

2018-10-29 15:56:46

I'm going to do a quick push to try to ru all images in the same space using ICA. Let's see what that gives us. First step is to convert all our images to TT_N27 space, and we can use that matrix later to put the EPI in the same space (https://afni.nimh.nih.gov/pub/dist/edu/latest/afni_handouts/afni10_volreg_talairach.pdf)

2018-10-30 08:00:53

```
cd ~/data/baseline_prediction/same_space/anat;
for m in `cut -d"," -f 1 ../../struct_rois_09062018_260timeDiff12mo.csv`;
do
    m2=`printf %04d $m`;
    mri_convert /Volumes/Shaw/freesurfer5.3_subjects/${m2}/mri/orig.mgz
    ./${m2}.nii.gz;
    @auto_tlrc -base TT_N27+tlrc -input ${m2}.nii.gz -no_pre -suffix
    _inTLRC;
done;
```

then, just so we can run everything in my desktop without depending on caterpie for data_by_maskID:

```
cd ~/data/baseline_prediction/same_space/epi;
for m2 in `cat ../../rsfmri_3minWithClinical.tsv`; do
    echo $m2;
    # mylink=`readlink /Volumes/Shaw/data_by_maskID/${m2} | sed
"s/\.\.\./Volumes/Shaw/";
    # 3dcopy
    ${mylink}/afni/${m2}.rest.subjectSpace.results/errts.${m2}.fanaticor+orig.
    HEAD ${m2}_epi.nii;
    @auto_tlrc -apar ../anat/${m2}_inTLRC.nii -input ${m2}_epi.nii -suffix
    _inTLRC -dxyz 2.5;
done;
```

Should probably check that the alignment worked well though... I'll go on with converting rsfMRI as well, and
NEED TO CHECK ALIGNMENT LATER!

dti

Let's go ahead and do this for DTI as well, which should be just a matter of making symlinks as we have all these data already generates for QC purposes:

```
for n in 223 272; do
    cd ~/data/baseline_prediction/same_space/dti/n${n}
    # ignore header
```

```

    for m in `tail -n +2
~/data/baseline_prediction/dti_gf_09212018_${n}timeDiff12mo.csv | cut -
d"," -f 2 -`; do
    maskid=`printf %04d $m`;
    for p in fa ad rd; do
        ln -s
/Volumes/Shaw/dti_robust_tsa/analysis_may2017/${maskid}_tensor_diffeo_${p}
.nii.gz .;
    done;
done;
done

```

But then I gunzipped and split everything by directory, because the Matlab tool was having lots of issues listing the directory.

2018-11-02 15:38:15

I've been generally running one version with 30 ICs, and then another one using the estimated number of ICs. Both, using MST500 for stability. I need to study a bit on what that actually does, because I don't see a stability metric like ICASSO. Of course, I could just do ICASSO if I want to. `

Another thing to look into is a better mask, because for the structural analysis I'm getting a lot of white matter and ventricle action. So, having a nicer gray matter mask would be better than just using the automask. That will be especially important for DTI too, if the property files didn't use the mask when being generated.

Of course, STILL NEED TO CHECK THAT TLRC ALIGNMENT WORKED WELL!

2018-11-05 11:16:22

I figured that since I'm doing a lot of this work, we might as well use a non-Linear transform, and also produce all the figures automatically. That's where the new SSwarper function comes in.

On the same step, I think we'll have better success in ICA if we use gray matter masks (or white matter, for DTI, if the files we're using are not restricted to that yet). So, for the structural data we'll need to use Freesurfer masks and put them into the common space.

Ideally I'd run this in Biowulf, but until I hear from them about Xfvb and netpbm, I'll run it locally:

```

cd ~/data/baseline_prediction/same_space/anat;
for m in `cut -d"," -f 1 ../../struct_rois_09062018_260timeDiff12mo.csv`;
do
    @SSwarper -input ${m2}.nii.gz -base TT_N27_SSW.nii.gz -subid ${m2};
    3dcalc -a
/Volumes/Shaw/freesurfer5.3_subjects/${m2}/SUMA/aparc+aseg_REN_gm.nii.gz \
    -prefix ${m2}_gm_mask.nii -expr "step(a)";
    3dNwarpApply -nwarp "anatQQ.${m2}_WARP.nii anatQQ.${m2}.aff12.1D" \
    -source ${m2}_gm_mask.nii -master anatQQ.${m2}.nii -inter NN \

```

```

-overwrite -prefix ${m2}_gm_mask_NL_inTLRC.nii;
done;

```

For DTI, they're all in the same space already, so I'll just use the FA skeleton mask we've been using for the voxelwise analysis.

Just needed to say that using the fa_skeleton masks causes the estimates of ICs to be quite low. For both n223 and n272 AD, I get only 4 ICs. For n223 FA, I got only 1, so I'm not running faEstimatedMaskedMST200. I'll check n272 and also rd, but I might end up skipping those estimated results in the end.

Yep, only one IC for FA n272 as well. Will only run IC20, and then check on RD.

2018-11-06 10:43:47

I got 10 and 11 for RD, so I'm running estimated there as well.

For anatomy, it's actually faster to swarm it:

```

cd ~/data/baseline_prediction/same_space/anat;
for m in `cut -d"," -f 1 ../../struct_rois_09062018_260timeDiff12mo.csv`;
do
    m2=`printf %04d $m`;
    echo "export OMP_NUM_THREADS=16; cd
~/data/baseline_prediction/same_space/anat; @SSwarper -input ${m2}.nii.gz
-base TT_N27_SSW.nii.gz -subid ${m2}" >> swarm.SSwarper
done;
swarm -g 10 -t 16 --job-name SSwarper --time 2:00:00 -f swarm.SSwarper -m
afni --partition quick --logdir trash

for m in `cut -d"," -f 1 ../../struct_rois_09062018_260timeDiff12mo.csv`;
do
    m2=`printf %04d $m`;
    3dcalc -a
/data/NCR_SBRB/freesurfer5.3_subjects/${m2}/SUMA/aparc+aseg_REN_gm.nii.gz
\
    -prefix ${m2}_gm_mask.nii -expr "step(a)";
    3dNwarpApply -nwarp "anatQQ.${m2}_WARP.nii anatQQ.${m2}.aff12.1D" \
    -source ${m2}_gm_mask.nii -master anatQQ.${m2}.nii -inter NN \
    -overwrite -prefix ${m2}_gm_mask_NL_inTLRC.nii;
done;

```

The last two commands run quite fast, but it's also a nice break to first go through the images to make sure alignment was successful... visually checked, and it looks good.

Back to DTI, let's convert the group ICA to a format that we can readily read into R:

```

load('rdEstimatedMaskedMST200Subject.mat')
load('rdEstimatedMaskedMST200_ica_c1-1.mat')

```

```

dlmwrite('~data/baseline_prediction/same_space/dti/n223/rdEstimatedMasked
MST200.csv', tc)
dlmwrite('~data/baseline_prediction/same_space/dti/n223/filelist.txt',fil
es.name, '')
load('adEstimatedMaskedMST200_ica_c1-1.mat')
dlmwrite('~data/baseline_prediction/same_space/dti/n223/adEstimatedMasked
MST200.csv', tc)
load('faIC20MaskedMST200_ica_c1-1.mat')
dlmwrite('~data/baseline_prediction/same_space/dti/n223/faIC20MaskedMST20
0.csv', tc)
load('adIC20MaskedMST200_ica_c1-1.mat')
dlmwrite('~data/baseline_prediction/same_space/dti/n223/adIC20MaskedMST20
0.csv', tc)
load('rdIC20MaskedMST200_ica_c1-1.mat')
dlmwrite('~data/baseline_prediction/same_space/dti/n223/rdIC20MaskedMST20
0.csv', tc)

```

Then, in R:

```

fnames =
readLines('~data/baseline_prediction/same_space/dti/n272/filelist.txt')
maskids = sapply(fnames, function(x)
gsub('/gpfs/gsf7/users/sudregp/baseline_prediction/same_space/dti/n272/rd
/', '', x)[[1]])
mask.id = sapply(maskids, function(x)
as.numeric(gsub('_tensor_diffeo_rd.nii,1', '', x)[[1]]))
for (f in c('adEstimatedMaskedMST200', 'rdEstimatedMaskedMST200',
            'adIC20MaskedMST200', 'rdIC20MaskedMST200',
            'faIC20MaskedMST200')) {
  fname =
  sprintf('~data/baseline_prediction/same_space/dti/n272/%s.csv', f)
  d = read.csv(fname, header=0)
  colnames(d) = sapply(1:ncol(d), function(x) sprintf('v_%02d',x))
  d2 = cbind(mask.id, d)
  data = merge(cclin, d2, by='mask.id')
  out_fname =
  sprintf('/data/NCR_SBRB/baseline_prediction/dti_n272_%s_11022018.RData.gz'
, f)
  save(data, file=out_fname, compress=T)
}

```

And then do the same thing for the other n.

Now, let's run the raw scripts for all those different permutations, and see what we get:

```

job_name=dti_sameSpace;
mydir=/data/NCR_SBRB/baseline_prediction/;
swarm_file=swarm.automl_${job_name};
rm -rf $swarm_file;

```

```

for f in dti_n223_adEstimatedMaskedMST200_11022018.RData.gz \
dti_n223_rdEstimatedMaskedMST200_11022018.RData.gz \
dti_n223_adIC20MaskedMST200_11022018.RData.gz \
dti_n223_rdIC20MaskedMST200_11022018.RData.gz \
dti_n223_faIC20MaskedMST200_11022018.RData.gz \
dti_n272_adEstimatedMaskedMST200_11022018.RData.gz \
dti_n272_rdEstimatedMaskedMST200_11022018.RData.gz \
dti_n272_adIC20MaskedMST200_11022018.RData.gz \
dti_n272_rdIC20MaskedMST200_11022018.RData.gz \
dti_n272_faIC20MaskedMST200_11022018.RData.gz; do
  for target in nvVSper nvVSrem perVSrem nvVSadhd; do
    for i in {1..100}; do
      echo "Rscript --vanilla
~/research_code/automl/raw_autoValidation.R ${mydir}/${f}
${mydir}/long_clin_0918.csv ${target}
${mydir}/models_spatial_within_DL/${USER} $RANDOM" >> $swarm_file;
    done;
  done;
done
sed -i -e "s/^/unset http_proxy; /g" $swarm_file;
split -l 1000 $swarm_file ${job_name}_split;
for f in `ls ${job_name}_split?`; do
  echo "ERROR" > swarm_wait_${USER}
  while grep -q ERROR swarm_wait_${USER}; do
    echo "Trying $f"
    swarm -f $f -g 30 -t 16 --time 3:00:00 --partition quick --logdir
trash_${job_name} --job-name ${job_name} -m R,afni --gres=lscratch:10 2>
swarm_wait_${USER};
    if grep -q ERROR swarm_wait_${USER}; then
      echo -e "\tError, sleeping..."
      sleep 10m;
    fi;
  done;
done

```

While we wait for the DTI stuff above running, let's go back to structural. Now, the challenge is making the grey matter masks. Using AFNI's 3dmasktool, let's create 3 masks: intersection, union, and a general one based on their examples (70% overlap plus some other shenanigans).

```

3dmask_tool -input ???_gm_mask_NL_inTLRC.nii -prefix
group_gm_mask_inter.nii -frac 1
3dmask_tool -input ???_gm_mask_NL_inTLRC.nii -prefix
group_gm_mask_union.nii -frac 0
3dmask_tool -input ???_gm_mask_NL_inTLRC.nii -prefix
group_gm_mask_fancy.nii \
  -dilate_input 5 -5 -frac 0.7 -fill_holes

```

Now we just need to run the Estimated and IC versions of ICA for each of the masks.

rsfMRI after nonlinear transform

The rsfMRI data still needs to go to the template space after all the no-linear transforms. This is how we'll do it:

```
cd ~/data/baseline_prediction/same_space/epi;
for m2 in `cat ../../rsfmri_3minWithClinical.tsv`; do
    echo $m2;
    3dNwarpApply -nwarp "../../anat/anatQQ.${m2}_WARP.nii
    ../../anat/anatQQ.${m2}.aff12.1D" \
        -source ${m2}_epi.nii -master ../../anat/anatQQ.${m2}.nii -dxyz 2.5\
        -overwrite -prefix ${m2}_epi_NL_inTLRC.nii;
done;
```

Also, I just confirmed that in our main analysis the EPI is warped to subject space, and not vice-versa. So, we should be good here to apply the non-linear transform to the ernts file.

But not all mask ids in fMRI are in the structural analysis, so we need to warp those still...

2018-11-07 08:59:20

```
cd ~/data/baseline_prediction/same_space/epi;
mkdir extra_NL
for m2 in `cat ../../rsfmri_3minWithClinical.tsv`; do
    if [ ! -e ${m2}_epi_NL_inTLRC.nii ]; then
        mri_convert
        /data/NCR_SBRB/freesurfer5.3_subjects/${m2}/mri/orig.mgz
        extra_NL/${m2}.nii.gz;
        echo "export OMP_NUM_THREADS=16; cd
        ~/data/baseline_prediction/same_space/epi/extra_NL; @SSwarper -input
        ${m2}.nii.gz -base TT_N27_SSW.nii.gz -subid ${m2}" >> swarm.SSwarper;
    fi;
done;
swarm -g 10 -t 16 --job-name SSwarper --time 2:00:00 -f swarm.SSwarper -m
afni --partition quick --logdir trash

# then, when it's finished
cd ~/data/baseline_prediction/same_space/epi;
for m2 in `cat ../../rsfmri_3minWithClinical.tsv`; do
    if [ ! -e ${m2}_epi_NL_inTLRC.nii ]; then
        3dNwarpApply -nwarp "extra_NL/anatQQ.${m2}_WARP.nii
        extra_NL/anatQQ.${m2}.aff12.1D" \
            -source ${m2}_epi.nii -master extra_NL/anatQQ.${m2}.nii -dxyz 2.5\
            -overwrite -prefix ${m2}_epi_NL_inTLRC.nii;
    fi;
done;
```

And let's also grab the DTI results:

```
echo "target,pheno,var,seed,nfeat,model,auc,f1,acc,ratio" >
dtiSameSpace_summary.csv;
dir=dti_sameSpace
for f in `ls -1 trash_${dir}/*o`; do
  phen=`head -n 2 $f | tail -1 | awk '{FS=" "; print $6}' | cut -d"/" -f
6`;
  target=`head -n 2 $f | tail -1 | awk '{FS=" "; print $8}'`;
  seed=`head -n 2 $f | tail -1 | awk '{FS=" "; print $10}'`;
  var=`head -n 2 $f | tail -1 | awk '{FS=" "; print $5}' | cut -d"/" -f
4 | sed -e "s/\.R//g"`;
  model=`grep -A 1 model_id $f | tail -1 | awk '{FS=" "; print $2}' |
cut -d"_" -f 1`;
  auc=`grep -A 1 model_id $f | tail -1 | awk '{FS=" "; print $3}'`;
  nfeat=`grep "Running model on" $f | awk '{FS=" "; print $5}'`;
  ratio=`grep -A 1 "Class distribution" $f | tail -1 | awk '{FS=" ";
{for (i=2; i<=NF; i++) printf $i ";"}}'`;
  f1=`grep -A 2 "Maximum Metrics:" $f | tail -1 | awk '{FS=" "; print
$5}'`;
  acc=`grep -A 5 "Maximum Metrics:" $f | tail -1 | awk '{FS=" "; print
$5}'`;
  echo $target,$phen,$var,$seed,$nfeat,$model,$auc,$f1,$acc,$ratio >>
dtiSameSpace_summary.csv;
done;
```

And we can make some plots in R:

```
data = read.csv('~/.tmp/dtiSameSpace_summary.csv')
target='nvVSper'
p1<-ggplot(data[data$target == target,], aes(x=pheno, y=acc, fill=pheno))
print(p1+geom_boxplot() + ggtitle(target))
```







