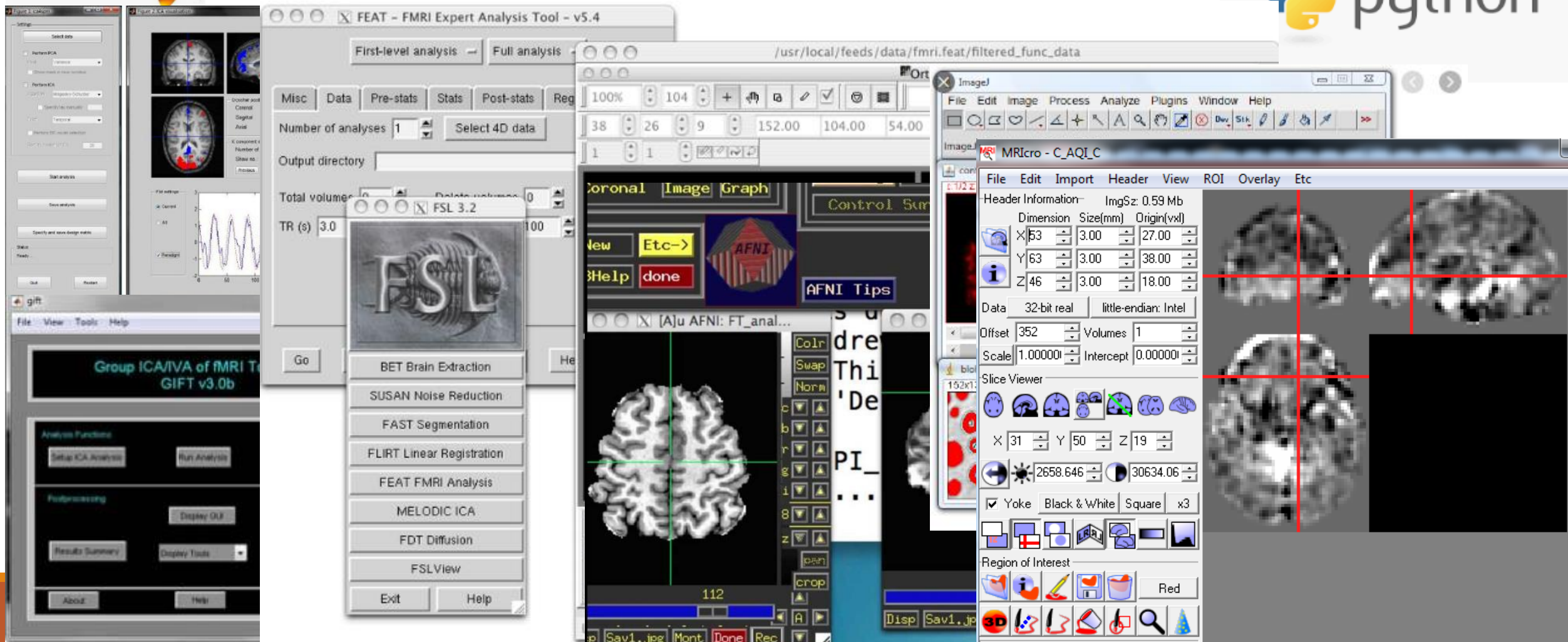Visualization using MATLAB, or another language (ImageJ, MRIcro, MRIcron, AFNI, FSL etc.)

So, lots of scripting, coding, batch processing
and using existing resources/software (versions, constantly evolving!)

Researchers are not necessarily efficient coders/programmers
So if there are resources (time or $$) we try to develop easy to use "toolboxes"
This is needed for continuity, transparency, replicate-ability, ...

Picking the right s/w, and developing (in-house/custom development) if necessary, is important!!!

- Takes a lot of effort ($$ and time)

- Are they well-documented, is support available?

- Hire/invest in/budget for a software developer if you can ($$?)

  or ask for help if they have the time (again $$? or time??)

- Not much grant $$ for that (even NSF?), even if there is, it is tough

- Most of the time we develop & distribute the toolboxes for free

  …. or in some cases it can be commercialized (FSL, etc.)

Fig 1: Brain 3D Voxel representation

CI/IT personnel:

- Be flexible with researchers, familiarize with their research

   Research s/w needs are varied, and continually evolve and change

- Planning and (prompt) communication is very important!!!

   Never start with phrase "Nope, it is not possible…" for a response   (never say never)

- In many cases researchers need & ask for high-level system access to be able to do modifications to the s/w (otherwise the red-tape takes just too long to add/install/modify any s/w, toolboxes, etc…), reasonably accommodate them

- Some research groups may have a dedicated IT/CI personnel & infrastructure that they want full/quick access/control of, and if they have the $$, they will establish a dedicated IT/CI personnel & infrastructure for better efficiency

Thank you!!!

- sakoglu@uhcl.edu if you have any further questions and/or comments

Fig 1: Brain 3D Voxel representation