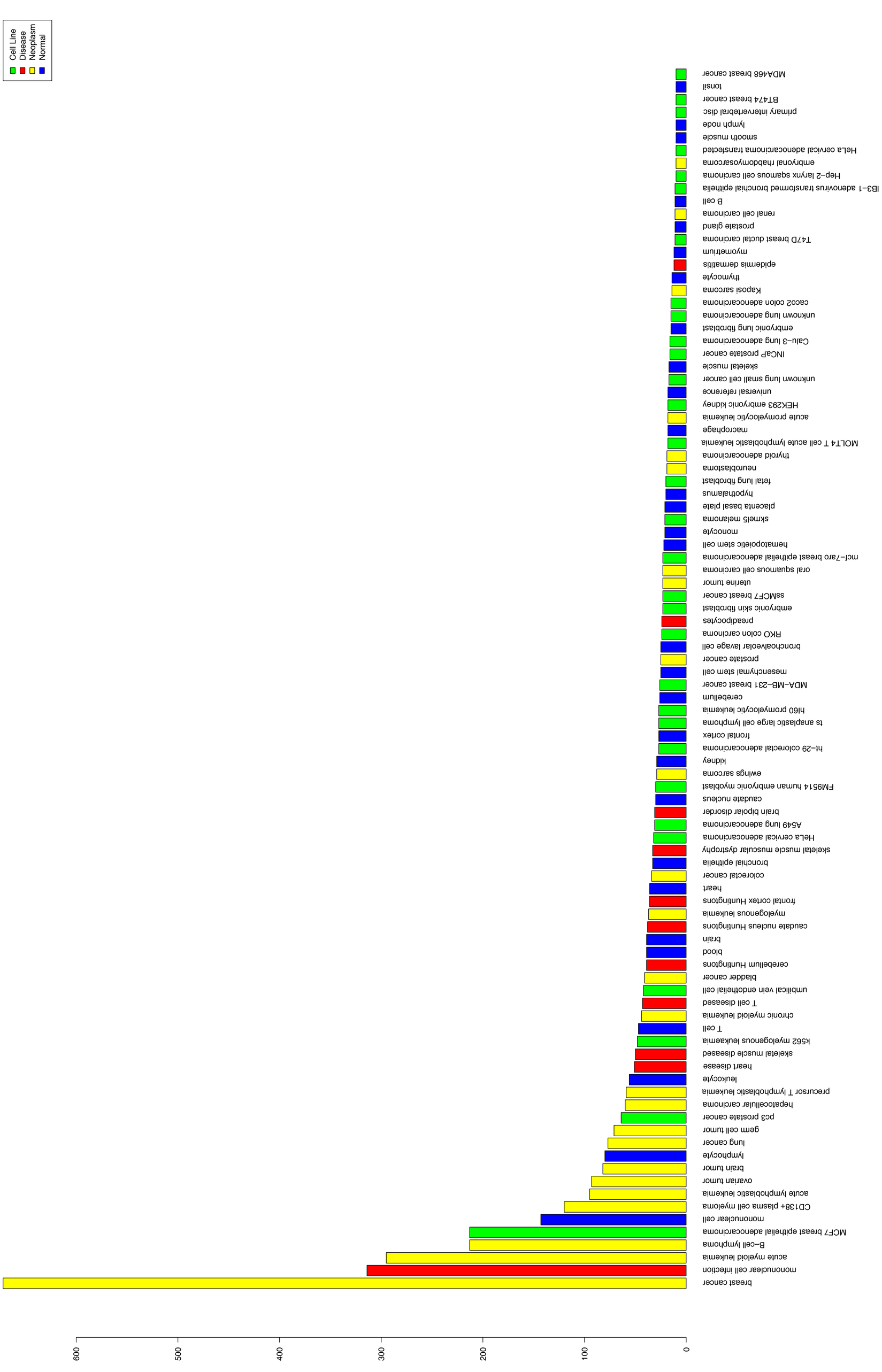
Cell Line
Disease
Neoplasm
Normal
0
100
200
300
400
500
600

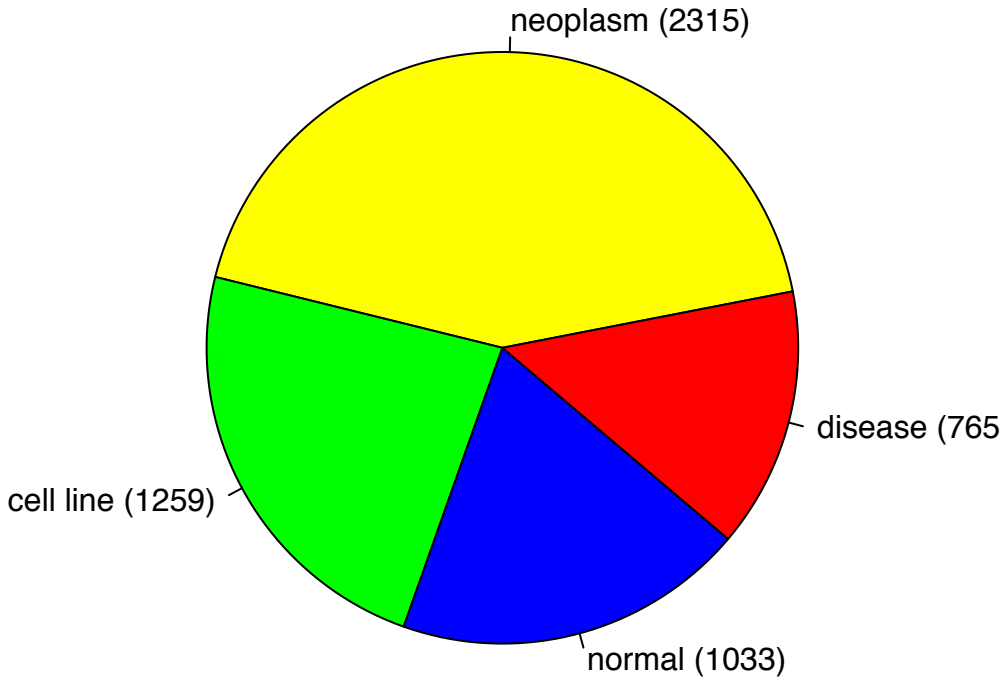
Supplementary Figure 1a:

Size distribution of 369 biological groups. Each bar represents the number of samples in the biological group. The bars are colored by 4 meta-groups (see legend).


Cell Line
Disease
Neoplasm
Normal
0
100
200
300
400
500
600
breast cancer
mononuclear cell infection
acute myeloid leukemia
B−cell lymphoma
MCF7 breast epithelial adenocarcinoma
mononuclear cell
CD138+ plasma cell myeloma
acute lymphoblastic leukemia
ovarian tumor
brain tumor
lymphocyte
lung cancer
germ cell tumor
pc3 prostate cancer
hepatocellular carcinoma
precursor T lymphoblastic leukemia
leukocyte
heart disease
skeletal muscle diseased
K562 myelogenous leukaemia
T cell
chronic myeloid leukemia
T cell diseased
umbilical vein endothelial cell
bladder cancer
cerebellum Huntingtons
blood
brain
caudate nucleus Huntingtons
myelogenous leukemia
frontal cortex Huntingtons
heart
colorectal cancer
bronchial epithelia
skeletal muscle muscular dystrophy
HeLa cervical adenocarcinoma
A549 lung adenocarcinoma
brain bipolar disorder
caudate nucleus
FM9514 human embryonic myoblast
ewings sarcoma
kidney
ht−29 colorectal adenocarcinoma
frontal cortex
ts anaplastic large cell lymphoma
hl60 promyelocytic leukemia
cerebellum
MDA−MB−231 breast cancer
mesenchymal stem cell
prostate cancer
bronchoalveolar lavage cell
RKO colon carcinoma
preadipocytes
embryonic skin fibroblast
ssMCF7 breast cancer
uterine tumor
oral squamous cell carcinoma
mcf−7aro breast epithelial adenocarcinoma
hematopoietic stem cell
monocyte
skmel5 melanoma
placenta basal plate
hypothalamus
fetal lung fibroblast
neuroblastoma
thyroid adenocarcinoma
MOLT4 T cell acute lymphoblastic leukemia
macrophage
acute promyelocytic leukemia
HEK293 embryonic kidney
universal reference
unknown lung small cell cancer
skeletal muscle
LNCaP prostate cancer
Calu−3 lung adenocarcinoma
embryonic lung fibroblast
unknown lung adenocarcinoma
caco2 colon adenocarcinoma
Kaposi sarcoma
thymocyte
epidermis dermatitis
myometrium
T47D breast ductal carcinoma
prostate gland
renal cell carcinoma
B cell
IB3−1 adenovirus transformed bronchial epithelia
Hep−2 larynx squamous cell carcinoma
embryonal rhabdomyosarcoma
HeLa cervical adenocarcinoma transfected
smooth muscle
lymph node
primary intervertebral disc
BT474 breast cancer
tonsil
MDA468 breast cancer

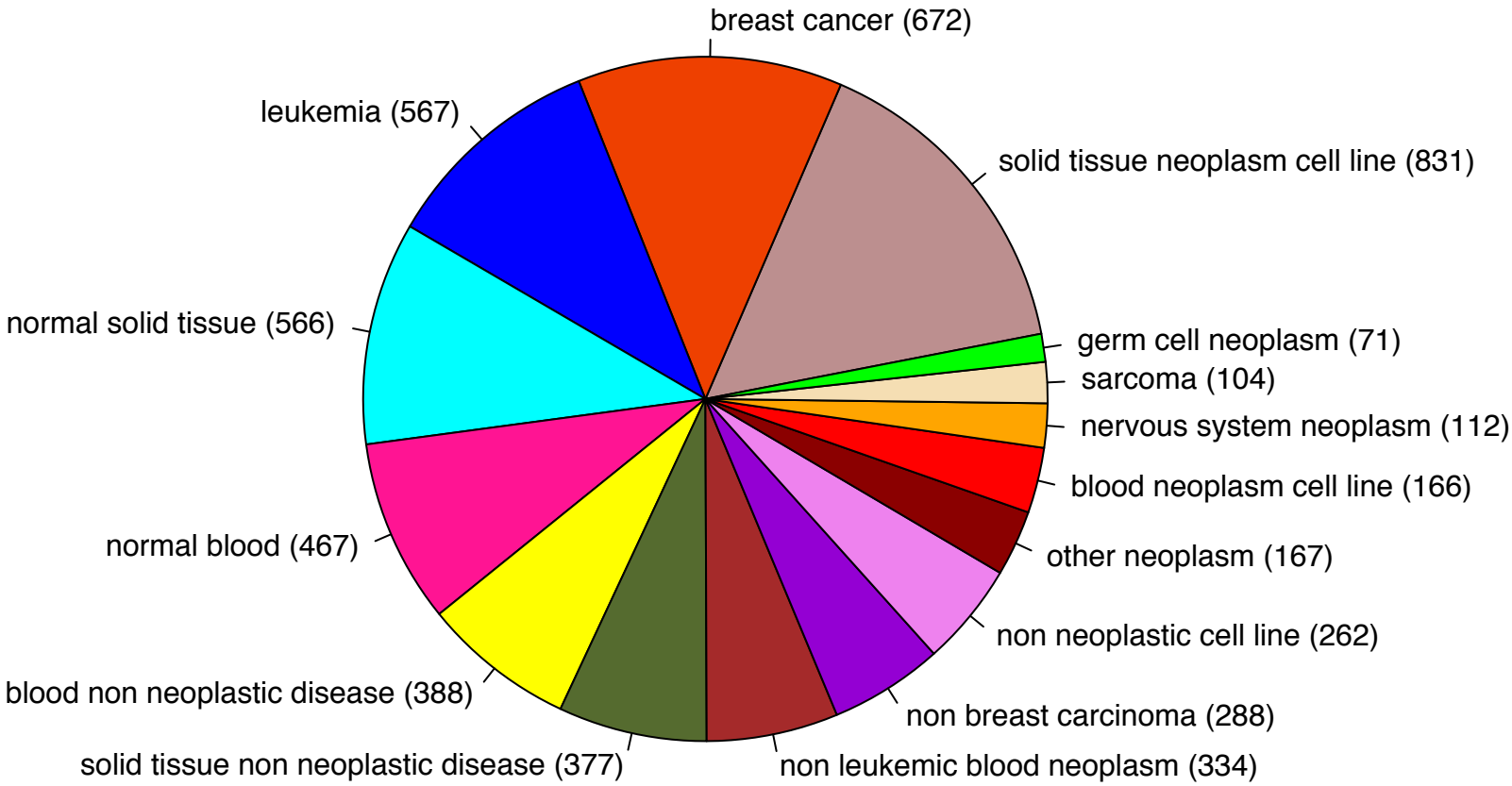
Supplementary Figure 1b:

Size distribution of the 96 larger biological groups, each containing at least 10 replicates representing the total of 4608 samples. The bars are colored by 4 meta-groups (see legend).


neoplasm (2315)
disease (765)
cell line (1259)
normal (1033)

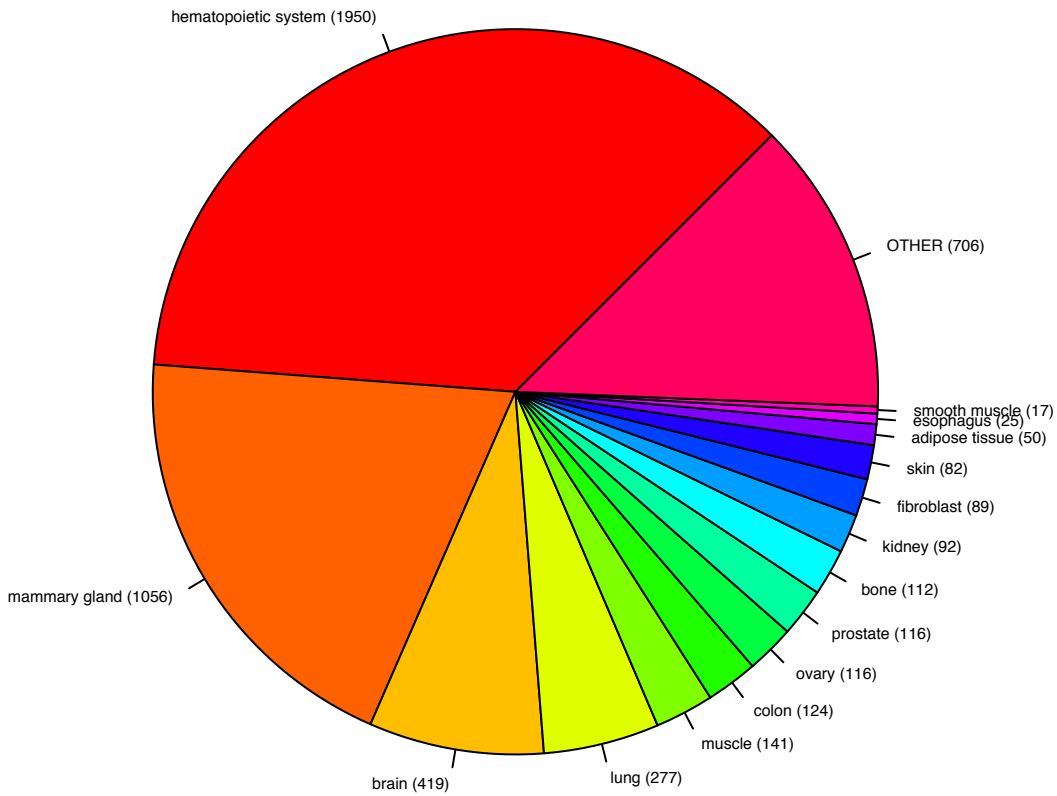
Supplementary Figure 1c:

The number of samples in each of the 4 meta-groups – cell line based samples, neoplastic primary tissue samples, non-neoplastic primary tissue samples, and 'normal' primary tissue samples.


breast cancer (672)
solid tissue neoplasm cell line (831)
germ cell neoplasm (71)
sarcoma (104)
nervous system neoplasm (112)
blood neoplasm cell line (166)
other neoplasm (167)
non neoplastic cell line (262)
non breast carcinoma (288)
non leukemic blood neoplasm (334)
solid tissue non neoplastic disease (377)
blood non neoplastic disease (388)
normal blood (467)
normal solid tissue (566)
leukemia (567)

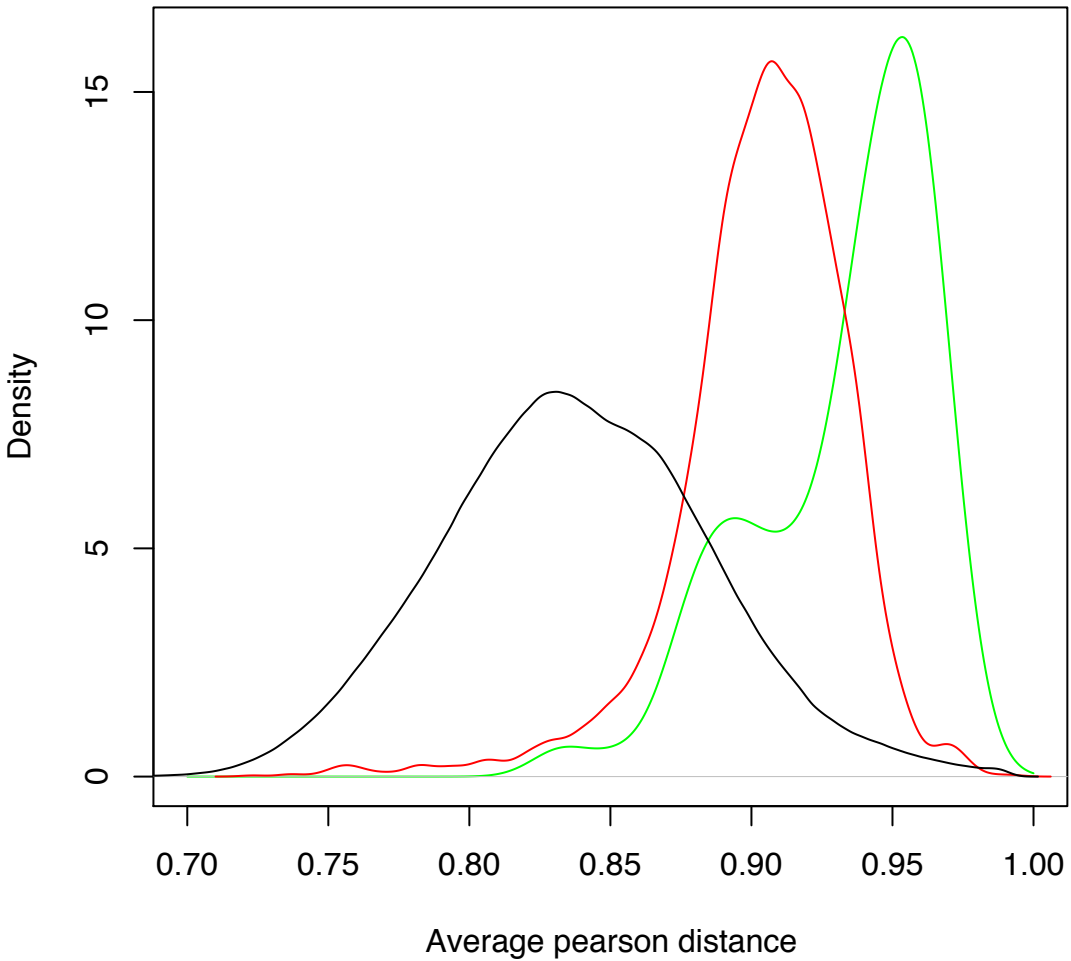
Supplementary Figure 1d:

The number of samples in each of the 15 meta-groups.


hematopoietic system (1950)
OTHER (706)
smooth muscle (17)
esophagus (25)
adipose tissue (50)
skin (82)
fibroblast (89)
kidney (92)
bone (112)
prostate (116)
ovary (116)
colon (124)
muscle (141)
lung (277)
brain (419)
mammary gland (1056)

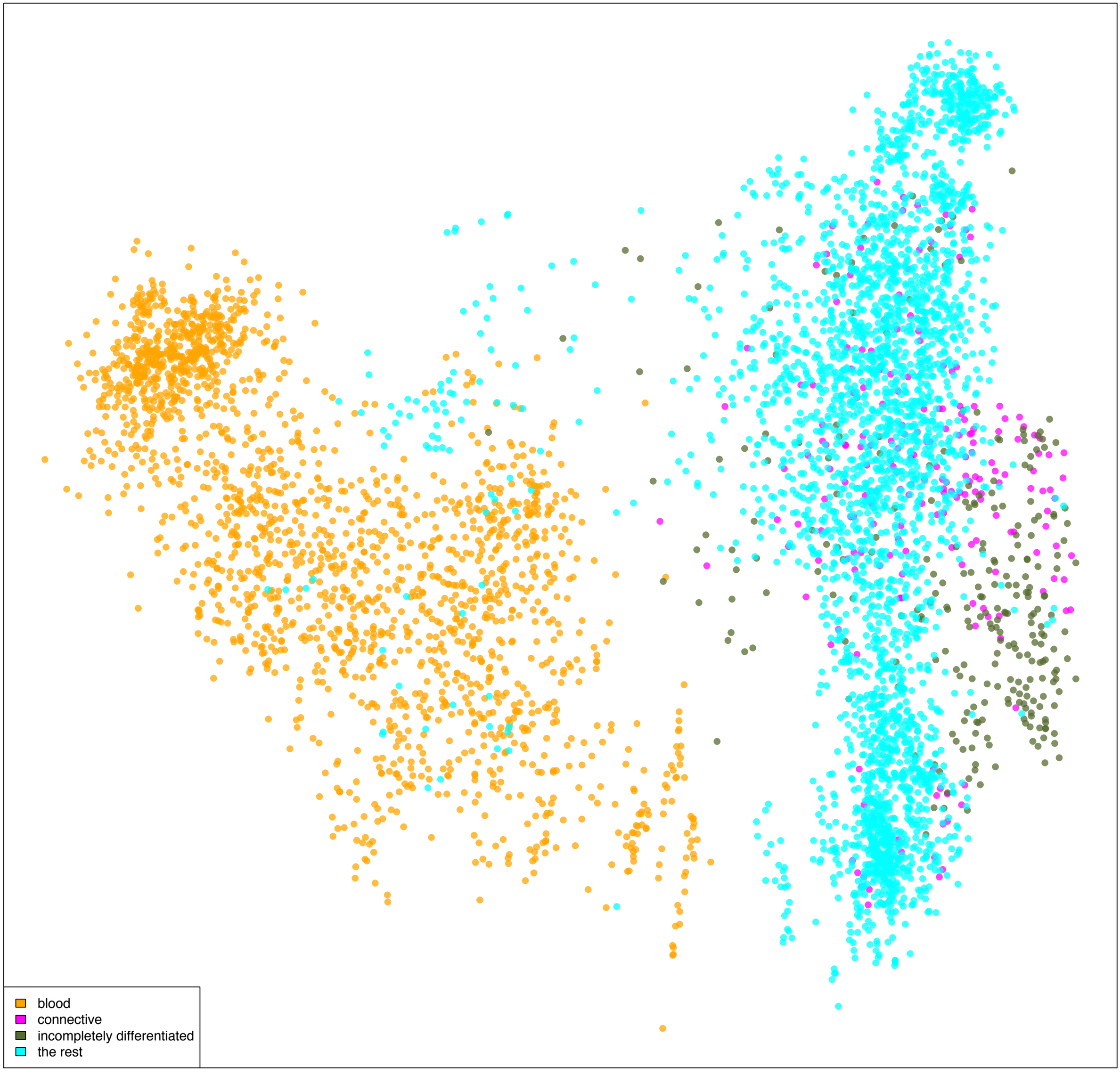
Supplementary Figure 1e:

The number of samples by different tissues of origin including both cell line and primary tissue samples. The meta-group 'other' contains samples for which tissue of origin including both cell line and primary tissues could not be constructed.


Density
0
5
10
15
0.70
0.75
0.80
0.85
0.90
0.95
1.00
Average pearson distance

Supplementary Figure 2:

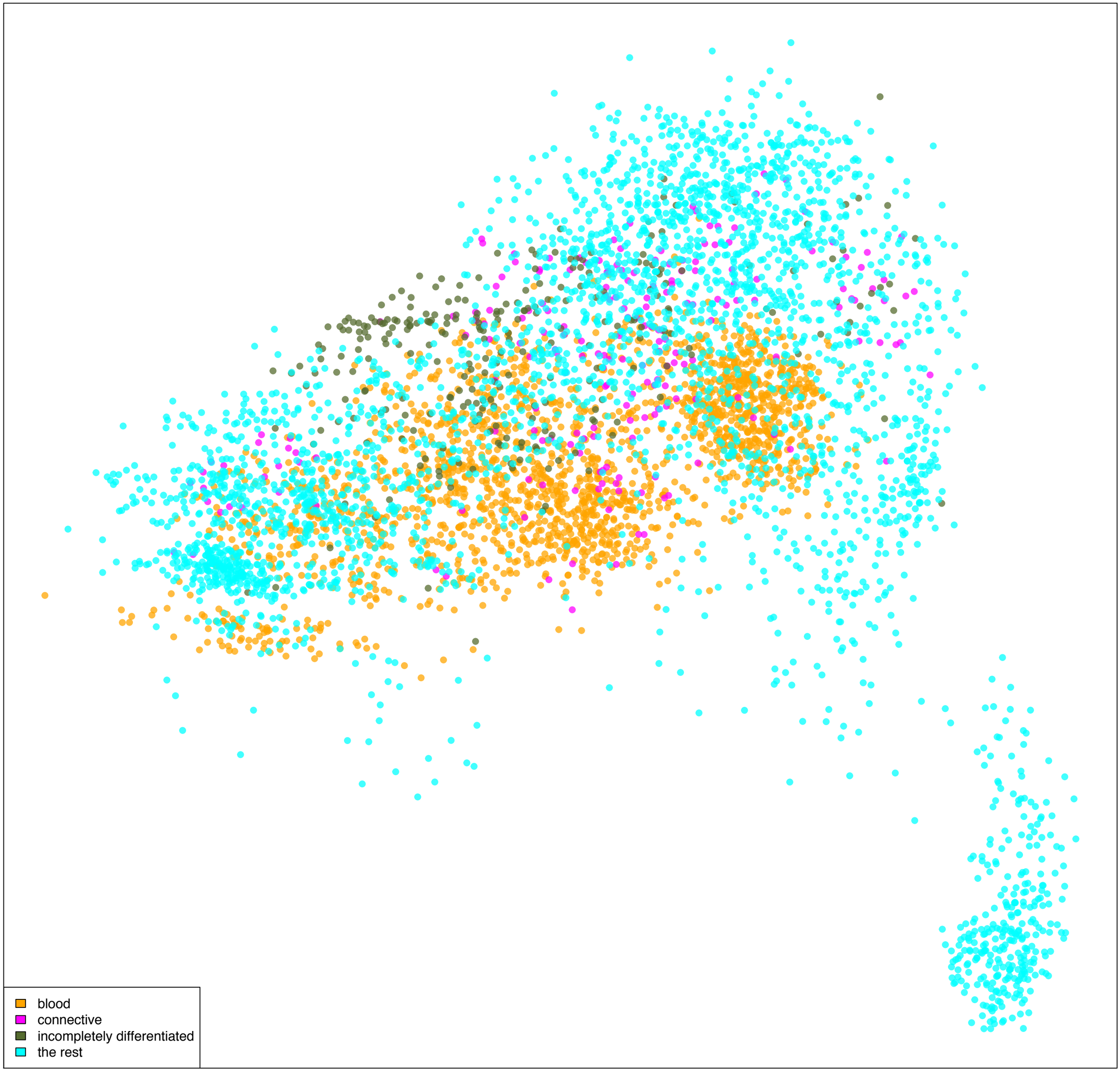
Pearson correlation coefficients between samples. Distribution of all pairwise correlations between the samples in the dataset (black), distribution of average similarities between the sample subgroups from different laboratories within the same biological group (green), distribution of average similarities between the sample subgroups from different biological groups within the same laboratory (red).


blood
connective
incompletely differentiated
the rest

Supplementary Figure 3a:

Principal component analysis. Each dot represents one of the 5372 samples in a multidimensional gene expression space projected on the 'principal' plane formed by pairs of first principal components. The dots are colored (semi-transparently) according to the biological group the sample belongs to.

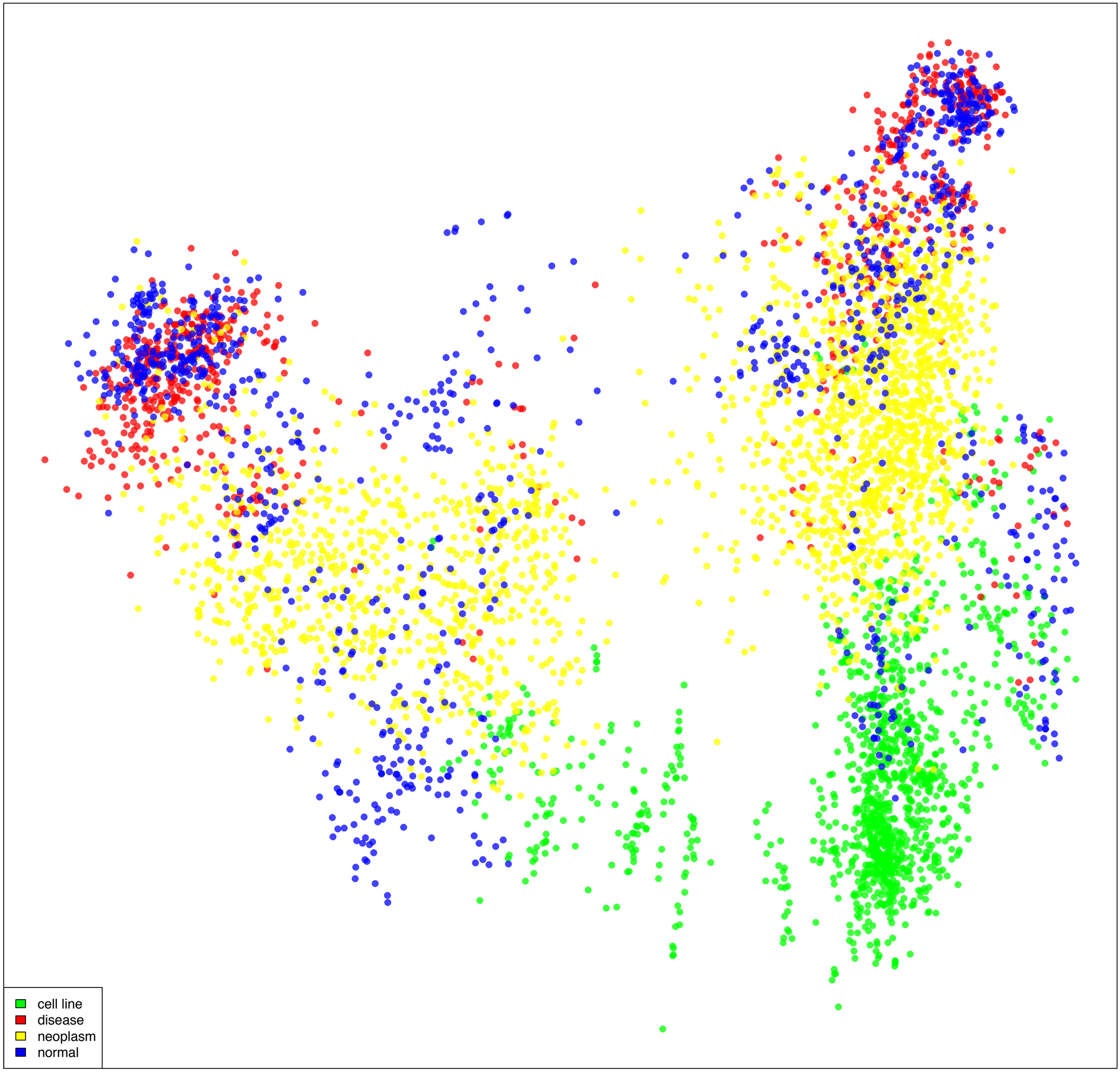
Samples of blood (orange), connective tissues (magenta), incompletely differentiated cells (green) and the rest (cyan) colored on the 1st and 2nd principal component plane.


blood
connective
incompletely differentiated
the rest

Supplementary Figure 3a23:

Principal component analysis. Each dot represents one of the 5372 samples in a multidimensional gene expression space projected on the 'principal' plane formed by pairs of first principal components. The dots are colored (semi-transparently) according to the biological group the sample belongs to.

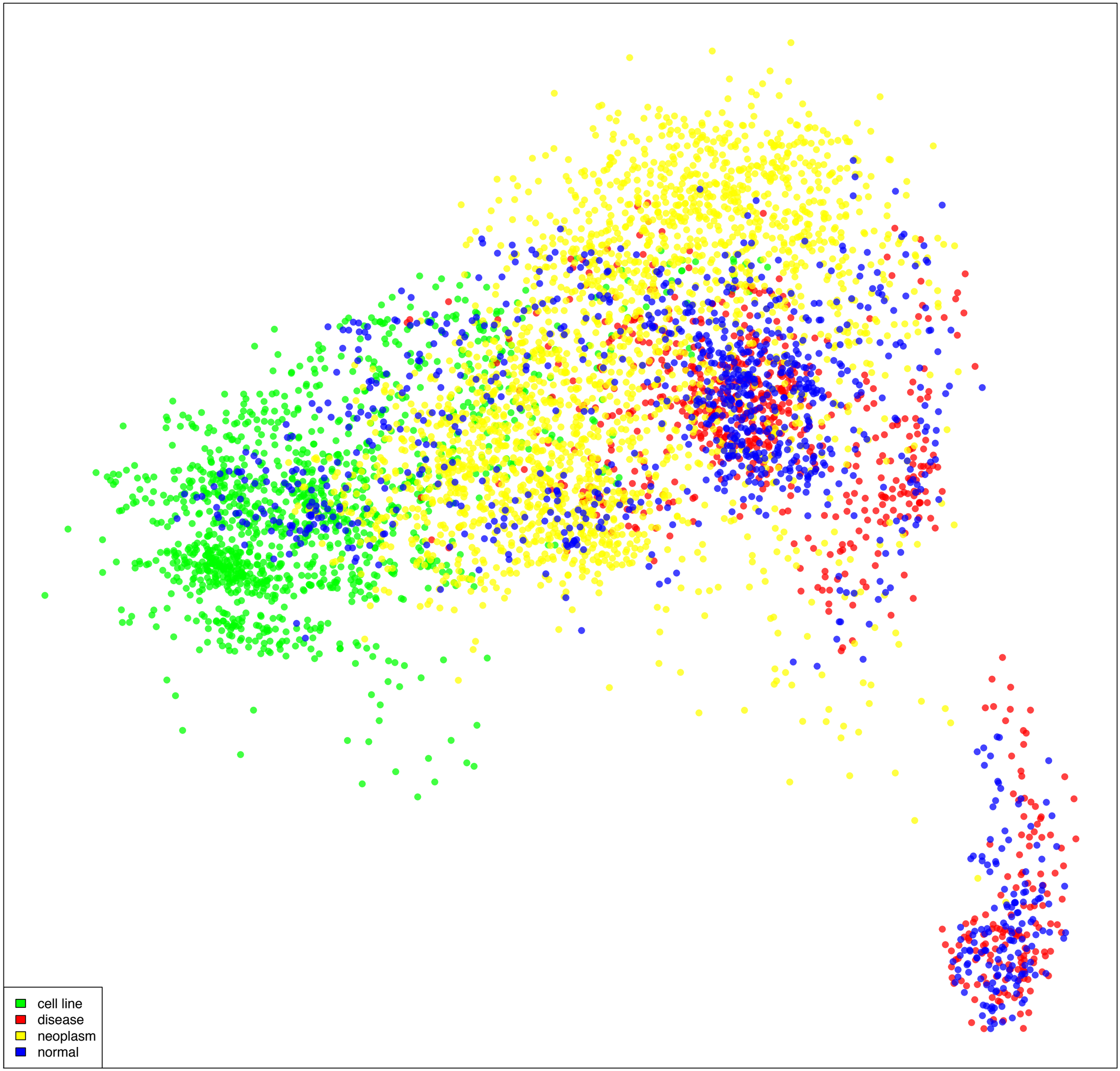
Samples of blood (orange), connective tissues (magenta), incompletely differentiated cells (green) and the rest (cyan) colored on the 2nd and 3rd principal component plane.


cell line
disease
neoplasm
normal

Supplementary Figure 3b:

Principal component analysis. Each dot represents one of the 5372 samples in a multidimensional gene expression space projected on the 'principal' plane formed by pairs of first principal components. The dots are colored (semi-transparently) according to the biological group the sample belongs to.

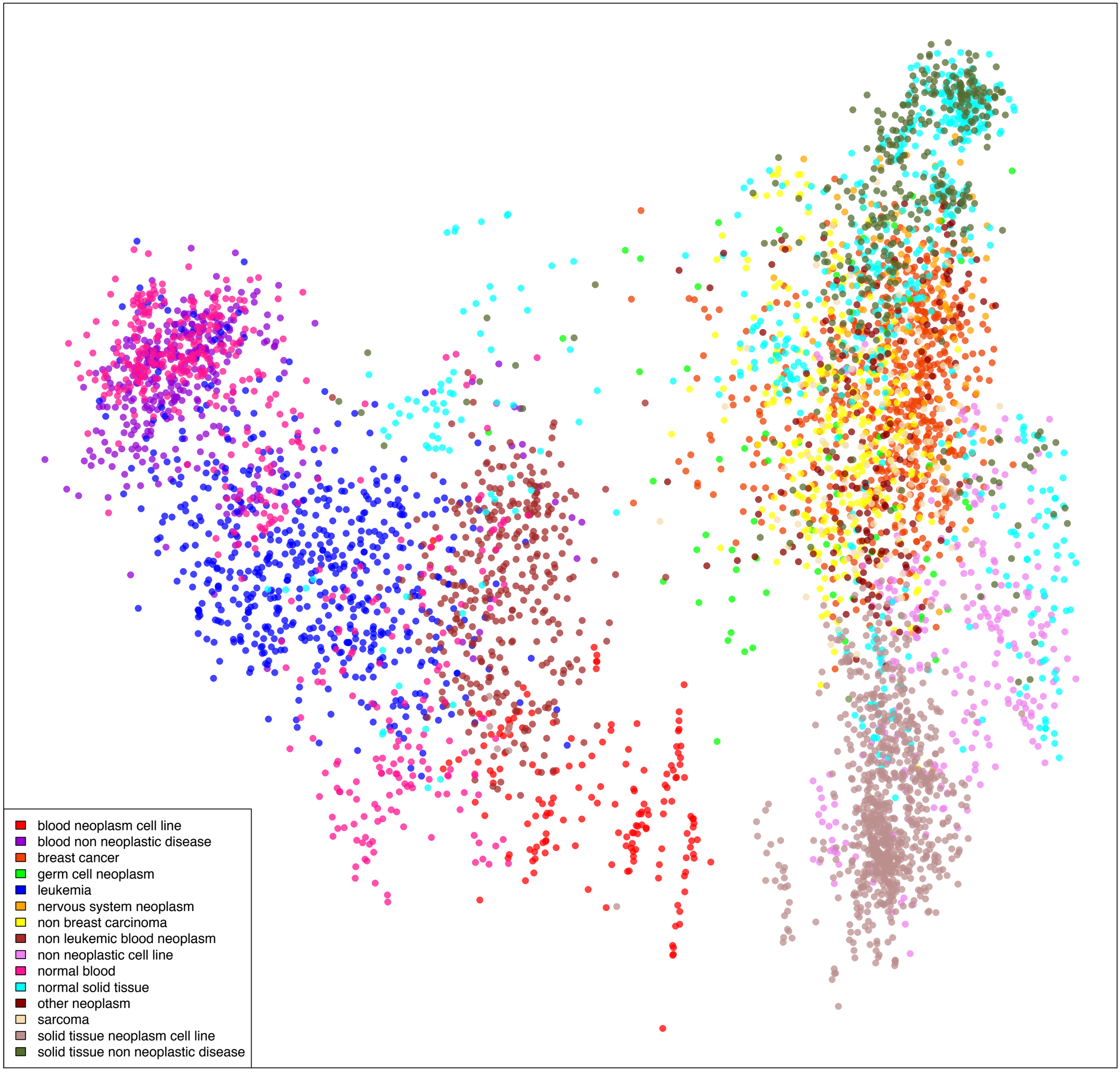
Samples colored by 4 meta-groups (see legend) on the 1st and 2nd principal component plane.


cell line
disease
neoplasm
normal

Supplementary Figure 3b23:

Principal component analysis. Each dot represents one of the 5372 samples in a multidimensional gene expression space projected on the 'principal' plane formed by pairs of first principal components. The dots are colored (semi-transparently) according to the biological group the sample belongs to.

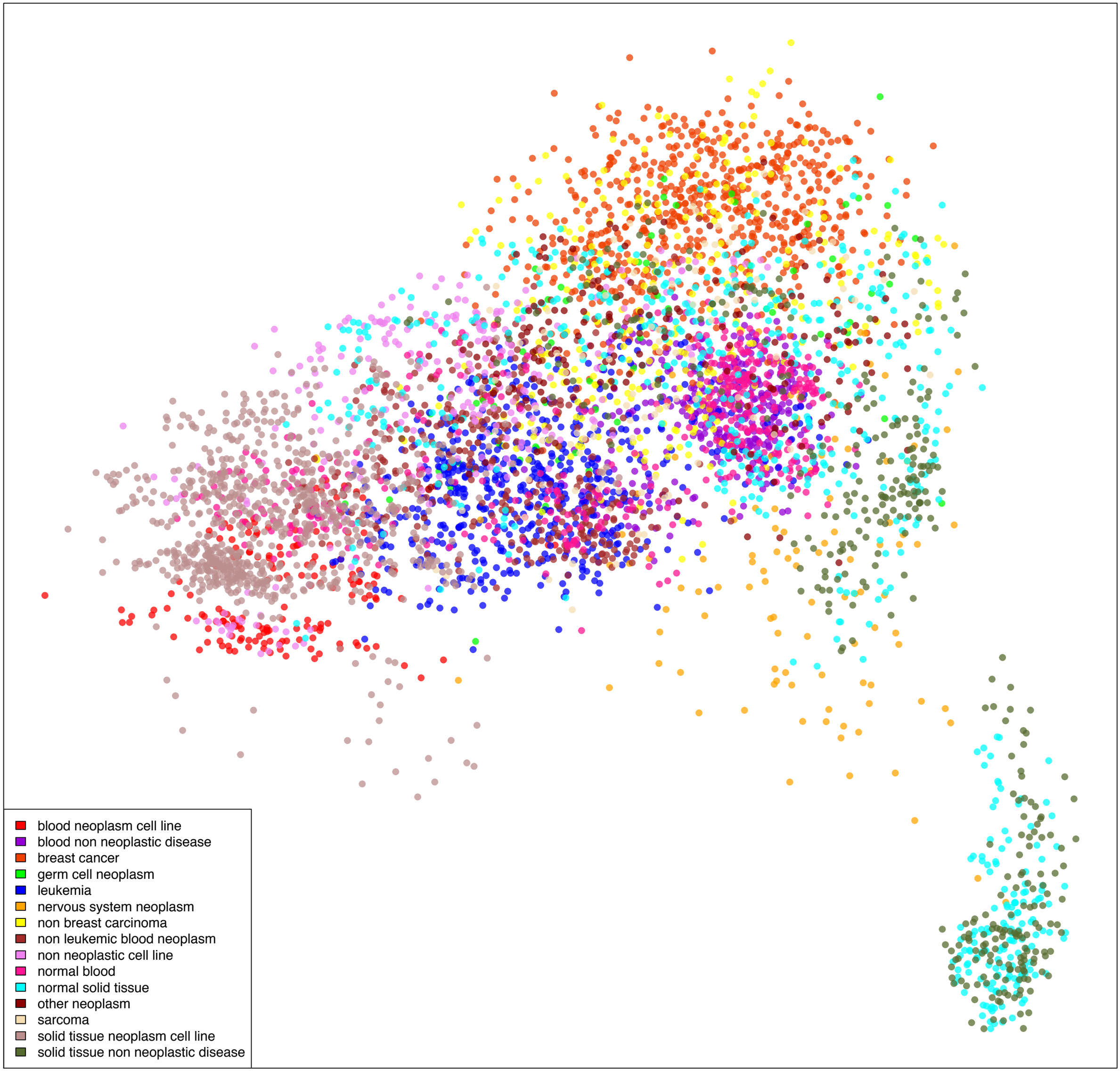
Samples colored by 4 meta-groups (see legend) on the 2nd and 3rd principal component plane.


blood neoplasm cell line
blood non neoplastic disease
breast cancer
germ cell neoplasm
leukemia
nervous system neoplasm
non breast carcinoma
non leukemic blood neoplasm
non neoplastic cell line
normal blood
normal solid tissue
other neoplasm
sarcoma
solid tissue neoplasm cell line
solid tissue non neoplastic disease

Supplementary Figure 3c:

Principal component analysis. Each dot represents one of the 5372 samples in a multidimensional gene expression space projected on the 'principal' plane formed by pairs of first principal components. The dots are colored (semi-transparently) according to the biological group the sample belongs to.

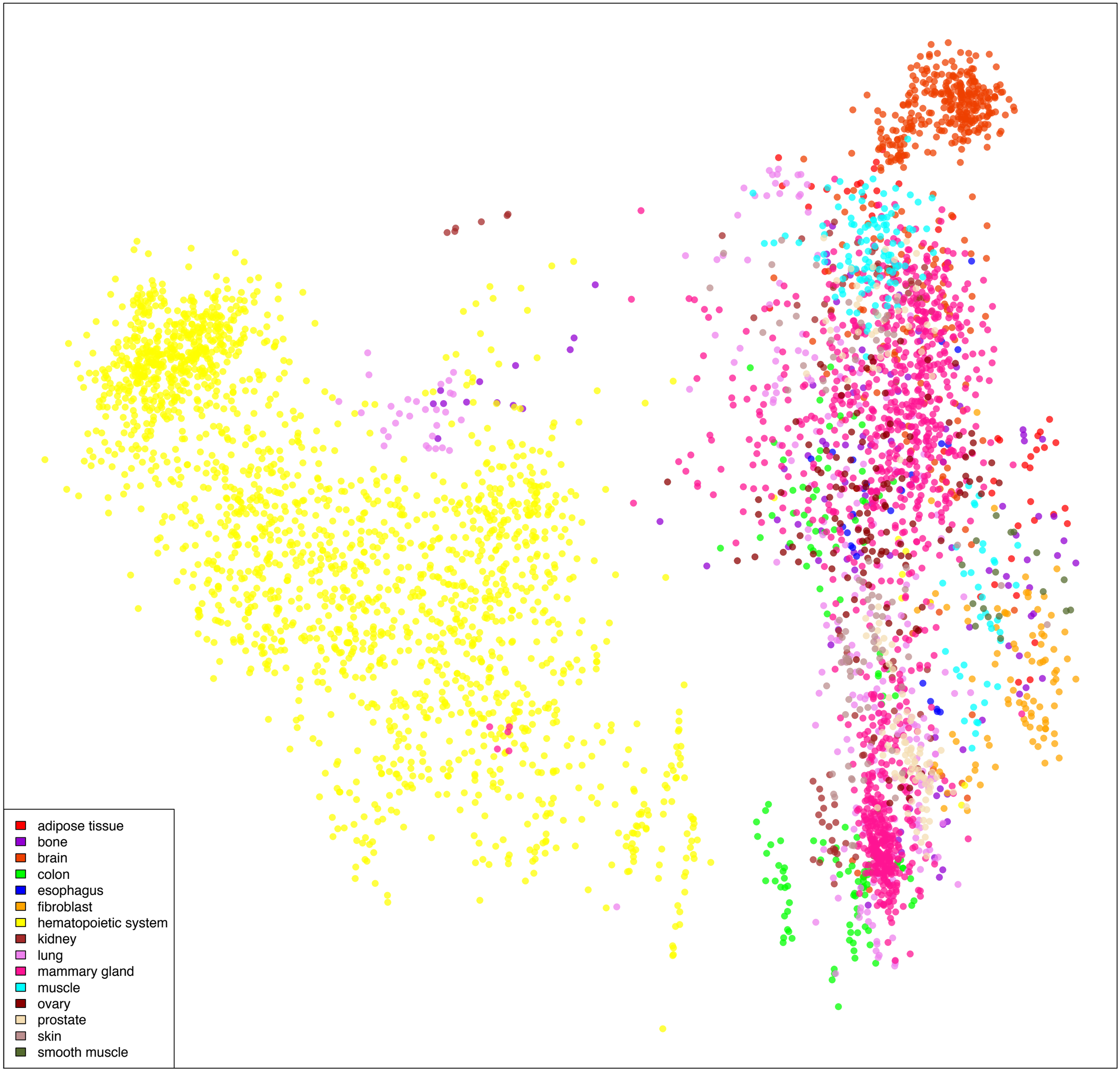
Samples colored by 15 meta-groups (see legend) on the 1st and 2nd principal component plane.


blood neoplasm cell line
blood non neoplastic disease
breast cancer
germ cell neoplasm
leukemia
nervous system neoplasm
non breast carcinoma
non leukemic blood neoplasm
non neoplastic cell line
normal blood
normal solid tissue
other neoplasm
sarcoma
solid tissue neoplasm cell line
solid tissue non neoplastic disease

Supplementary Figure 3c23:

Principal component analysis. Each dot represents one of the 5372 samples in a multidimensional gene expression space projected on the 'principal' plane formed by pairs of first principal components. The dots are colored (semi-transparently) according to the biological group the sample belongs to.

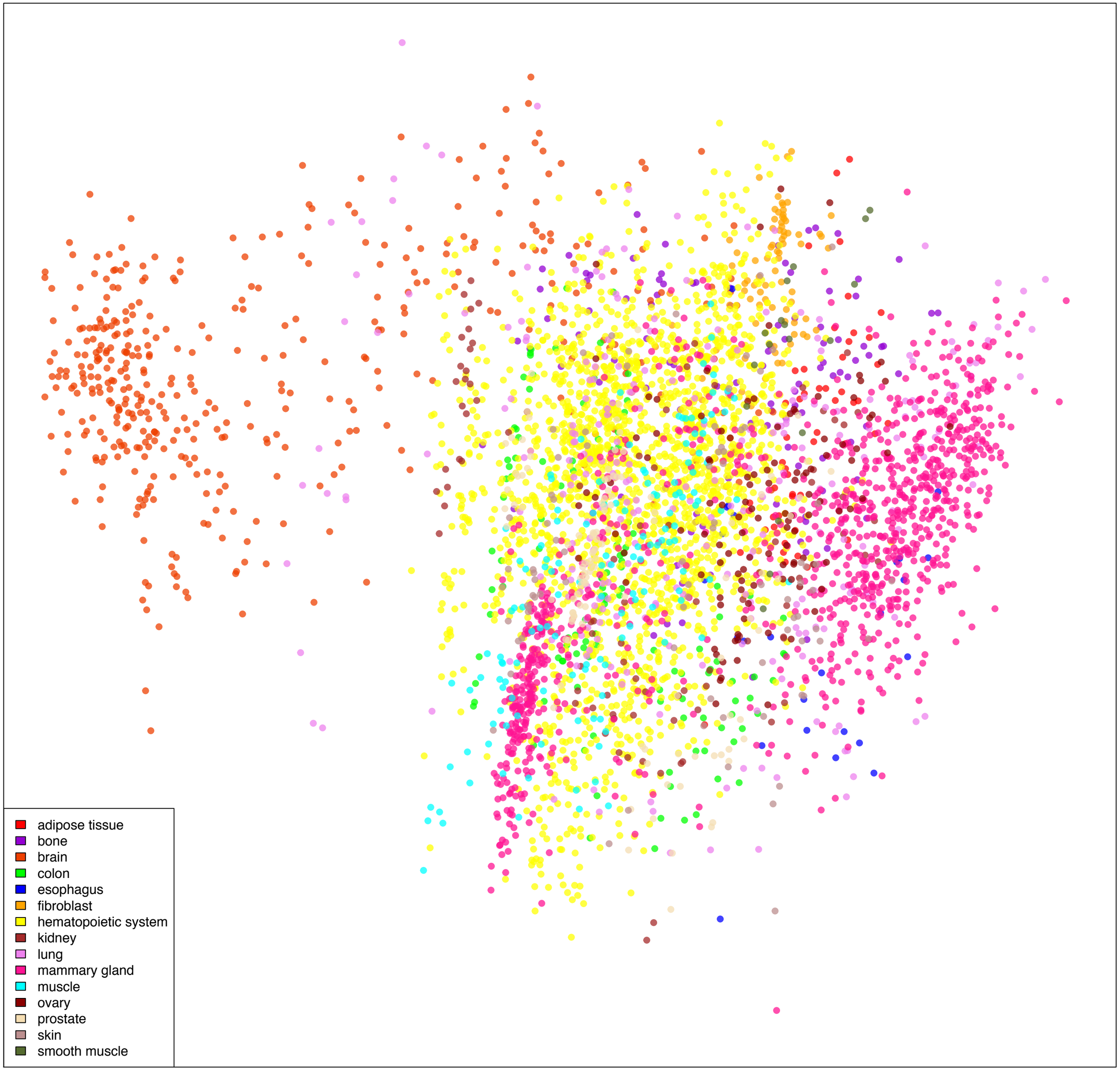
Samples colored by 15 meta-groups (see legend) on the 2nd and 3rd principal component plane.


adipose tissue
bone
brain
colon
esophagus
fibroblast
hematopoietic system
kidney
lung
mammary gland
muscle
ovary
prostate
skin
smooth muscle

Supplementary Figure 3d:

Principal component analysis. Each dot represents one of the 5372 samples in a multidimensional gene expression space projected on the 'principal' plane formed by pairs of first principal components. The dots are colored (semi-transparently) according to the biological group the sample belongs to.

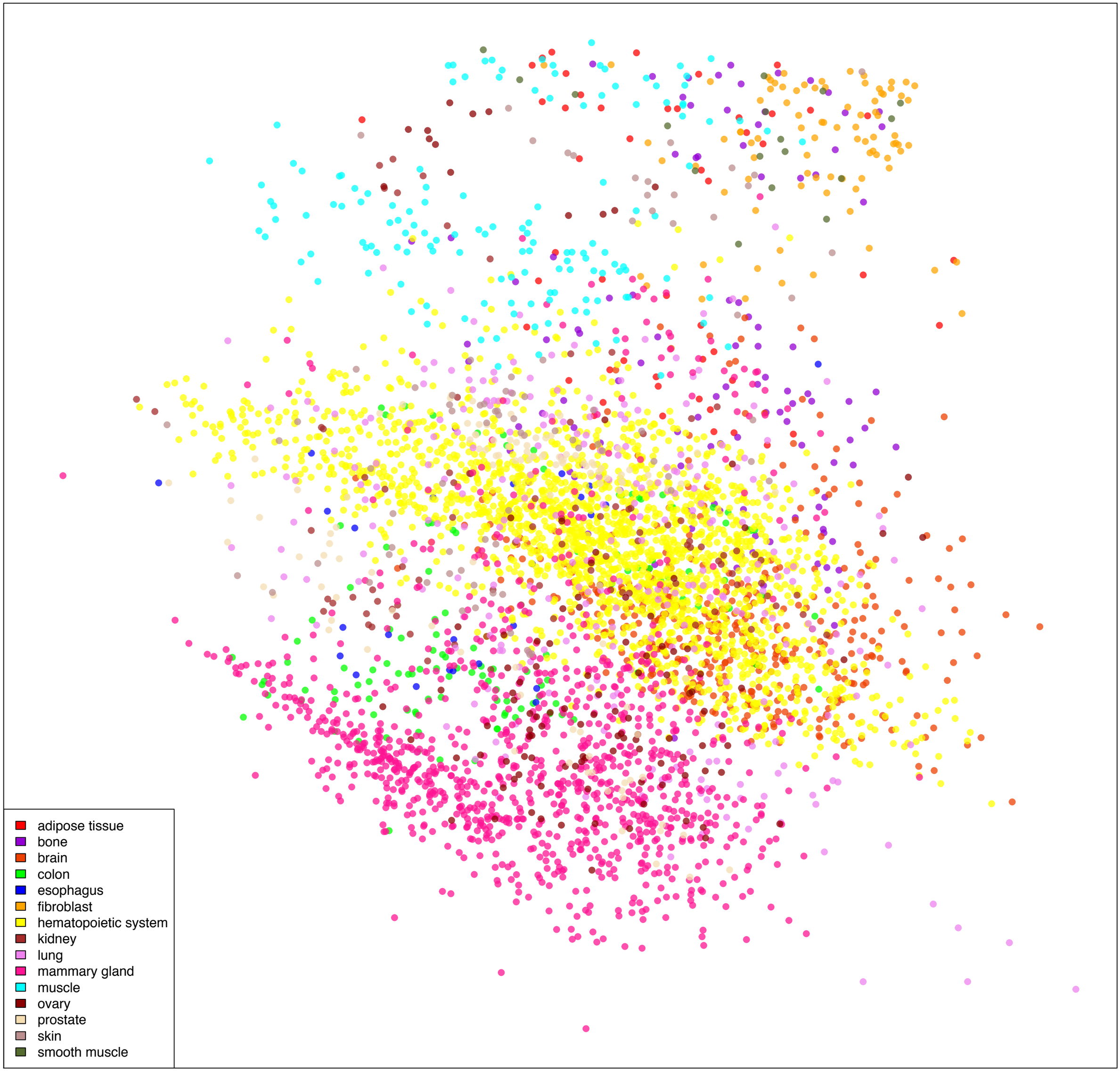
Samples colored by 15 groups of tissues of origin (see legend) on the 1st and 2nd principal component plane.


adipose tissue
bone
brain
colon
esophagus
fibroblast
hematopoietic system
kidney
lung
mammary gland
muscle
ovary
prostate
skin
smooth muscle

Supplementary Figure 3d34:

Principal component analysis. Each dot represents one of the 5372 samples in a multidimensional gene expression space projected on the 'principal' plane formed by pairs of first principal components. The dots are colored (semi-transparently) according to the biological group the sample belongs to.

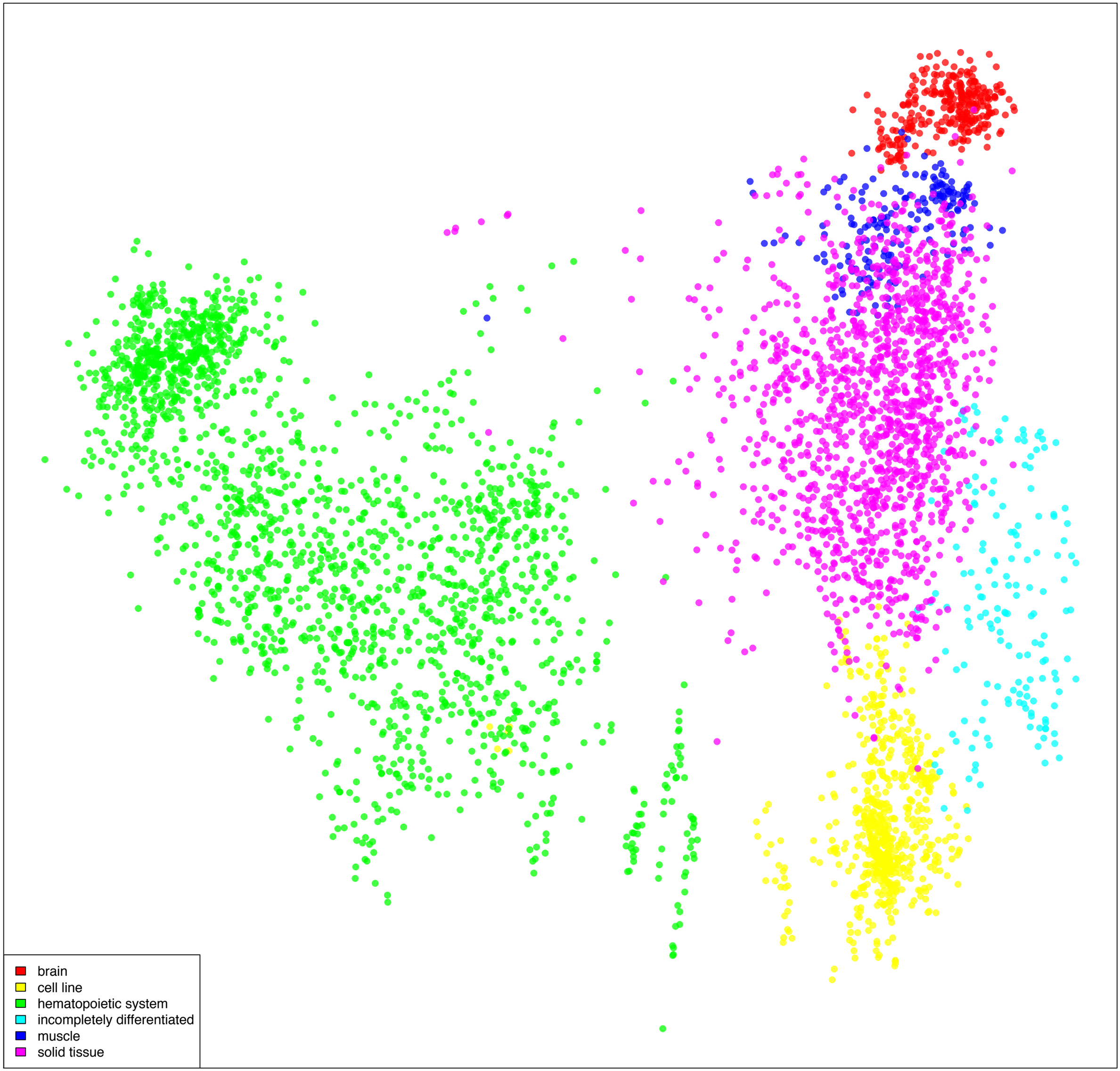
Samples colored by 15 groups of tissues of origin (see legend) on the 3rd and 4th principal component plane.


adipose tissue
bone
brain
colon
esophagus
fibroblast
hematopoietic system
kidney
lung
mammary gland
muscle
ovary
prostate
skin
smooth muscle

Supplementary Figure 3d45:

Principal component analysis. Each dot represents one of the 5372 samples in a multidimensional gene expression space projected on the 'principal' plane formed by pairs of first principal components. The dots are colored (semi-transparently) according to the biological group the sample belongs to.

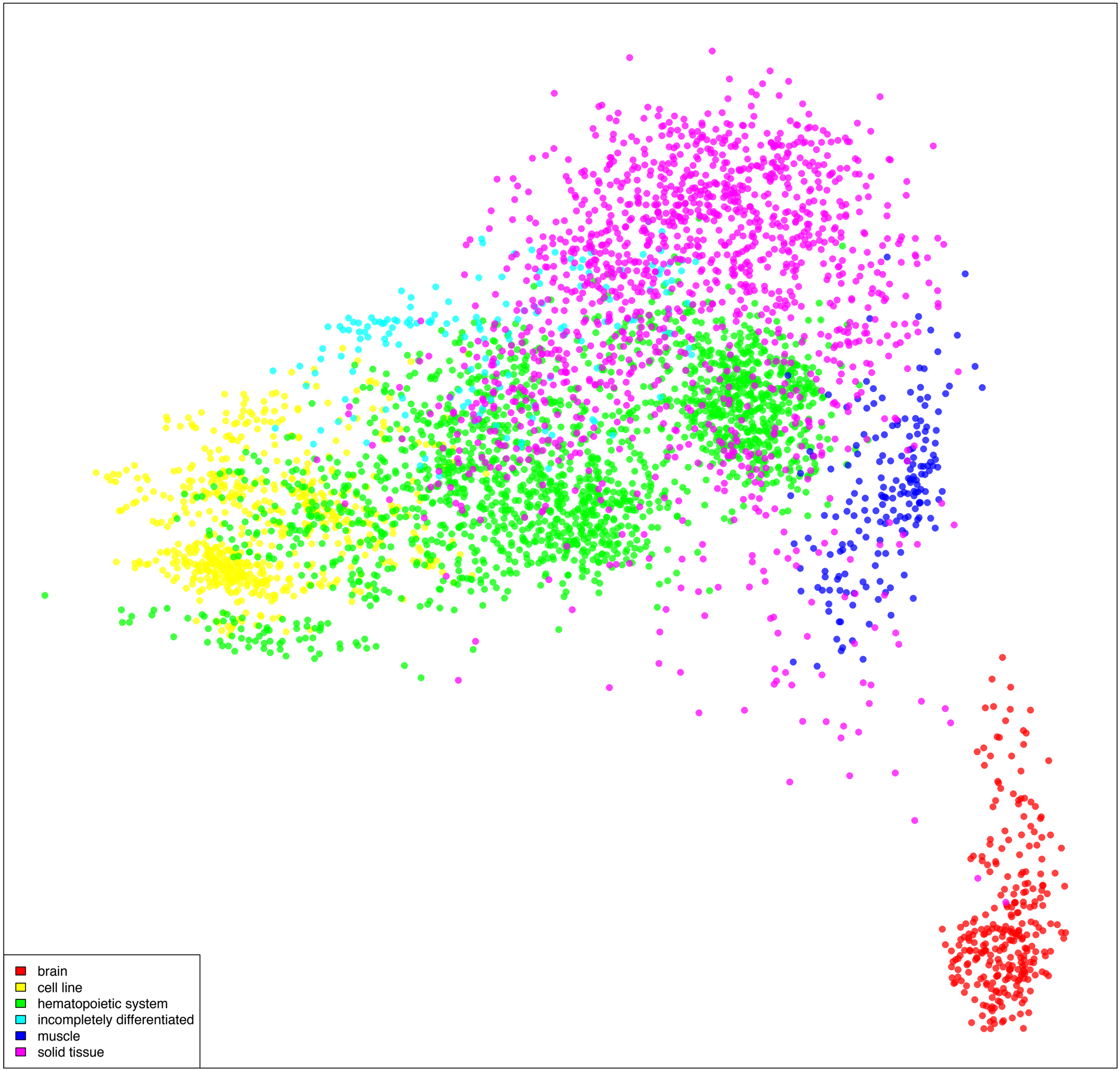
Samples colored by 15 groups of tissues of origin (see legend) on the 4th and 5th principal component plane. One can observe, that the 5th component mammary gland cell line and primary tissue samples seem to be brought closer together. This is true also for several other tissue types, however we were not able to prove this statistically.


brain
cell line
hematopoietic system
incompletely differentiated
muscle
solid tissue

Supplementary Figure 3e:

Principal component analysis. Each dot represents one of the 5372 samples in a multidimensional gene expression space projected on the 'principal' plane formed by pairs of first principal components. The dots are colored (semi-transparently) according to the biological group the sample belongs to.

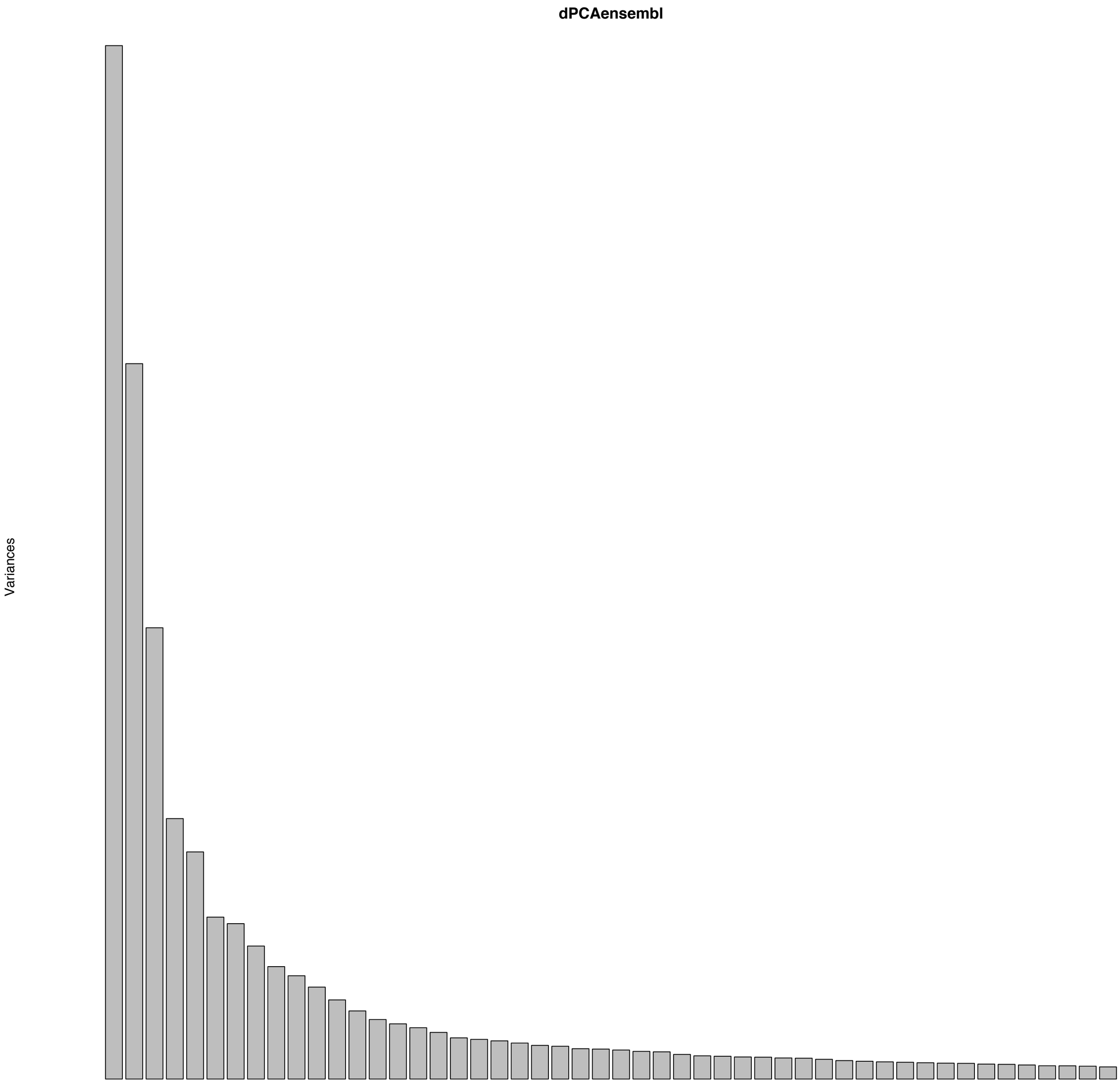
Samples colored by 6 clusters (see legend) identified on self-vs-self heat-map of 96 biological groups. The samples are visualized on the 1st and 2nd principal component plane.


brain
cell line
hematopoietic system
incompletely differentiated
muscle
solid tissue

Supplementary Figure 3e23:

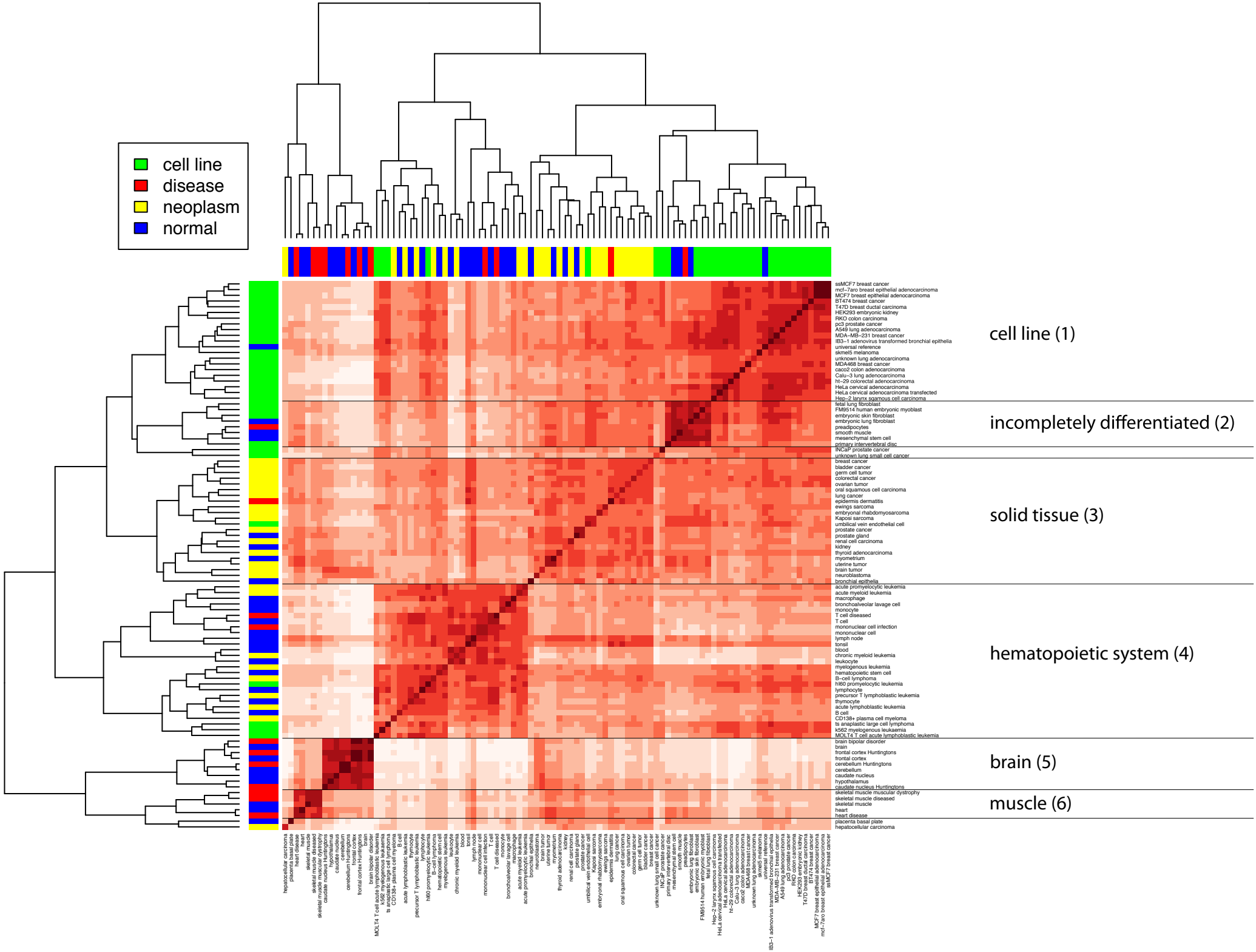
Principal component analysis. Each dot represents one of the 5372 samples in a multidimensional gene expression space projected on the 'principal' plane formed by pairs of first principal components. The dots are colored (semi-transparently) according to the biological group the sample belongs to.

Samples colored by 6 clusters (see legend) identified on self-vs-self heat-map of 96 biological groups. The samples are visualized on the 2nd and 3rd principal component plane.


dPCAensembl
Variances

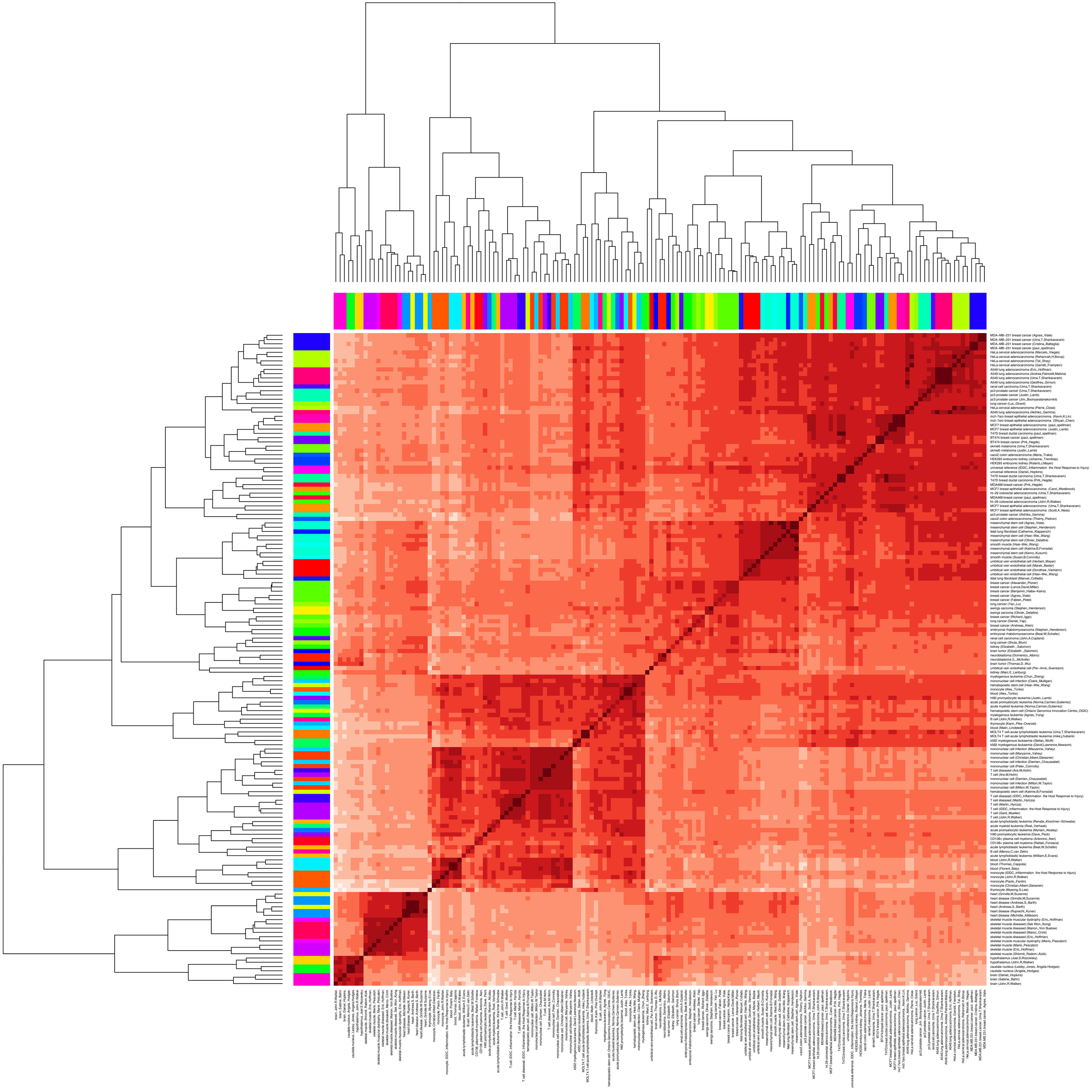
Supplementary Figure 3f:

The variances captured by first 50 principal components. The first 3 principal components explain ~37% of variability.


cell line
disease
neoplasm
normal
cell line (1)
incompletely differentiated (2)
solid tissue (3)
hematopoietic system (4)
brain (5)
muscle (6)

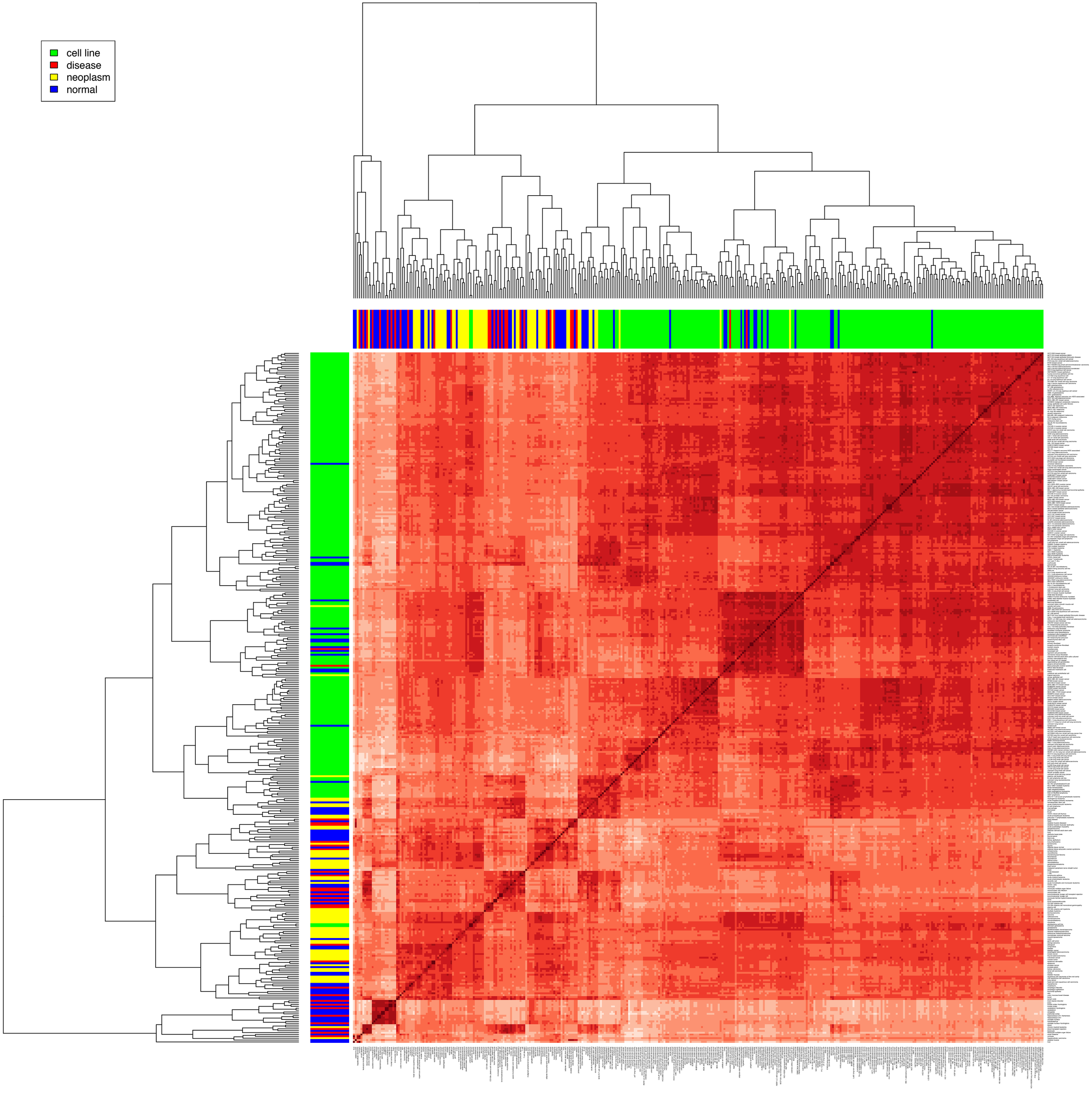
Supplementary Figure 4a:

Heat map of hierarchical clustering of the 96 larger biological groups compared to each other. Each group consists of 10 or more biological replicates. Similarity between groups is defined as the mean correlation coefficient between the sample expression profiles. Clustering is applied to sample groups. The diagonal represents the mean similarity between the elements within a particular group (thus, the more compact the group, the higher the value of its diagonal element). The heatmap uses a false color assigning dark red to maximum similarity and white to minimum. The legend top left indicates which major class of samples the particular group belongs to. Six major clusters can be observed: (1) solid tissue cell lines, (2) incompletely differentiated cell type and connective tissue cluster, (3) solid normal and neoplastic tissues, (4) hematopoietic system, (5) brain, (6) muscle and heart. Additionally there are 4 outliers - INCaP prostate cancer, lung small cell cancer, placenta basal plate, and hepatocellular carcinoma.



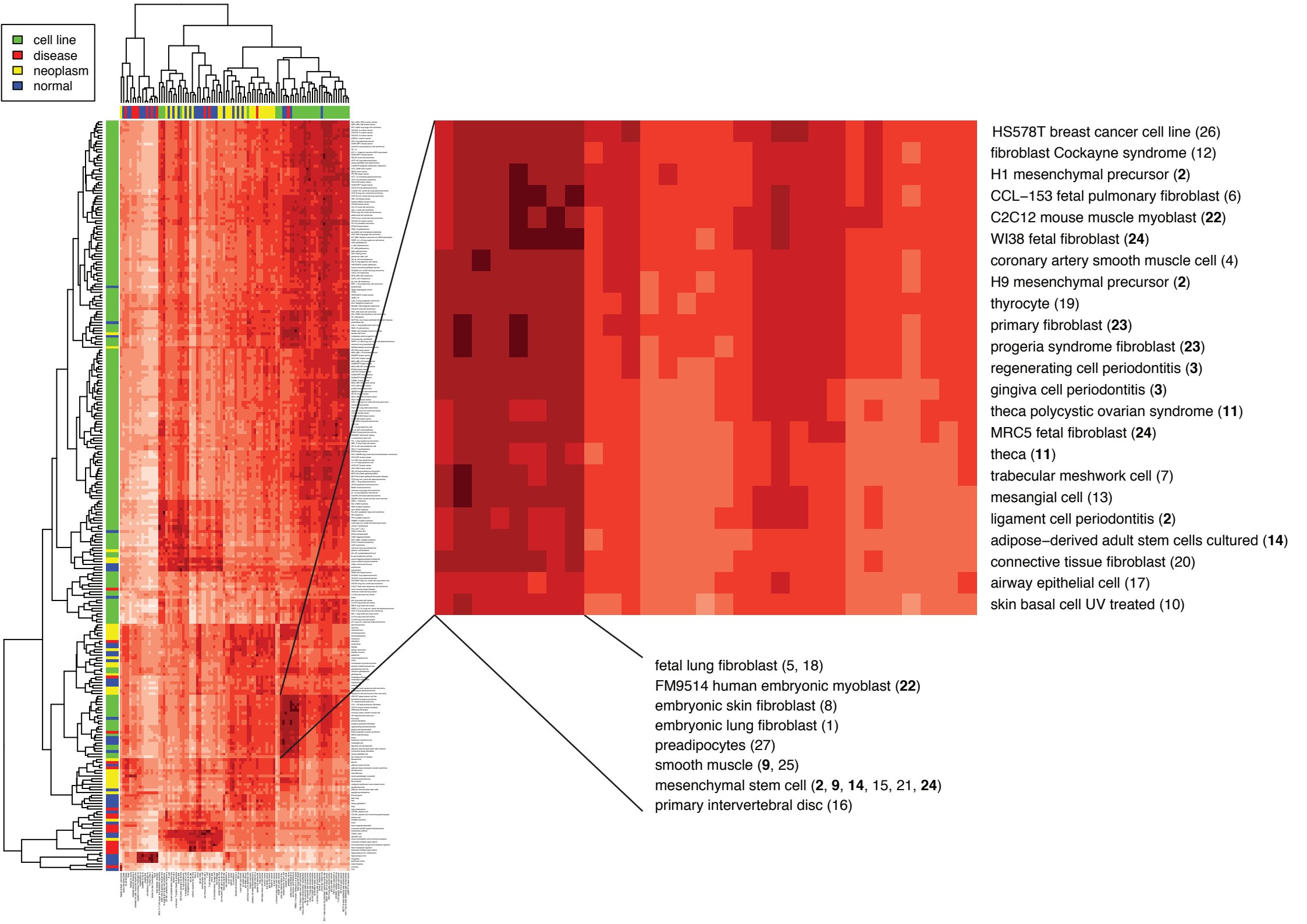
Supplementary Figure 4b:

Heat map comparison of 51 biological groups of at least 10 samples and 2 laboratories where the biological subgroups of laboratories are kept separate. Information about the laboratory is given in parenthesis. The colors at the left and top of the heat map depict the biological groups.


cell line
disease
neoplasm
normal

Supplementary Figure 4c:

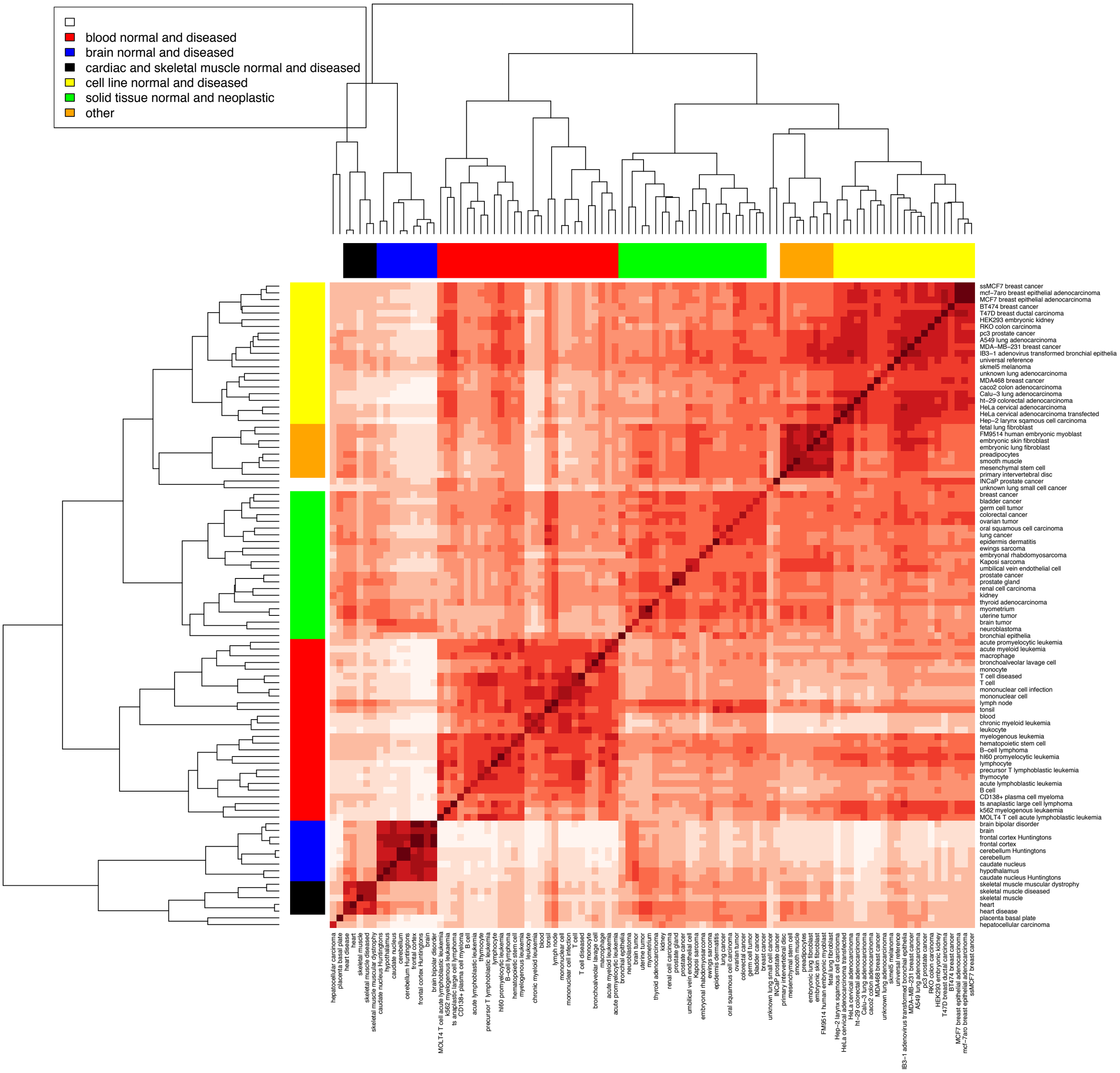
Heat map comparison of all 369 biological groups. The four colors at the left and top of the heat map depict which major class of samples the particular group belongs to (see legend).

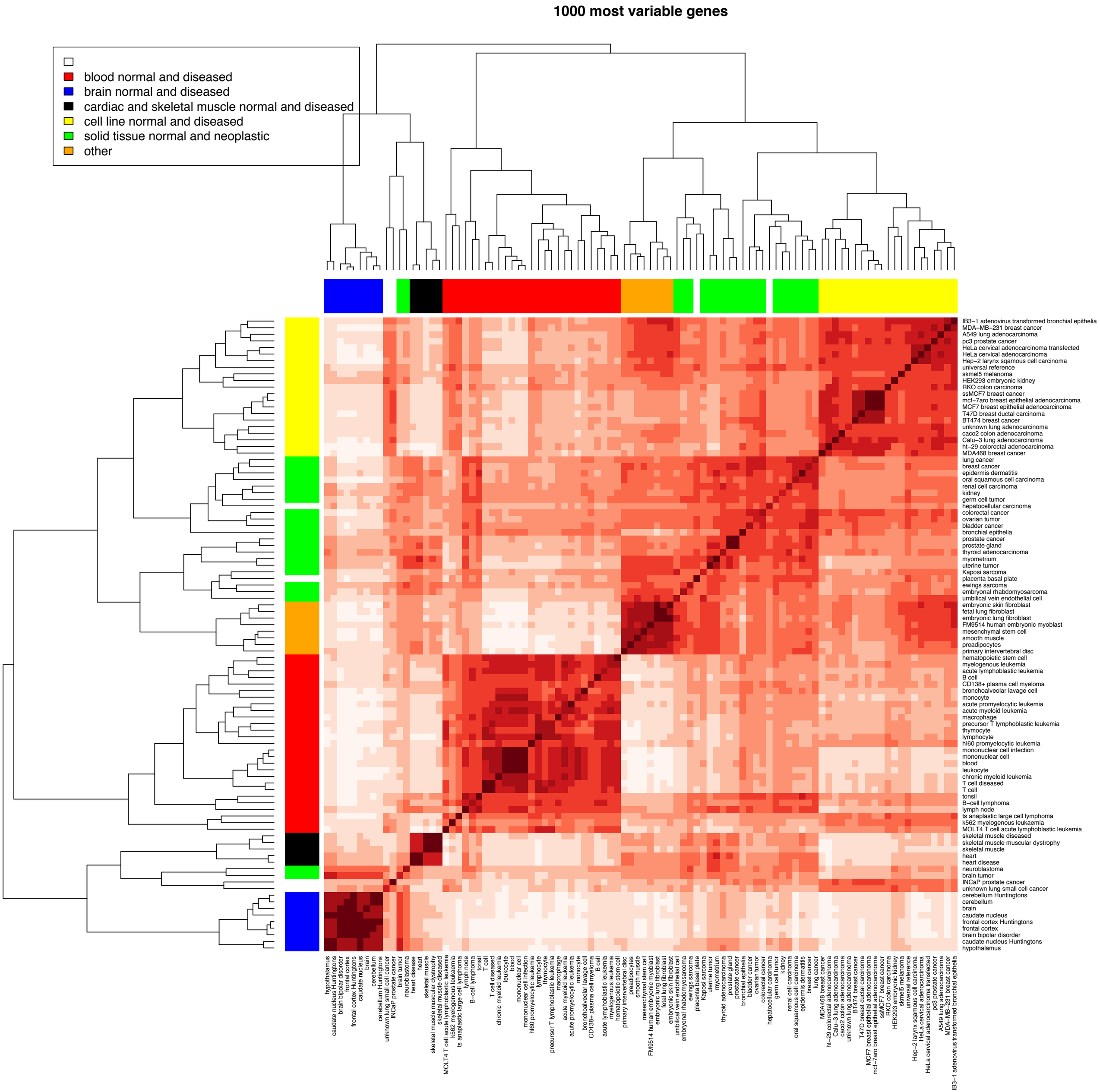

cell line
disease
neoplasm
normal
HS578T breast cancer cell line (26)
fibroblast Cockayne syndrome (12)
H1 mesenchymal precursor (**2**)
CCL−153 fetal pulmonary fibroblast (6)
C2C12 mouse muscle myoblast (**22**)
WI38 fetal fibroblast (**24**)
coronary artery smooth muscle cell (4)
H9 mesenchymal precursor (**2**)
thyrocyte (19)
primary fibroblast (**23**)
progeria syndrome fibroblast (**23**)
regenerating cell periodontitis (**3**)
gingiva cell periodontitis (**3**)
theca polycystic ovarian syndrome (**11**)
MRC5 fetal fibroblast (**24**)
theca (**11**)
trabecular meshwork cell (7)
mesangial cell (13)
ligament cell periodontitis (**2**)
adipose−derived adult stem cells cultured (**14**)
connective tissue fibroblast (20)
airway epithelial cell (17)
skin basal cell UV treated (10)
fetal lung fibroblast (5, 18)
FM9514 human embryonic myoblast (**22**)
embryonic skin fibroblast (8)
embryonic lung fibroblast (1)
preadipocytes (27)
smooth muscle (**9**, 25)
mesenchymal stem cell (**2**, **9**, **14**, 15, 21, **24**)
primary intervertebral disc (16)

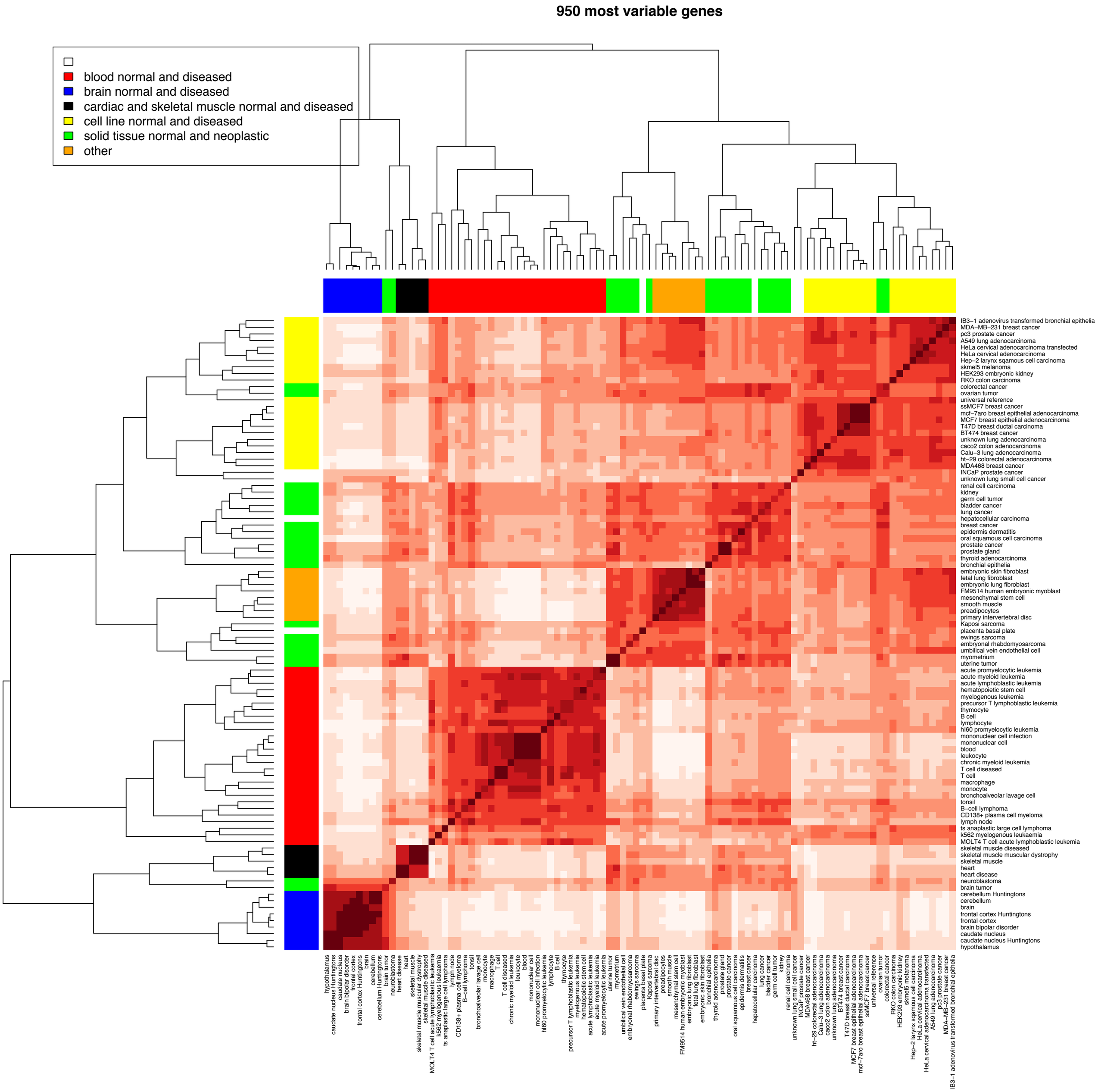
Supplementary Figure 4d:

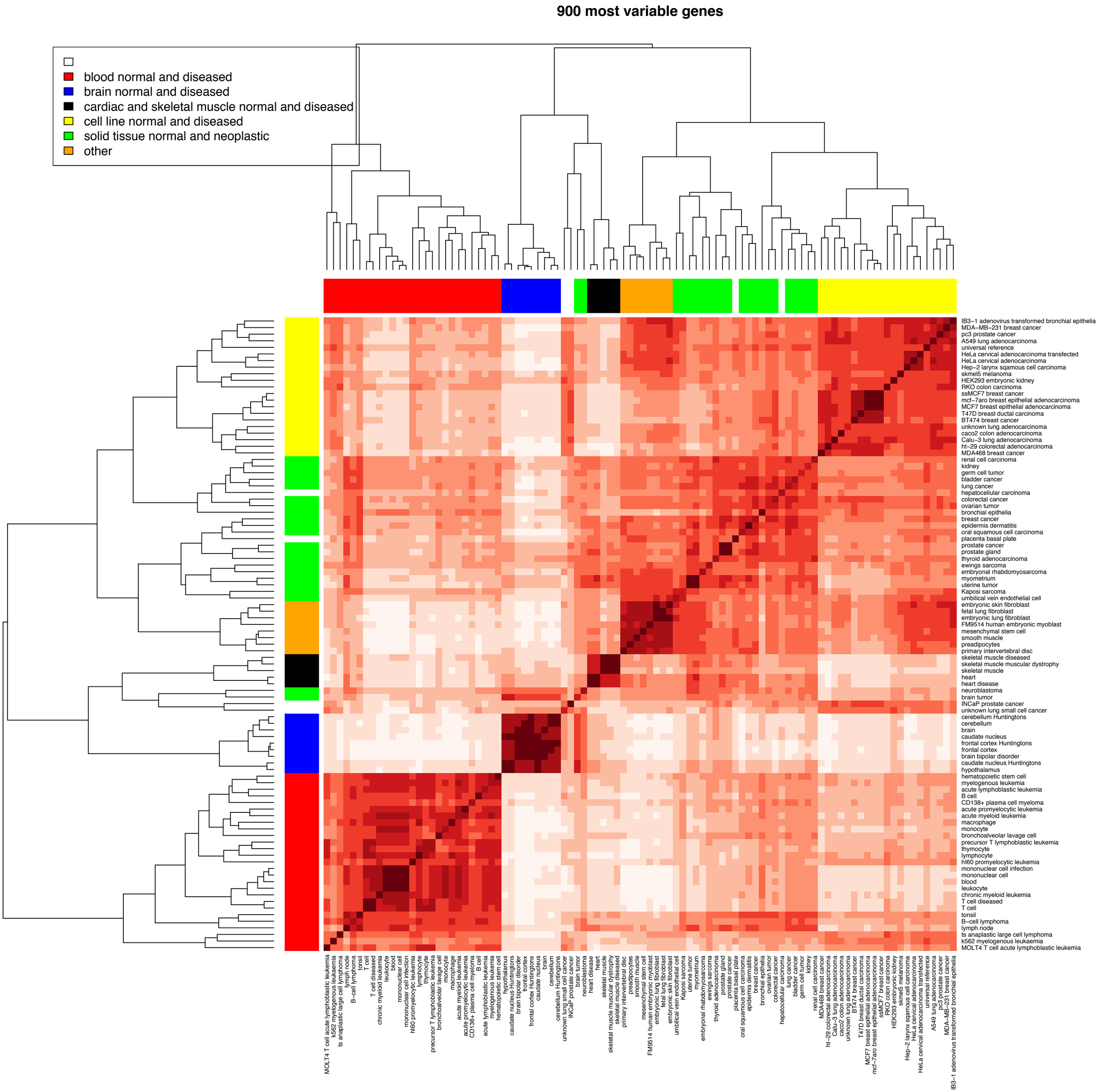
Heat map comparison of the 96 larger biological groups with the 273 smaller ones.

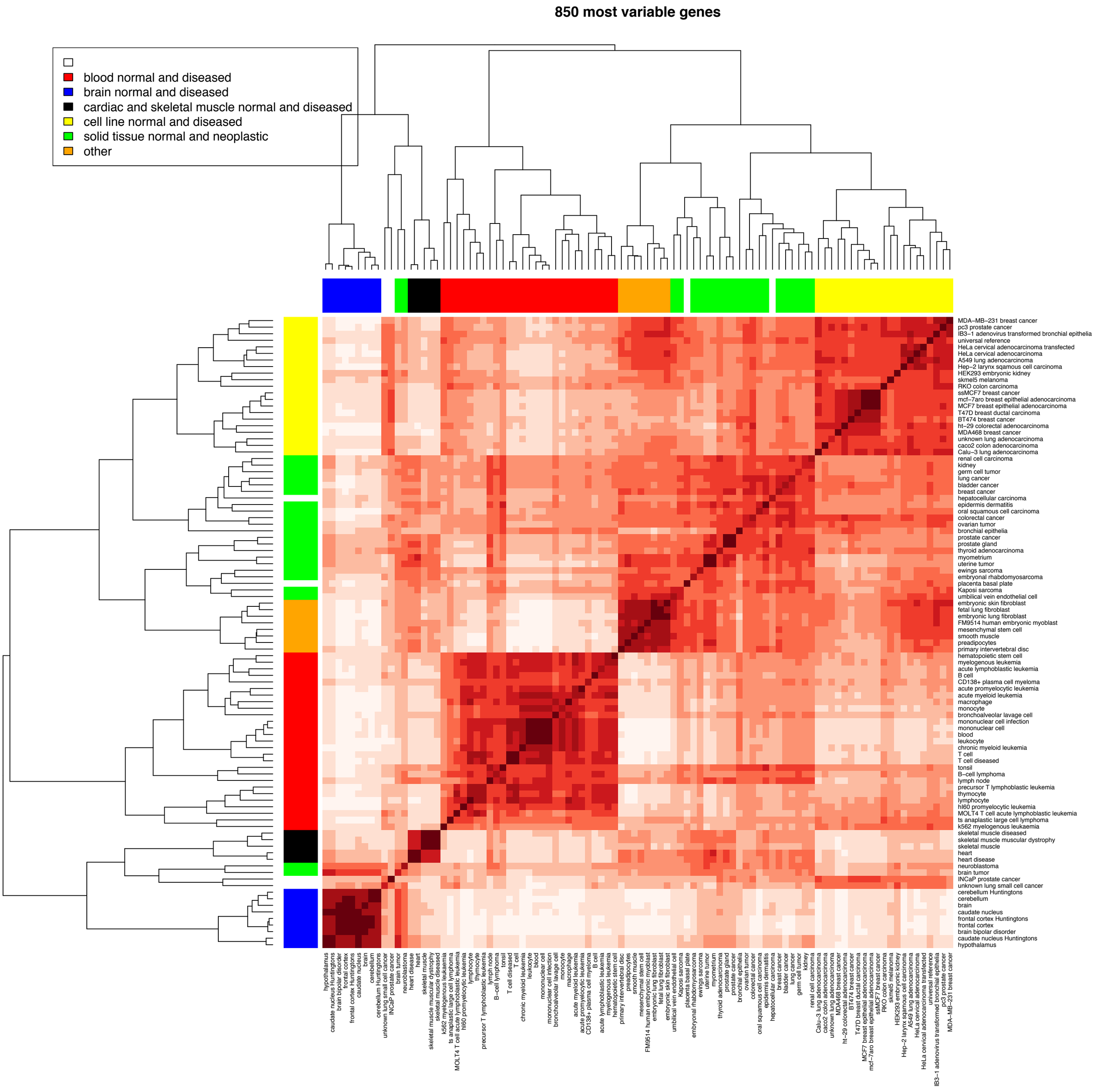
ALL genes

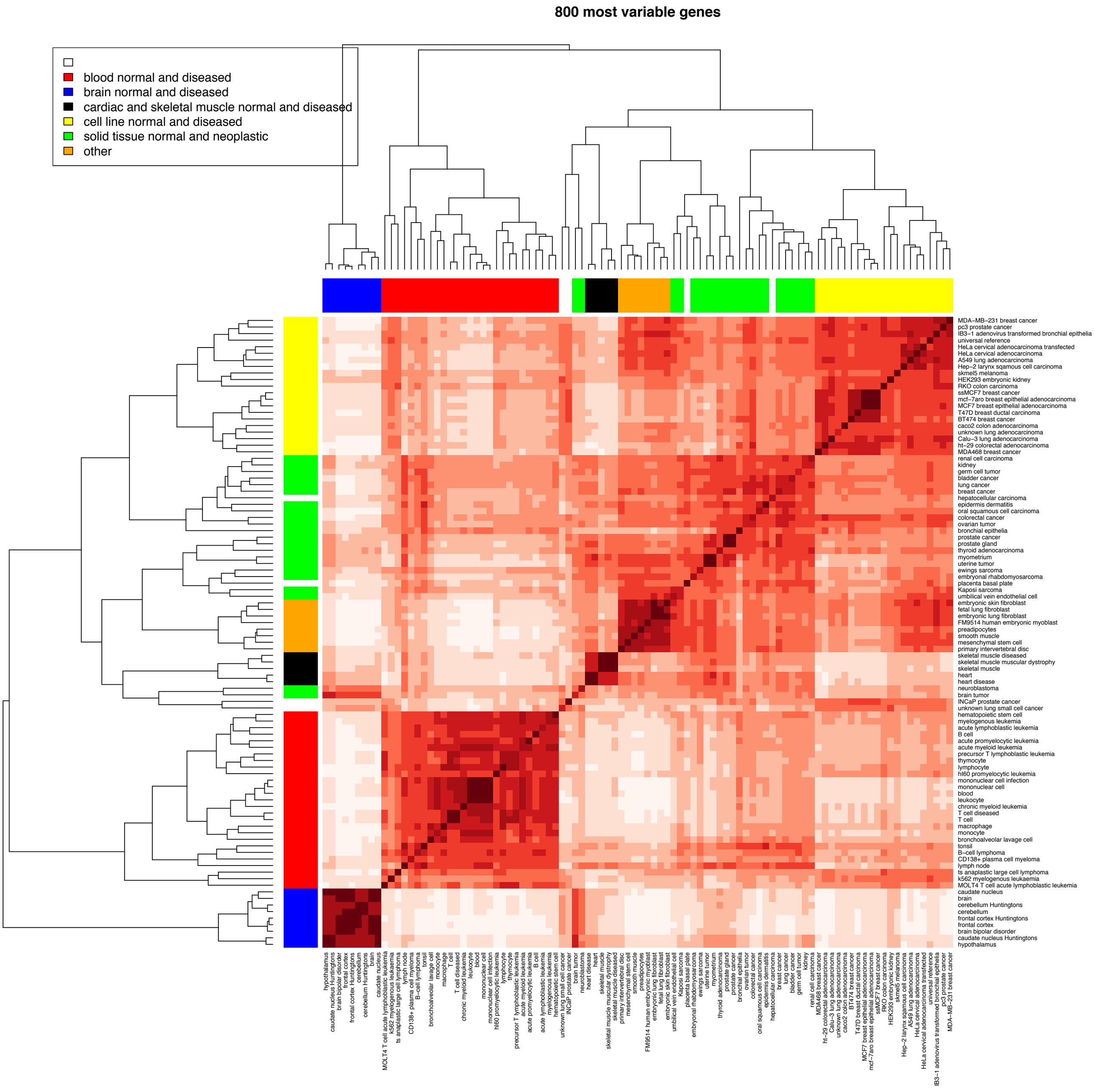

blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
ssMCF7 breast cancer
mcf−7aro breast epithelial adenocarcinoma
MCF7 breast epithelial adenocarcinoma
BT474 breast cancer
T47D breast ductal carcinoma
HEK293 embryonic kidney
RKO colon carcinoma
pc3 prostate cancer
A549 lung adenocarcinoma
MDA−MB−231 breast cancer
IB3−1 adenovirus transformed bronchial epithelia
universal reference
skmel5 melanoma
unknown lung adenocarcinoma
MDA468 breast cancer
caco2 colon adenocarcinoma
Calu−3 lung adenocarcinoma
ht−29 colorectal adenocarcinoma
HeLa cervical adenocarcinoma
HeLa cervical adenocarcinoma transfected
Hep−2 larynx sqamous cell carcinoma
fetal lung fibroblast
FM9514 human embryonic myoblast
embryonic skin fibroblast
embryonic lung fibroblast
preadipocytes
smooth muscle
mesenchymal stem cell
primary intervertebral disc
INCaP prostate cancer
unknown lung small cell cancer
breast cancer
bladder cancer
germ cell tumor
colorectal cancer
ovarian tumor
oral squamous cell carcinoma
lung cancer
epidermis dermatitis
ewings sarcoma
embryonal rhabdomyosarcoma
Kaposi sarcoma
umbilical vein endothelial cell
prostate cancer
prostate gland
renal cell carcinoma
kidney
thyroid adenocarcinoma
myometrium
uterine tumor
brain tumor
neuroblastoma
bronchial epithelia
acute promyelocytic leukemia
acute myeloid leukemia
macrophage
bronchoalveolar lavage cell
monocyte
T cell diseased
T cell
mononuclear cell infection
mononuclear cell
lymph node
tonsil
blood
chronic myeloid leukemia
leukocyte
myelogenous leukemia
hematopoietic stem cell
B−cell lymphoma
hl60 promyelocytic leukemia
lymphocyte
precursor T lymphoblastic leukemia
thymocyte
acute lymphoblastic leukemia
B cell
CD138+ plasma cell myeloma
ts anaplastic large cell lymphoma
k562 myelogenous leukaemia
MOLT4 T cell acute lymphoblastic leukemia
brain bipolar disorder
brain
frontal cortex Huntingtons
frontal cortex
cerebellum Huntingtons
cerebellum
caudate nucleus
hypothalamus
caudate nucleus Huntingtons
skeletal muscle muscular dystrophy
skeletal muscle diseased
skeletal muscle
heart
heart disease
placenta basal plate
hepatocellular carcinoma

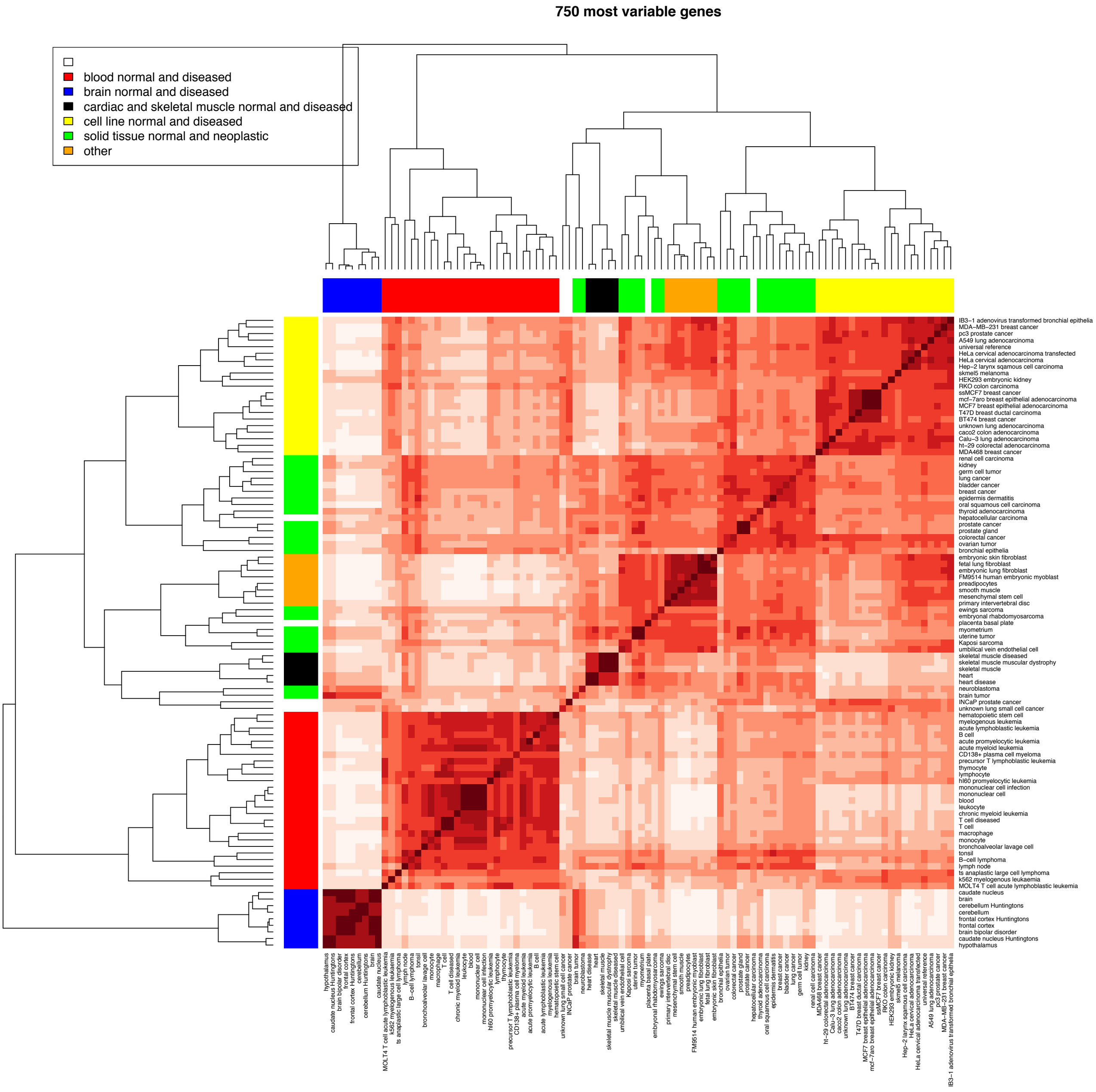

1000 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
IB3−1 adenovirus transformed bronchial epithelia
MDA−MB−231 breast cancer
A549 lung adenocarcinoma
pc3 prostate cancer
HeLa cervical adenocarcinoma transfected
HeLa cervical adenocarcinoma
Hep−2 larynx sqamous cell carcinoma
universal reference
skmel5 melanoma
HEK293 embryonic kidney
RKO colon carcinoma
ssMCF7 breast cancer
mcf−7aro breast epithelial adenocarcinoma
MCF7 breast epithelial adenocarcinoma
T47D breast ductal carcinoma
BT474 breast cancer
unknown lung adenocarcinoma
caco2 colon adenocarcinoma
Calu−3 lung adenocarcinoma
ht−29 colorectal adenocarcinoma
MDA468 breast cancer
lung cancer
breast cancer
epidermis dermatitis
oral squamous cell carcinoma
renal cell carcinoma
kidney
germ cell tumor
hepatocellular carcinoma
colorectal cancer
ovarian tumor
bladder cancer
bronchial epithelia
prostate cancer
prostate gland
thyroid adenocarcinoma
myometrium
uterine tumor
Kaposi sarcoma
placenta basal plate
ewings sarcoma
embryonal rhabdomyosarcoma
umbilical vein endothelial cell
embryonic skin fibroblast
fetal lung fibroblast
embryonic lung fibroblast
FM9514 human embryonic myoblast
mesenchymal stem cell
smooth muscle
preadipocytes
primary intervertebral disc
hematopoietic stem cell
myelogenous leukemia
acute lymphoblastic leukemia
B cell
CD138+ plasma cell myeloma
bronchoalveolar lavage cell
monocyte
acute promyelocytic leukemia
acute myeloid leukemia
macrophage
precursor T lymphoblastic leukemia
thymocyte
lymphocyte
hl60 promyelocytic leukemia
mononuclear cell infection
mononuclear cell
blood
leukocyte
chronic myeloid leukemia
T cell diseased
T cell
tonsil
B−cell lymphoma
lymph node
ts anaplastic large cell lymphoma
k562 myelogenous leukaemia
MOLT4 T cell acute lymphoblastic leukemia
skeletal muscle diseased
skeletal muscle muscular dystrophy
skeletal muscle
heart
heart disease
neuroblastoma
brain tumor
INCaP prostate cancer
unknown lung small cell cancer
cerebellum Huntingtons
cerebellum
brain
caudate nucleus
frontal cortex Huntingtons
frontal cortex
brain bipolar disorder
caudate nucleus Huntingtons
hypothalamus

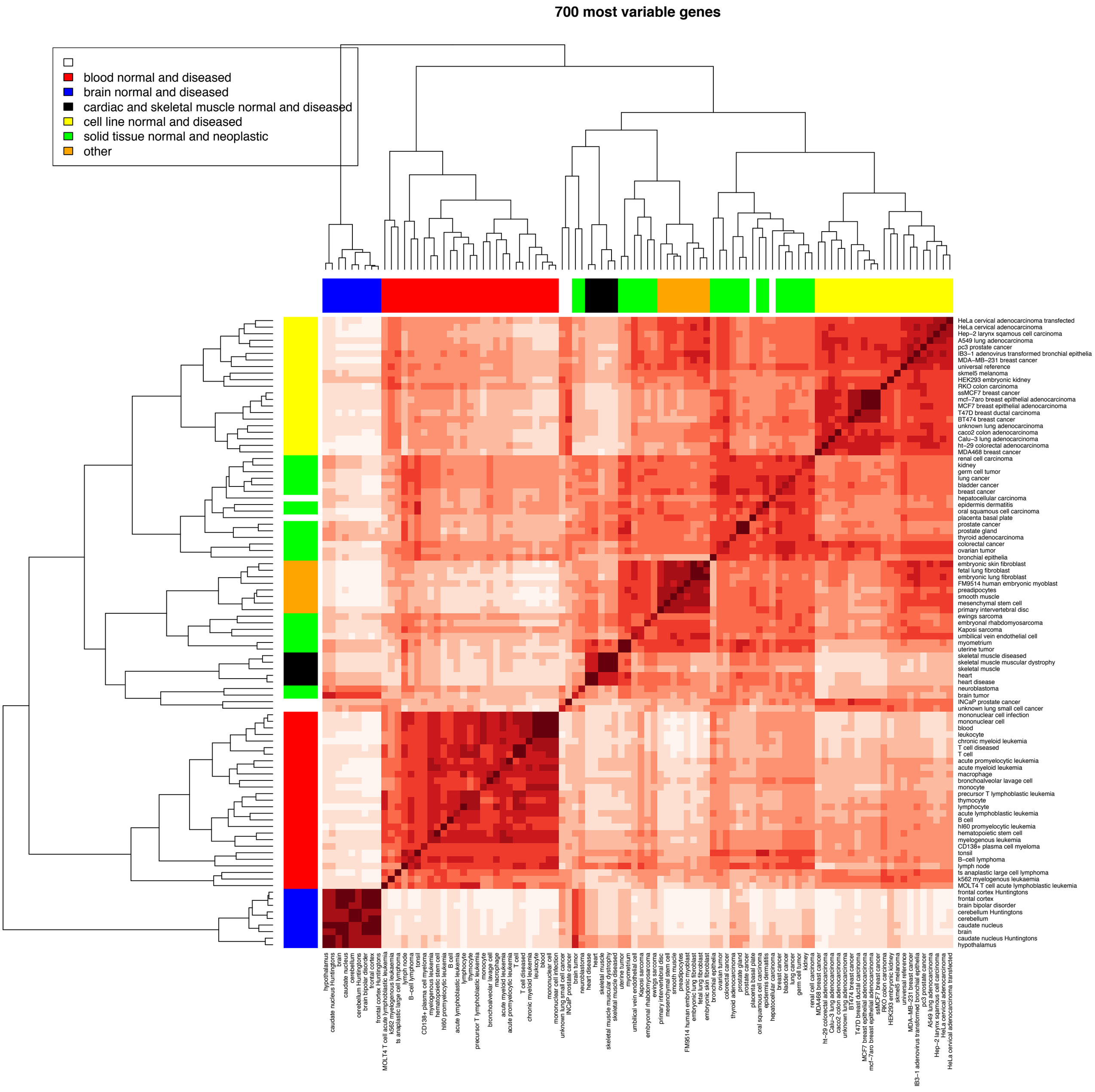

950 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
IB3−1 adenovirus transformed bronchial epithelia
MDA−MB−231 breast cancer
pc3 prostate cancer
A549 lung adenocarcinoma
HeLa cervical adenocarcinoma transfected
HeLa cervical adenocarcinoma
Hep−2 larynx sqamous cell carcinoma
skmel5 melanoma
HEK293 embryonic kidney
RKO colon carcinoma
colorectal cancer
ovarian tumor
universal reference
ssMCF7 breast cancer
mcf−7aro breast epithelial adenocarcinoma
MCF7 breast epithelial adenocarcinoma
T47D breast ductal carcinoma
BT474 breast cancer
unknown lung adenocarcinoma
caco2 colon adenocarcinoma
Calu−3 lung adenocarcinoma
ht−29 colorectal adenocarcinoma
MDA468 breast cancer
INCaP prostate cancer
unknown lung small cell cancer
renal cell carcinoma
kidney
germ cell tumor
bladder cancer
lung cancer
hepatocellular carcinoma
breast cancer
epidermis dermatitis
oral squamous cell carcinoma
prostate cancer
prostate gland
thyroid adenocarcinoma
bronchial epithelia
embryonic skin fibroblast
fetal lung fibroblast
embryonic lung fibroblast
FM9514 human embryonic myoblast
mesenchymal stem cell
smooth muscle
preadipocytes
primary intervertebral disc
Kaposi sarcoma
placenta basal plate
ewings sarcoma
embryonal rhabdomyosarcoma
umbilical vein endothelial cell
myometrium
uterine tumor
acute promyelocytic leukemia
acute myeloid leukemia
acute lymphoblastic leukemia
hematopoietic stem cell
myelogenous leukemia
precursor T lymphoblastic leukemia
thymocyte
B cell
lymphocyte
hl60 promyelocytic leukemia
mononuclear cell infection
mononuclear cell
blood
leukocyte
chronic myeloid leukemia
T cell diseased
T cell
macrophage
monocyte
bronchoalveolar lavage cell
tonsil
B−cell lymphoma
CD138+ plasma cell myeloma
lymph node
ts anaplastic large cell lymphoma
k562 myelogenous leukaemia
MOLT4 T cell acute lymphoblastic leukemia
skeletal muscle diseased
skeletal muscle muscular dystrophy
skeletal muscle
heart
heart disease
neuroblastoma
brain tumor
cerebellum Huntingtons
cerebellum
brain
frontal cortex Huntingtons
frontal cortex
brain bipolar disorder
caudate nucleus
caudate nucleus Huntingtons
hypothalamus

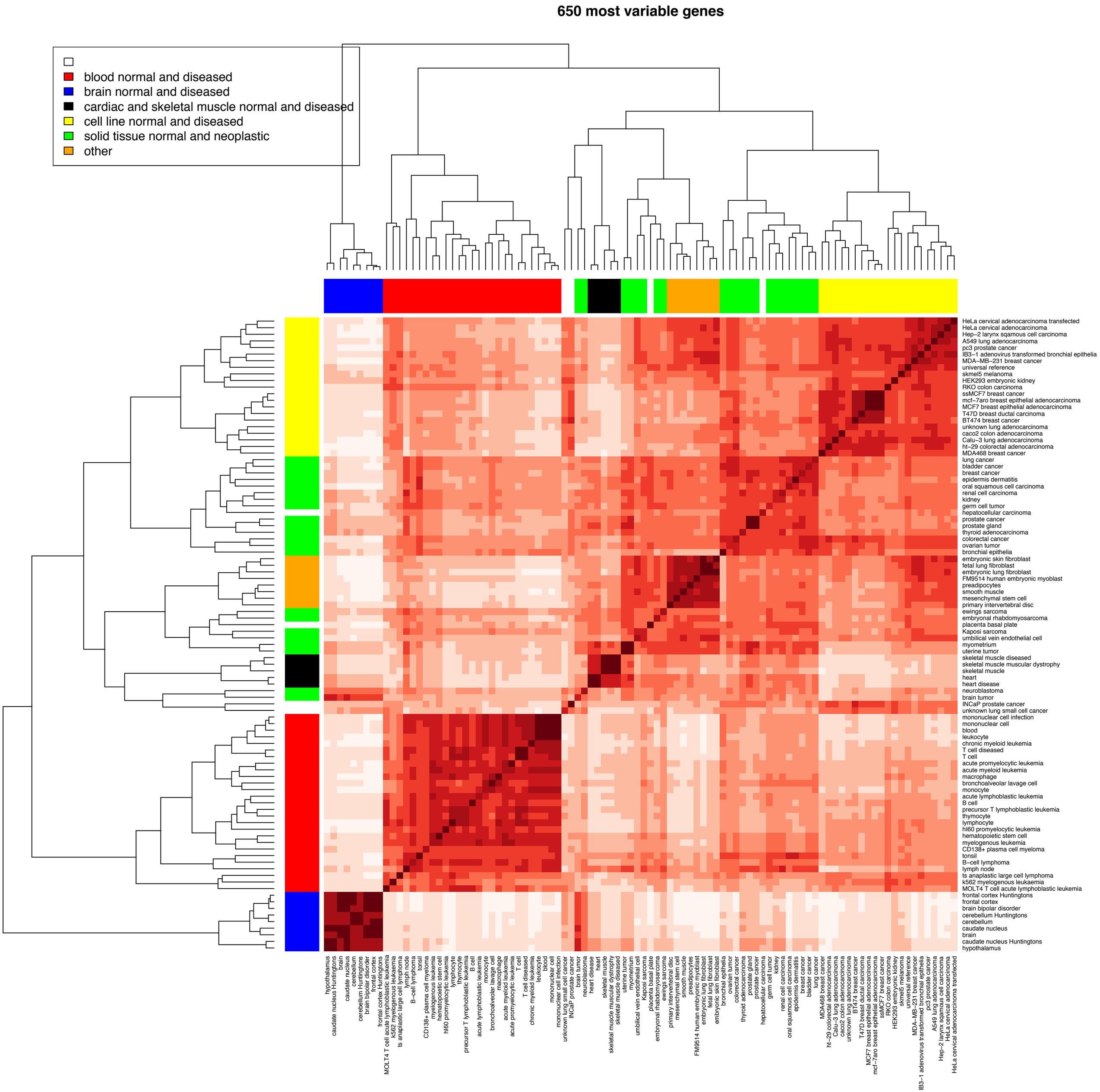

900 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
IB3−1 adenovirus transformed bronchial epithelia
MDA−MB−231 breast cancer
pc3 prostate cancer
A549 lung adenocarcinoma
universal reference
HeLa cervical adenocarcinoma transfected
HeLa cervical adenocarcinoma
Hep−2 larynx sqamous cell carcinoma
skmel5 melanoma
HEK293 embryonic kidney
RKO colon carcinoma
ssMCF7 breast cancer
mcf−7aro breast epithelial adenocarcinoma
MCF7 breast epithelial adenocarcinoma
T47D breast ductal carcinoma
BT474 breast cancer
unknown lung adenocarcinoma
caco2 colon adenocarcinoma
Calu−3 lung adenocarcinoma
ht−29 colorectal adenocarcinoma
MDA468 breast cancer
renal cell carcinoma
kidney
germ cell tumor
bladder cancer
lung cancer
hepatocellular carcinoma
colorectal cancer
ovarian tumor
bronchial epithelia
breast cancer
epidermis dermatitis
oral squamous cell carcinoma
placenta basal plate
prostate cancer
prostate gland
thyroid adenocarcinoma
ewings sarcoma
embryonal rhabdomyosarcoma
myometrium
uterine tumor
Kaposi sarcoma
umbilical vein endothelial cell
embryonic skin fibroblast
fetal lung fibroblast
embryonic lung fibroblast
FM9514 human embryonic myoblast
mesenchymal stem cell
smooth muscle
preadipocytes
primary intervertebral disc
skeletal muscle diseased
skeletal muscle muscular dystrophy
skeletal muscle
heart
heart disease
neuroblastoma
brain tumor
INCaP prostate cancer
unknown lung small cell cancer
cerebellum Huntingtons
cerebellum
brain
caudate nucleus
frontal cortex Huntingtons
frontal cortex
brain bipolar disorder
caudate nucleus Huntingtons
hypothalamus
hematopoietic stem cell
myelogenous leukemia
acute lymphoblastic leukemia
B cell
CD138+ plasma cell myeloma
acute promyelocytic leukemia
acute myeloid leukemia
macrophage
monocyte
bronchoalveolar lavage cell
precursor T lymphoblastic leukemia
thymocyte
lymphocyte
hl60 promyelocytic leukemia
mononuclear cell infection
mononuclear cell
blood
leukocyte
chronic myeloid leukemia
T cell diseased
T cell
tonsil
B−cell lymphoma
lymph node
ts anaplastic large cell lymphoma
k562 myelogenous leukaemia
MOLT4 T cell acute lymphoblastic leukemia

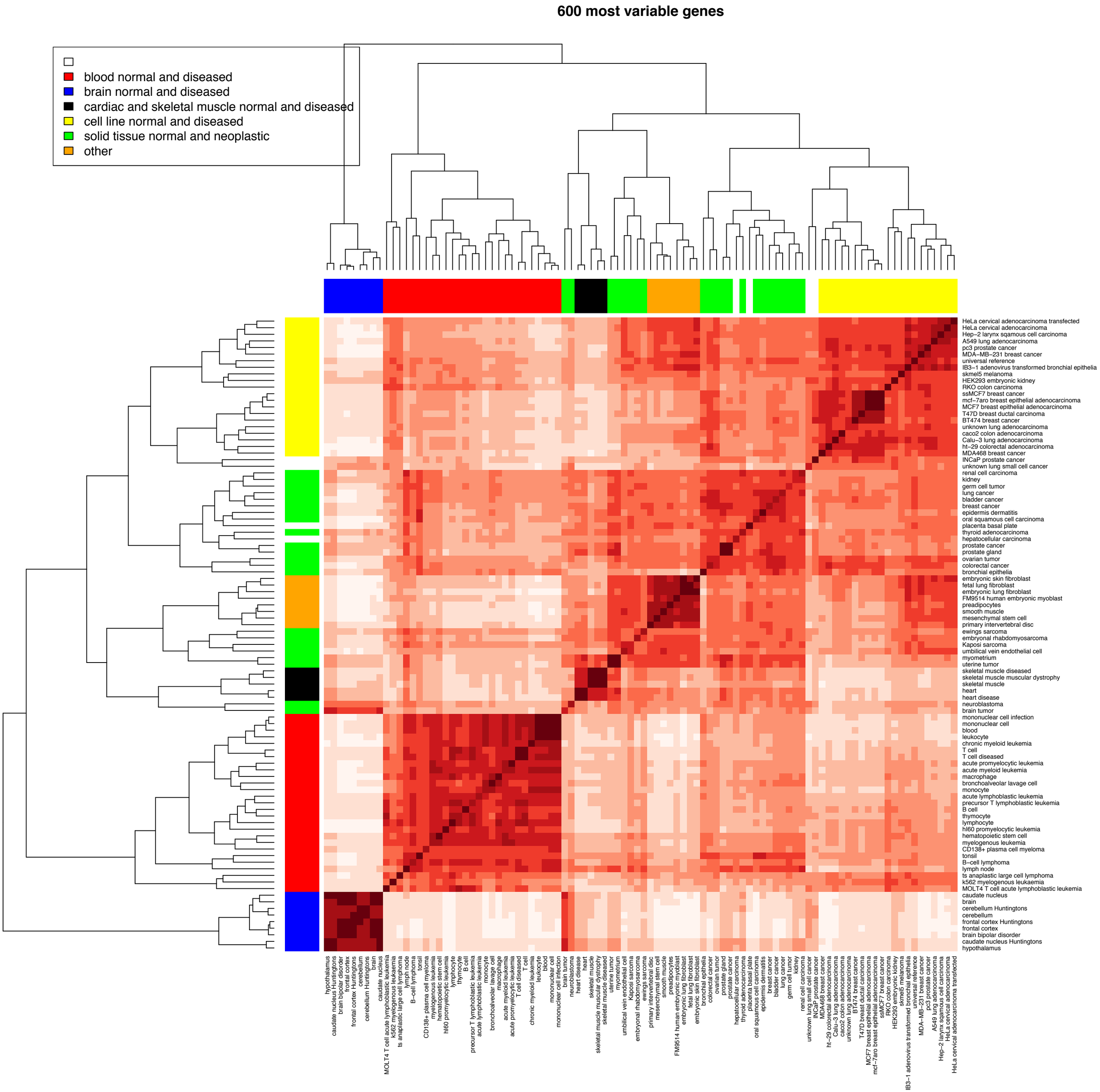

850 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
MDA−MB−231 breast cancer
pc3 prostate cancer
IB3−1 adenovirus transformed bronchial epithelia
universal reference
HeLa cervical adenocarcinoma transfected
HeLa cervical adenocarcinoma
A549 lung adenocarcinoma
Hep−2 larynx sqamous cell carcinoma
HEK293 embryonic kidney
skmel5 melanoma
RKO colon carcinoma
ssMCF7 breast cancer
mcf−7aro breast epithelial adenocarcinoma
MCF7 breast epithelial adenocarcinoma
T47D breast ductal carcinoma
BT474 breast cancer
ht−29 colorectal adenocarcinoma
MDA468 breast cancer
unknown lung adenocarcinoma
caco2 colon adenocarcinoma
Calu−3 lung adenocarcinoma
renal cell carcinoma
kidney
germ cell tumor
lung cancer
bladder cancer
breast cancer
hepatocellular carcinoma
epidermis dermatitis
oral squamous cell carcinoma
colorectal cancer
ovarian tumor
bronchial epithelia
prostate cancer
prostate gland
thyroid adenocarcinoma
myometrium
uterine tumor
ewings sarcoma
embryonal rhabdomyosarcoma
placenta basal plate
Kaposi sarcoma
umbilical vein endothelial cell
embryonic skin fibroblast
fetal lung fibroblast
embryonic lung fibroblast
FM9514 human embryonic myoblast
mesenchymal stem cell
smooth muscle
preadipocytes
primary intervertebral disc
hematopoietic stem cell
myelogenous leukemia
acute lymphoblastic leukemia
B cell
CD138+ plasma cell myeloma
acute promyelocytic leukemia
acute myeloid leukemia
macrophage
monocyte
bronchoalveolar lavage cell
mononuclear cell infection
mononuclear cell
blood
leukocyte
chronic myeloid leukemia
T cell
T cell diseased
tonsil
B−cell lymphoma
lymph node
precursor T lymphoblastic leukemia
thymocyte
lymphocyte
hl60 promyelocytic leukemia
MOLT4 T cell acute lymphoblastic leukemia
ts anaplastic large cell lymphoma
k562 myelogenous leukaemia
skeletal muscle diseased
skeletal muscle muscular dystrophy
skeletal muscle
heart
heart disease
neuroblastoma
brain tumor
INCaP prostate cancer
unknown lung small cell cancer
cerebellum Huntingtons
cerebellum
brain
caudate nucleus
frontal cortex Huntingtons
frontal cortex
brain bipolar disorder
caudate nucleus Huntingtons
hypothalamus

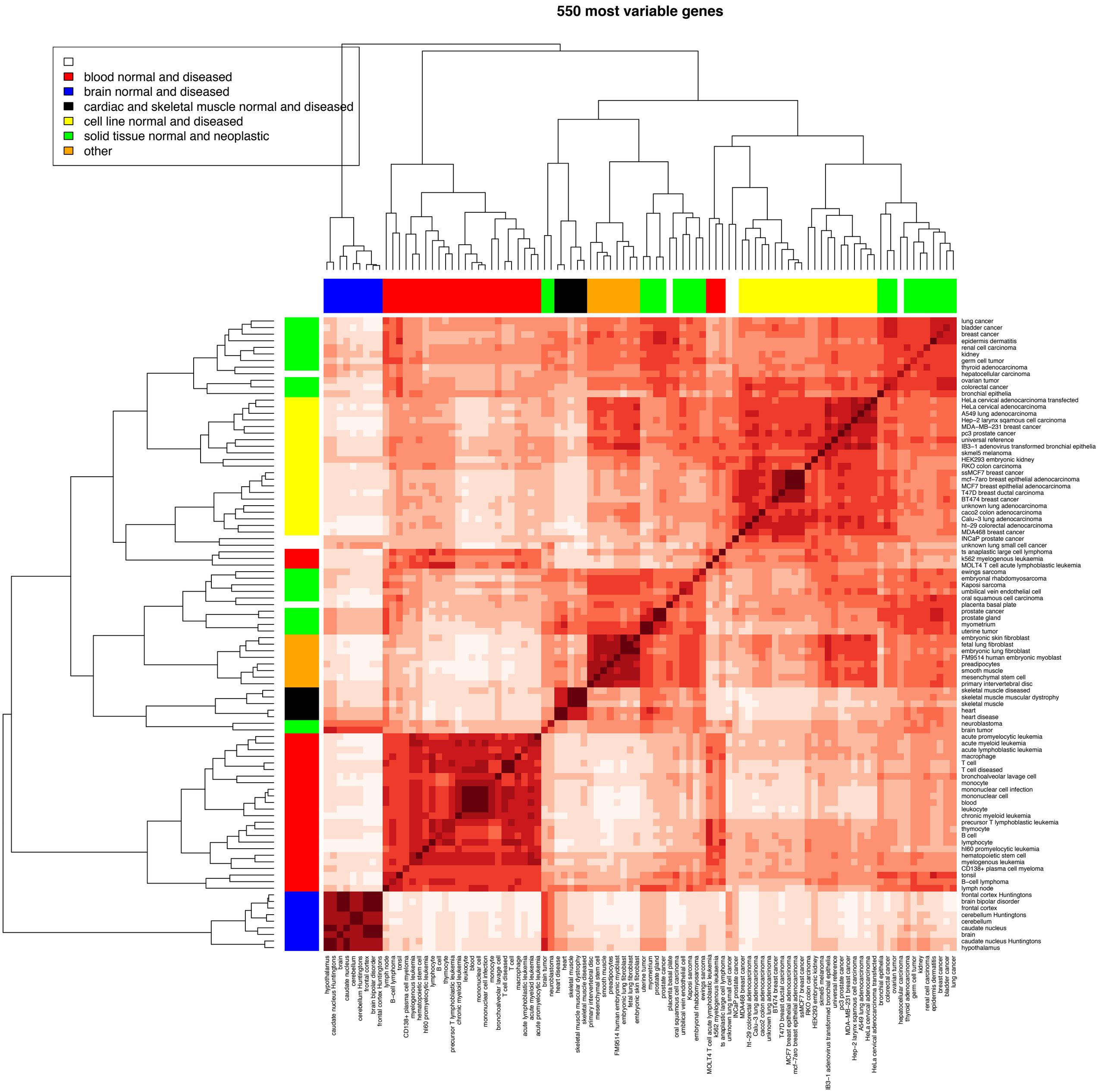

800 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
MDA−MB−231 breast cancer
pc3 prostate cancer
IB3−1 adenovirus transformed bronchial epithelia
universal reference
HeLa cervical adenocarcinoma transfected
HeLa cervical adenocarcinoma
A549 lung adenocarcinoma
Hep−2 larynx sqamous cell carcinoma
skmel5 melanoma
HEK293 embryonic kidney
RKO colon carcinoma
ssMCF7 breast cancer
mcf−7aro breast epithelial adenocarcinoma
MCF7 breast epithelial adenocarcinoma
T47D breast ductal carcinoma
BT474 breast cancer
caco2 colon adenocarcinoma
unknown lung adenocarcinoma
Calu−3 lung adenocarcinoma
ht−29 colorectal adenocarcinoma
MDA468 breast cancer
renal cell carcinoma
kidney
germ cell tumor
bladder cancer
lung cancer
breast cancer
hepatocellular carcinoma
epidermis dermatitis
oral squamous cell carcinoma
colorectal cancer
ovarian tumor
bronchial epithelia
prostate cancer
prostate gland
thyroid adenocarcinoma
myometrium
uterine tumor
ewings sarcoma
embryonal rhabdomyosarcoma
placenta basal plate
Kaposi sarcoma
umbilical vein endothelial cell
embryonic skin fibroblast
fetal lung fibroblast
embryonic lung fibroblast
FM9514 human embryonic myoblast
preadipocytes
smooth muscle
mesenchymal stem cell
primary intervertebral disc
skeletal muscle diseased
skeletal muscle muscular dystrophy
skeletal muscle
heart
heart disease
neuroblastoma
brain tumor
INCaP prostate cancer
unknown lung small cell cancer
hematopoietic stem cell
myelogenous leukemia
acute lymphoblastic leukemia
B cell
acute promyelocytic leukemia
acute myeloid leukemia
precursor T lymphoblastic leukemia
thymocyte
lymphocyte
hl60 promyelocytic leukemia
mononuclear cell infection
mononuclear cell
blood
leukocyte
chronic myeloid leukemia
T cell diseased
T cell
macrophage
monocyte
bronchoalveolar lavage cell
tonsil
B−cell lymphoma
CD138+ plasma cell myeloma
lymph node
ts anaplastic large cell lymphoma
k562 myelogenous leukaemia
MOLT4 T cell acute lymphoblastic leukemia
caudate nucleus
brain
cerebellum Huntingtons
cerebellum
frontal cortex Huntingtons
frontal cortex
brain bipolar disorder
caudate nucleus Huntingtons
hypothalamus

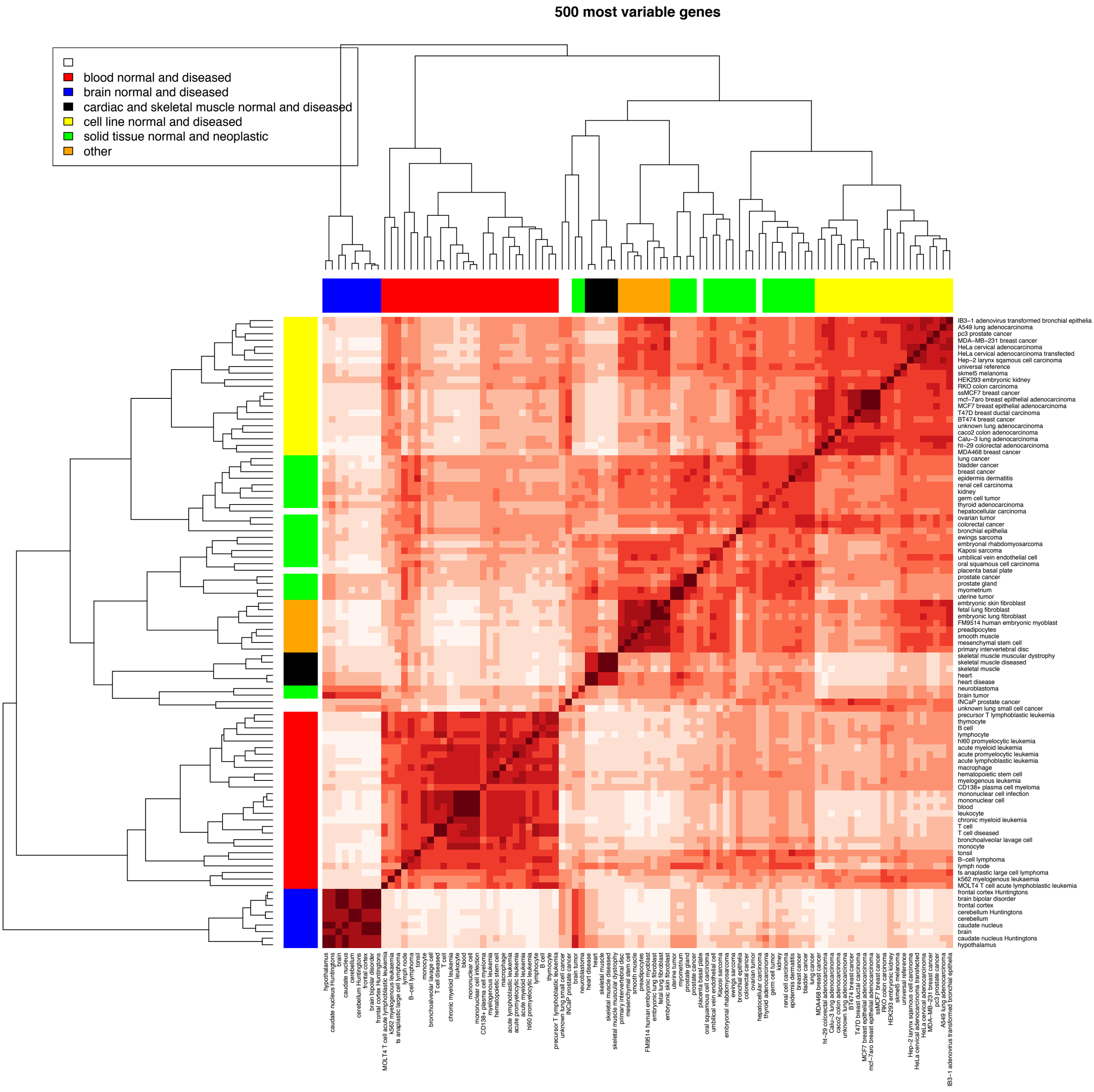

750 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
IB3−1 adenovirus transformed bronchial epithelia
MDA−MB−231 breast cancer
pc3 prostate cancer
A549 lung adenocarcinoma
universal reference
HeLa cervical adenocarcinoma transfected
HeLa cervical adenocarcinoma
Hep−2 larynx sqamous cell carcinoma
skmel5 melanoma
HEK293 embryonic kidney
RKO colon carcinoma
ssMCF7 breast cancer
mcf−7aro breast epithelial adenocarcinoma
MCF7 breast epithelial adenocarcinoma
T47D breast ductal carcinoma
BT474 breast cancer
unknown lung adenocarcinoma
caco2 colon adenocarcinoma
Calu−3 lung adenocarcinoma
ht−29 colorectal adenocarcinoma
MDA468 breast cancer
renal cell carcinoma
kidney
germ cell tumor
lung cancer
bladder cancer
breast cancer
epidermis dermatitis
oral squamous cell carcinoma
thyroid adenocarcinoma
hepatocellular carcinoma
prostate cancer
prostate gland
colorectal cancer
ovarian tumor
bronchial epithelia
embryonic skin fibroblast
fetal lung fibroblast
embryonic lung fibroblast
FM9514 human embryonic myoblast
preadipocytes
smooth muscle
mesenchymal stem cell
primary intervertebral disc
ewings sarcoma
embryonal rhabdomyosarcoma
placenta basal plate
myometrium
uterine tumor
Kaposi sarcoma
umbilical vein endothelial cell
skeletal muscle diseased
skeletal muscle muscular dystrophy
skeletal muscle
heart
heart disease
neuroblastoma
brain tumor
INCaP prostate cancer
unknown lung small cell cancer
hematopoietic stem cell
myelogenous leukemia
acute lymphoblastic leukemia
B cell
acute promyelocytic leukemia
acute myeloid leukemia
CD138+ plasma cell myeloma
precursor T lymphoblastic leukemia
thymocyte
lymphocyte
hl60 promyelocytic leukemia
mononuclear cell infection
mononuclear cell
blood
leukocyte
chronic myeloid leukemia
T cell diseased
T cell
macrophage
monocyte
bronchoalveolar lavage cell
tonsil
B−cell lymphoma
lymph node
ts anaplastic large cell lymphoma
k562 myelogenous leukaemia
MOLT4 T cell acute lymphoblastic leukemia
caudate nucleus
brain
cerebellum Huntingtons
cerebellum
frontal cortex Huntingtons
frontal cortex
brain bipolar disorder
caudate nucleus Huntingtons
hypothalamus

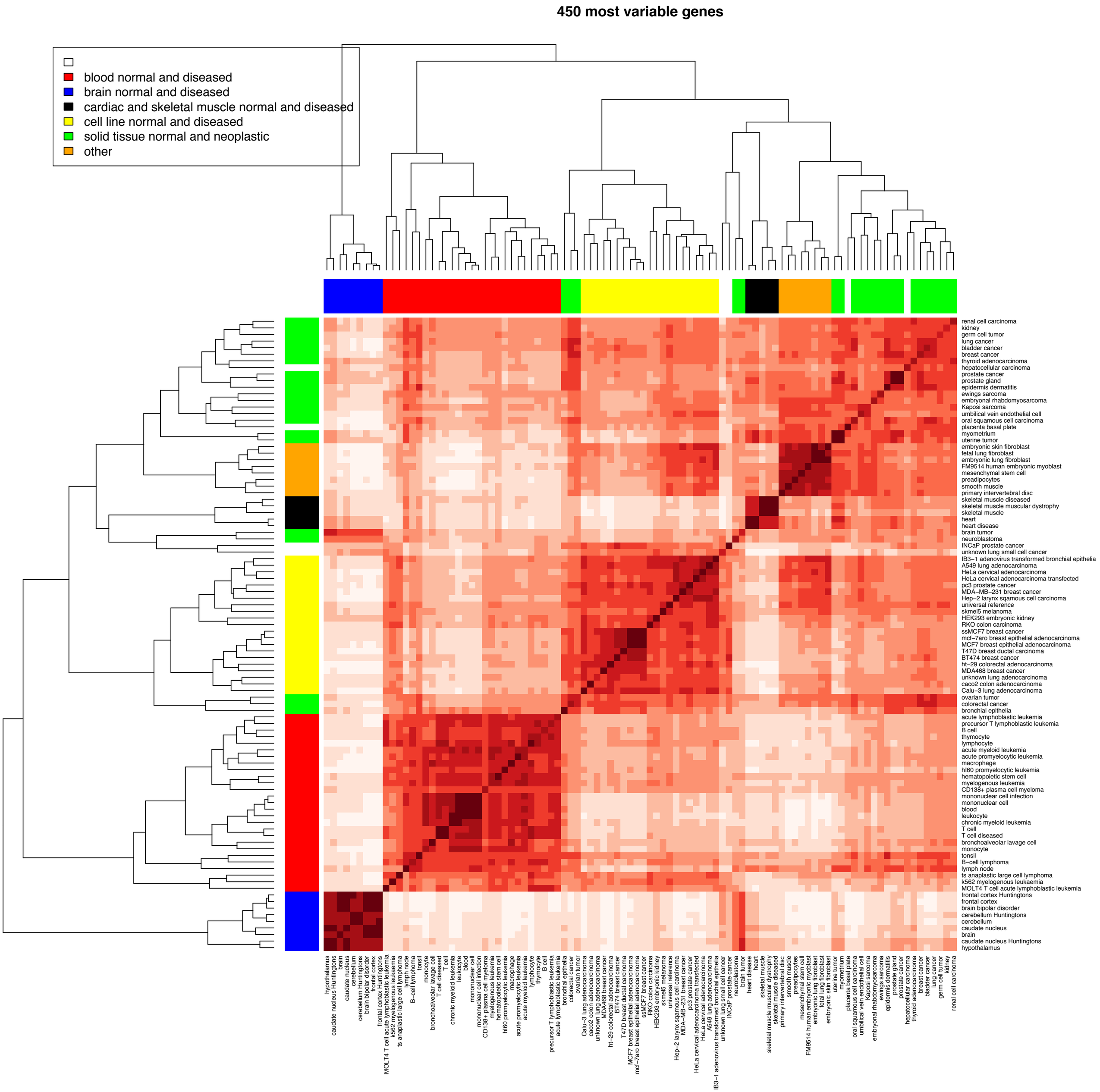

700 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
HeLa cervical adenocarcinoma transfected
HeLa cervical adenocarcinoma
Hep−2 larynx sqamous cell carcinoma
A549 lung adenocarcinoma
pc3 prostate cancer
IB3−1 adenovirus transformed bronchial epithelia
MDA−MB−231 breast cancer
universal reference
skmel5 melanoma
HEK293 embryonic kidney
RKO colon carcinoma
ssMCF7 breast cancer
mcf−7aro breast epithelial adenocarcinoma
MCF7 breast epithelial adenocarcinoma
T47D breast ductal carcinoma
BT474 breast cancer
unknown lung adenocarcinoma
caco2 colon adenocarcinoma
Calu−3 lung adenocarcinoma
ht−29 colorectal adenocarcinoma
MDA468 breast cancer
renal cell carcinoma
kidney
germ cell tumor
lung cancer
bladder cancer
breast cancer
hepatocellular carcinoma
epidermis dermatitis
oral squamous cell carcinoma
placenta basal plate
prostate cancer
prostate gland
thyroid adenocarcinoma
colorectal cancer
ovarian tumor
bronchial epithelia
embryonic skin fibroblast
fetal lung fibroblast
embryonic lung fibroblast
FM9514 human embryonic myoblast
preadipocytes
smooth muscle
mesenchymal stem cell
primary intervertebral disc
ewings sarcoma
embryonal rhabdomyosarcoma
Kaposi sarcoma
umbilical vein endothelial cell
myometrium
uterine tumor
skeletal muscle diseased
skeletal muscle muscular dystrophy
skeletal muscle
heart
heart disease
neuroblastoma
brain tumor
INCaP prostate cancer
unknown lung small cell cancer
mononuclear cell infection
mononuclear cell
blood
leukocyte
chronic myeloid leukemia
T cell diseased
T cell
acute promyelocytic leukemia
acute myeloid leukemia
macrophage
bronchoalveolar lavage cell
monocyte
precursor T lymphoblastic leukemia
thymocyte
lymphocyte
acute lymphoblastic leukemia
B cell
hl60 promyelocytic leukemia
hematopoietic stem cell
myelogenous leukemia
CD138+ plasma cell myeloma
tonsil
B−cell lymphoma
lymph node
ts anaplastic large cell lymphoma
k562 myelogenous leukaemia
MOLT4 T cell acute lymphoblastic leukemia
frontal cortex Huntingtons
frontal cortex
brain bipolar disorder
cerebellum Huntingtons
cerebellum
caudate nucleus
brain
caudate nucleus Huntingtons
hypothalamus

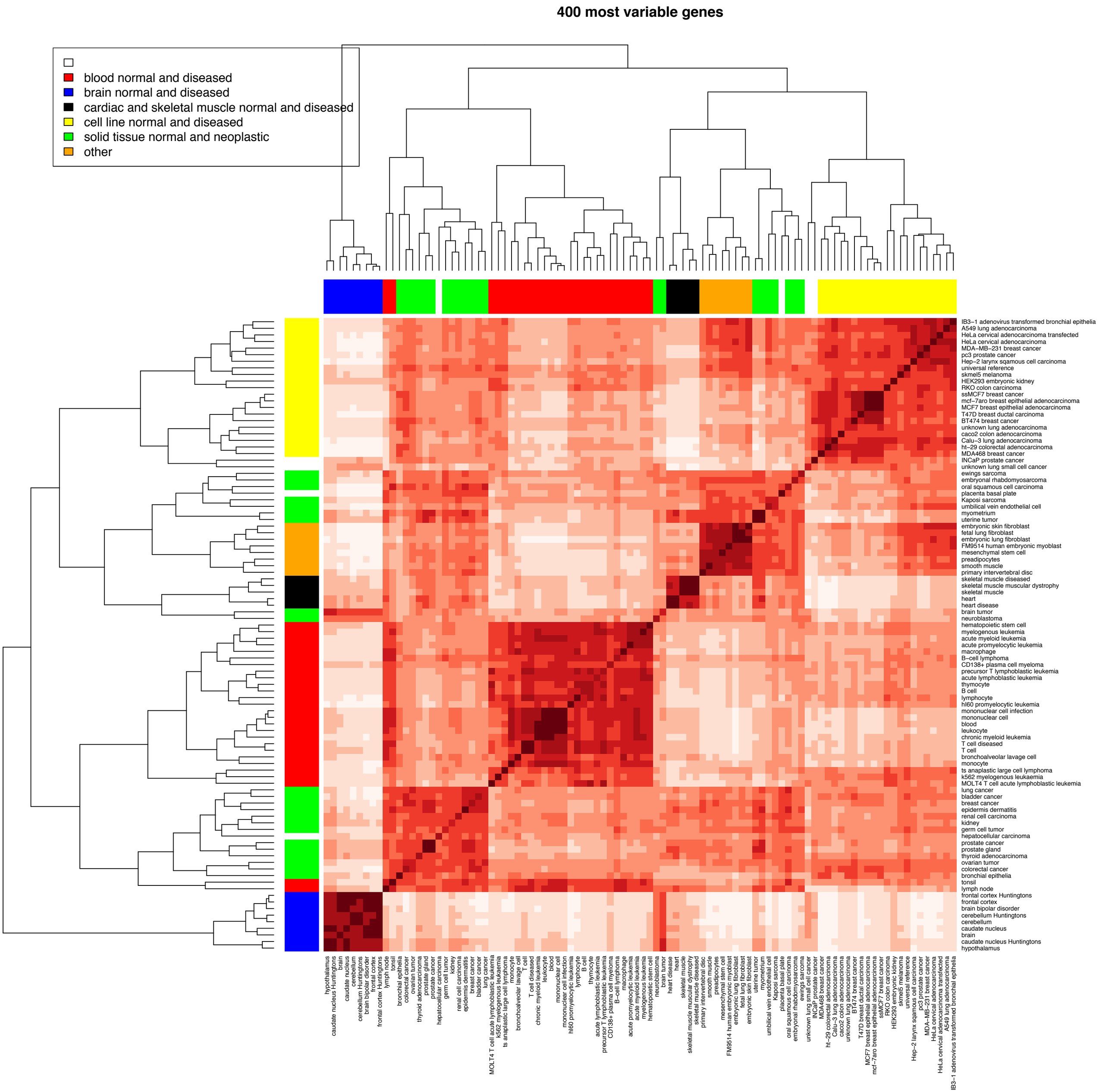

650 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
HeLa cervical adenocarcinoma transfected
HeLa cervical adenocarcinoma
Hep−2 larynx sqamous cell carcinoma
A549 lung adenocarcinoma
pc3 prostate cancer
IB3−1 adenovirus transformed bronchial epithelia
MDA−MB−231 breast cancer
universal reference
skmel5 melanoma
HEK293 embryonic kidney
RKO colon carcinoma
ssMCF7 breast cancer
mcf−7aro breast epithelial adenocarcinoma
MCF7 breast epithelial adenocarcinoma
T47D breast ductal carcinoma
BT474 breast cancer
unknown lung adenocarcinoma
caco2 colon adenocarcinoma
Calu−3 lung adenocarcinoma
ht−29 colorectal adenocarcinoma
MDA468 breast cancer
lung cancer
bladder cancer
breast cancer
epidermis dermatitis
oral squamous cell carcinoma
renal cell carcinoma
kidney
germ cell tumor
hepatocellular carcinoma
prostate cancer
prostate gland
thyroid adenocarcinoma
colorectal cancer
ovarian tumor
bronchial epithelia
embryonic skin fibroblast
fetal lung fibroblast
embryonic lung fibroblast
FM9514 human embryonic myoblast
preadipocytes
smooth muscle
mesenchymal stem cell
primary intervertebral disc
ewings sarcoma
embryonal rhabdomyosarcoma
placenta basal plate
Kaposi sarcoma
umbilical vein endothelial cell
myometrium
uterine tumor
skeletal muscle diseased
skeletal muscle muscular dystrophy
skeletal muscle
heart
heart disease
neuroblastoma
brain tumor
INCaP prostate cancer
unknown lung small cell cancer
mononuclear cell infection
mononuclear cell
blood
leukocyte
chronic myeloid leukemia
T cell diseased
T cell
acute promyelocytic leukemia
acute myeloid leukemia
macrophage
bronchoalveolar lavage cell
monocyte
acute lymphoblastic leukemia
B cell
precursor T lymphoblastic leukemia
thymocyte
lymphocyte
hl60 promyelocytic leukemia
hematopoietic stem cell
myelogenous leukemia
CD138+ plasma cell myeloma
tonsil
B−cell lymphoma
lymph node
ts anaplastic large cell lymphoma
k562 myelogenous leukaemia
MOLT4 T cell acute lymphoblastic leukemia
frontal cortex Huntingtons
frontal cortex
brain bipolar disorder
cerebellum Huntingtons
cerebellum
caudate nucleus
brain
caudate nucleus Huntingtons
hypothalamus

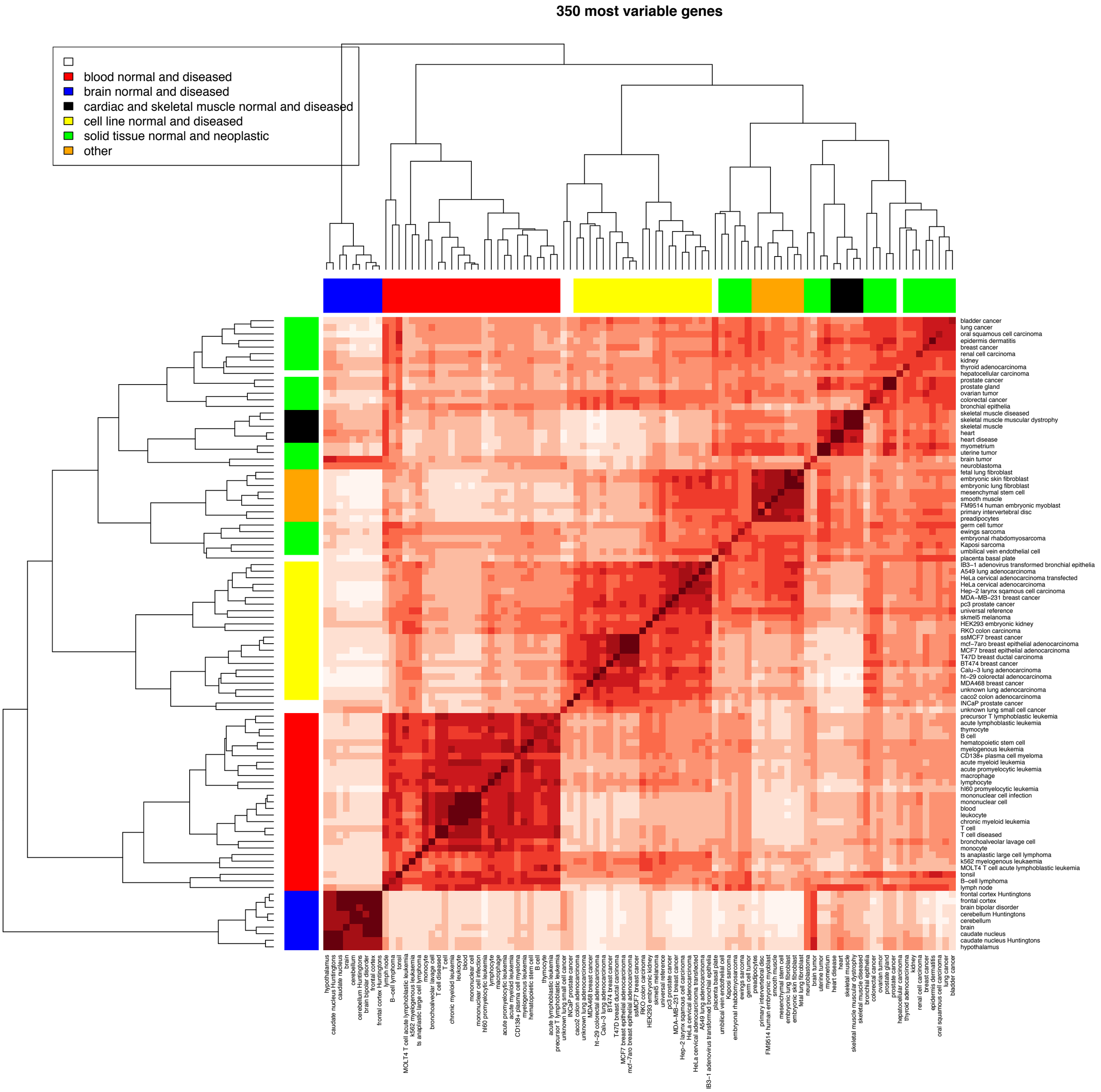

600 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
HeLa cervical adenocarcinoma transfected
HeLa cervical adenocarcinoma
Hep−2 larynx sqamous cell carcinoma
A549 lung adenocarcinoma
pc3 prostate cancer
MDA−MB−231 breast cancer
universal reference
IB3−1 adenovirus transformed bronchial epithelia
skmel5 melanoma
HEK293 embryonic kidney
RKO colon carcinoma
ssMCF7 breast cancer
mcf−7aro breast epithelial adenocarcinoma
MCF7 breast epithelial adenocarcinoma
T47D breast ductal carcinoma
BT474 breast cancer
unknown lung adenocarcinoma
caco2 colon adenocarcinoma
Calu−3 lung adenocarcinoma
ht−29 colorectal adenocarcinoma
MDA468 breast cancer
INCaP prostate cancer
unknown lung small cell cancer
renal cell carcinoma
kidney
germ cell tumor
lung cancer
bladder cancer
breast cancer
epidermis dermatitis
oral squamous cell carcinoma
placenta basal plate
thyroid adenocarcinoma
hepatocellular carcinoma
prostate cancer
prostate gland
ovarian tumor
colorectal cancer
bronchial epithelia
embryonic skin fibroblast
fetal lung fibroblast
embryonic lung fibroblast
FM9514 human embryonic myoblast
preadipocytes
smooth muscle
mesenchymal stem cell
primary intervertebral disc
ewings sarcoma
embryonal rhabdomyosarcoma
Kaposi sarcoma
umbilical vein endothelial cell
myometrium
uterine tumor
skeletal muscle diseased
skeletal muscle muscular dystrophy
skeletal muscle
heart
heart disease
neuroblastoma
brain tumor
mononuclear cell infection
mononuclear cell
blood
leukocyte
chronic myeloid leukemia
T cell
T cell diseased
acute promyelocytic leukemia
acute myeloid leukemia
macrophage
bronchoalveolar lavage cell
monocyte
acute lymphoblastic leukemia
precursor T lymphoblastic leukemia
B cell
thymocyte
lymphocyte
hl60 promyelocytic leukemia
hematopoietic stem cell
myelogenous leukemia
CD138+ plasma cell myeloma
tonsil
B−cell lymphoma
lymph node
ts anaplastic large cell lymphoma
k562 myelogenous leukaemia
MOLT4 T cell acute lymphoblastic leukemia
caudate nucleus
brain
cerebellum Huntingtons
cerebellum
frontal cortex Huntingtons
frontal cortex
brain bipolar disorder
caudate nucleus Huntingtons
hypothalamus

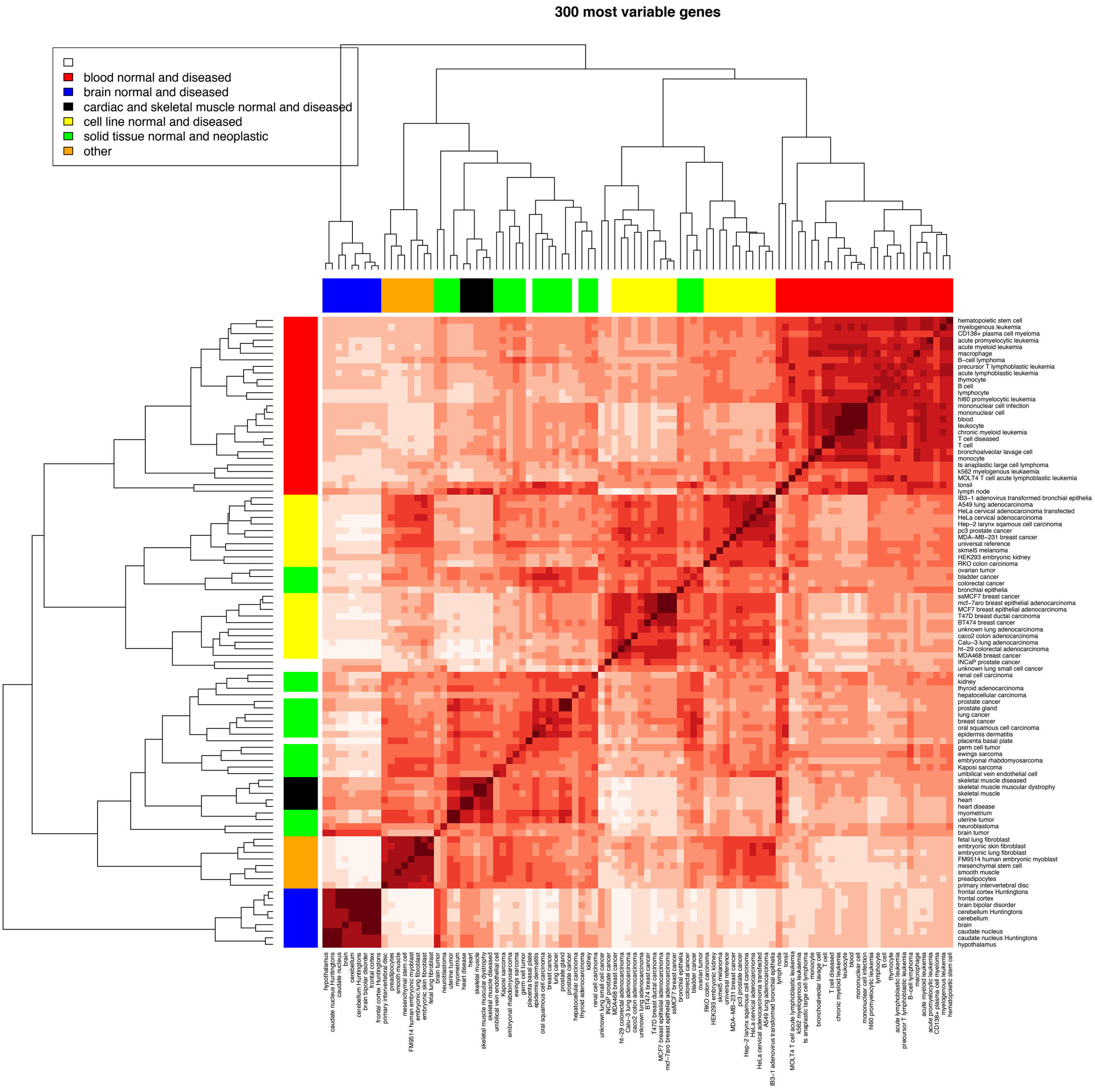

550 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
lung cancer
bladder cancer
breast cancer
epidermis dermatitis
renal cell carcinoma
kidney
germ cell tumor
thyroid adenocarcinoma
hepatocellular carcinoma
ovarian tumor
colorectal cancer
bronchial epithelia
HeLa cervical adenocarcinoma transfected
HeLa cervical adenocarcinoma
A549 lung adenocarcinoma
Hep−2 larynx sqamous cell carcinoma
MDA−MB−231 breast cancer
pc3 prostate cancer
universal reference
IB3−1 adenovirus transformed bronchial epithelia
skmel5 melanoma
HEK293 embryonic kidney
RKO colon carcinoma
ssMCF7 breast cancer
mcf−7aro breast epithelial adenocarcinoma
MCF7 breast epithelial adenocarcinoma
T47D breast ductal carcinoma
BT474 breast cancer
unknown lung adenocarcinoma
caco2 colon adenocarcinoma
Calu−3 lung adenocarcinoma
ht−29 colorectal adenocarcinoma
MDA468 breast cancer
INCaP prostate cancer
unknown lung small cell cancer
ts anaplastic large cell lymphoma
k562 myelogenous leukaemia
MOLT4 T cell acute lymphoblastic leukemia
ewings sarcoma
embryonal rhabdomyosarcoma
Kaposi sarcoma
umbilical vein endothelial cell
oral squamous cell carcinoma
placenta basal plate
prostate cancer
prostate gland
myometrium
uterine tumor
embryonic skin fibroblast
fetal lung fibroblast
embryonic lung fibroblast
FM9514 human embryonic myoblast
preadipocytes
smooth muscle
mesenchymal stem cell
primary intervertebral disc
skeletal muscle diseased
skeletal muscle muscular dystrophy
skeletal muscle
heart
heart disease
neuroblastoma
brain tumor
acute promyelocytic leukemia
acute myeloid leukemia
acute lymphoblastic leukemia
macrophage
T cell
T cell diseased
bronchoalveolar lavage cell
monocyte
mononuclear cell infection
mononuclear cell
blood
leukocyte
chronic myeloid leukemia
precursor T lymphoblastic leukemia
thymocyte
B cell
lymphocyte
hl60 promyelocytic leukemia
hematopoietic stem cell
myelogenous leukemia
CD138+ plasma cell myeloma
tonsil
B−cell lymphoma
lymph node
frontal cortex Huntingtons
brain bipolar disorder
frontal cortex
cerebellum Huntingtons
cerebellum
caudate nucleus
brain
caudate nucleus Huntingtons
hypothalamus

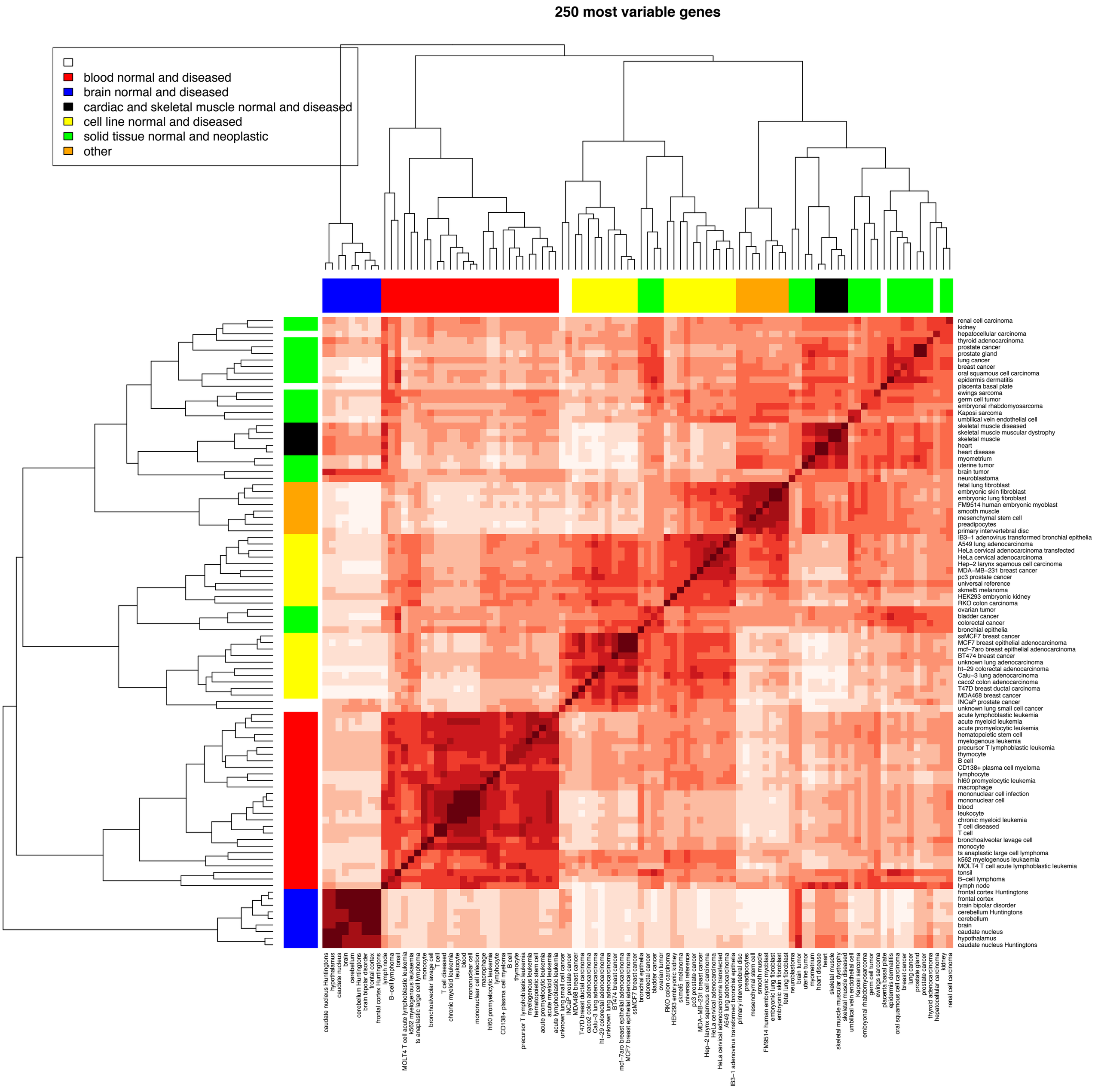

500 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
IB3−1 adenovirus transformed bronchial epithelia
A549 lung adenocarcinoma
pc3 prostate cancer
MDA−MB−231 breast cancer
HeLa cervical adenocarcinoma
HeLa cervical adenocarcinoma transfected
Hep−2 larynx sqamous cell carcinoma
universal reference
skmel5 melanoma
HEK293 embryonic kidney
RKO colon carcinoma
ssMCF7 breast cancer
mcf−7aro breast epithelial adenocarcinoma
MCF7 breast epithelial adenocarcinoma
T47D breast ductal carcinoma
BT474 breast cancer
unknown lung adenocarcinoma
caco2 colon adenocarcinoma
Calu−3 lung adenocarcinoma
ht−29 colorectal adenocarcinoma
MDA468 breast cancer
lung cancer
bladder cancer
breast cancer
epidermis dermatitis
renal cell carcinoma
kidney
germ cell tumor
thyroid adenocarcinoma
hepatocellular carcinoma
ovarian tumor
colorectal cancer
bronchial epithelia
ewings sarcoma
embryonal rhabdomyosarcoma
Kaposi sarcoma
umbilical vein endothelial cell
oral squamous cell carcinoma
placenta basal plate
prostate cancer
prostate gland
myometrium
uterine tumor
embryonic skin fibroblast
fetal lung fibroblast
embryonic lung fibroblast
FM9514 human embryonic myoblast
preadipocytes
smooth muscle
mesenchymal stem cell
primary intervertebral disc
skeletal muscle muscular dystrophy
skeletal muscle diseased
skeletal muscle
heart
heart disease
neuroblastoma
brain tumor
INCaP prostate cancer
unknown lung small cell cancer
precursor T lymphoblastic leukemia
thymocyte
B cell
lymphocyte
hl60 promyelocytic leukemia
acute myeloid leukemia
acute promyelocytic leukemia
acute lymphoblastic leukemia
macrophage
hematopoietic stem cell
myelogenous leukemia
CD138+ plasma cell myeloma
mononuclear cell infection
mononuclear cell
blood
leukocyte
chronic myeloid leukemia
T cell
T cell diseased
bronchoalveolar lavage cell
monocyte
tonsil
B−cell lymphoma
lymph node
ts anaplastic large cell lymphoma
k562 myelogenous leukaemia
MOLT4 T cell acute lymphoblastic leukemia
frontal cortex Huntingtons
brain bipolar disorder
frontal cortex
cerebellum Huntingtons
cerebellum
caudate nucleus
brain
caudate nucleus Huntingtons
hypothalamus
hypothalamus
caudate nucleus Huntingtons
brain
caudate nucleus
cerebellum
cerebellum Huntingtons
frontal cortex
brain bipolar disorder
frontal cortex Huntingtons
MOLT4 T cell acute lymphoblastic leukemia
k562 myelogenous leukaemia
ts anaplastic large cell lymphoma
lymph node
tonsil
B−cell lymphoma
monocyte
bronchoalveolar lavage cell
T cell diseased
T cell
chronic myeloid leukemia
leukocyte
blood
mononuclear cell
mononuclear cell infection
CD138+ plasma cell myeloma
myelogenous leukemia
hematopoietic stem cell
macrophage
acute lymphoblastic leukemia
acute promyelocytic leukemia
acute myeloid leukemia
hl60 promyelocytic leukemia
lymphocyte
B cell
thymocyte
precursor T lymphoblastic leukemia
unknown lung small cell cancer
INCaP prostate cancer
brain tumor
neuroblastoma
heart disease
heart
skeletal muscle
skeletal muscle diseased
skeletal muscle muscular dystrophy
primary intervertebral disc
mesenchymal stem cell
smooth muscle
preadipocytes
FM9514 human embryonic myoblast
embryonic lung fibroblast
fetal lung fibroblast
embryonic skin fibroblast
uterine tumor
myometrium
prostate gland
prostate cancer
placenta basal plate
oral squamous cell carcinoma
umbilical vein endothelial cell
Kaposi sarcoma
embryonal rhabdomyosarcoma
ewings sarcoma
bronchial epithelia
colorectal cancer
ovarian tumor
hepatocellular carcinoma
thyroid adenocarcinoma
germ cell tumor
kidney
renal cell carcinoma
epidermis dermatitis
breast cancer
bladder cancer
lung cancer
MDA468 breast cancer
ht−29 colorectal adenocarcinoma
Calu−3 lung adenocarcinoma
caco2 colon adenocarcinoma
unknown lung adenocarcinoma
BT474 breast cancer
T47D breast ductal carcinoma
MCF7 breast epithelial adenocarcinoma
mcf−7aro breast epithelial adenocarcinoma
ssMCF7 breast cancer
RKO colon carcinoma
HEK293 embryonic kidney
skmel5 melanoma
universal reference
Hep−2 larynx sqamous cell carcinoma
HeLa cervical adenocarcinoma transfected
HeLa cervical adenocarcinoma
MDA−MB−231 breast cancer
pc3 prostate cancer
A549 lung adenocarcinoma
IB3−1 adenovirus transformed bronchial epithelia

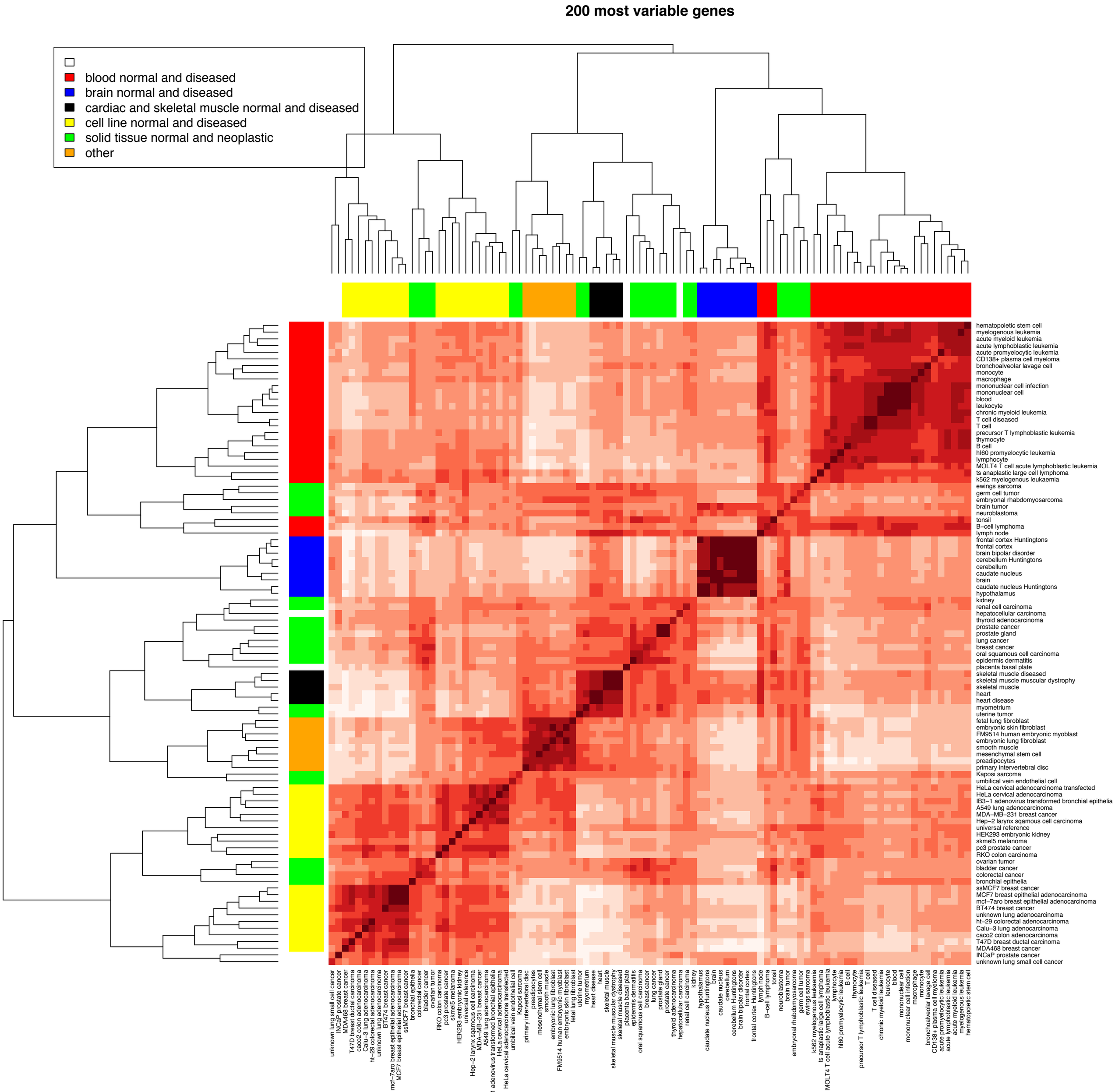

450 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
renal cell carcinoma
kidney
germ cell tumor
lung cancer
bladder cancer
breast cancer
thyroid adenocarcinoma
hepatocellular carcinoma
prostate cancer
prostate gland
epidermis dermatitis
ewings sarcoma
embryonal rhabdomyosarcoma
Kaposi sarcoma
umbilical vein endothelial cell
oral squamous cell carcinoma
placenta basal plate
myometrium
uterine tumor
embryonic skin fibroblast
fetal lung fibroblast
embryonic lung fibroblast
FM9514 human embryonic myoblast
mesenchymal stem cell
preadipocytes
smooth muscle
primary intervertebral disc
skeletal muscle diseased
skeletal muscle muscular dystrophy
skeletal muscle
heart
heart disease
brain tumor
neuroblastoma
INCaP prostate cancer
unknown lung small cell cancer
IB3−1 adenovirus transformed bronchial epithelia
A549 lung adenocarcinoma
HeLa cervical adenocarcinoma
HeLa cervical adenocarcinoma transfected
pc3 prostate cancer
MDA−MB−231 breast cancer
Hep−2 larynx sqamous cell carcinoma
universal reference
skmel5 melanoma
HEK293 embryonic kidney
RKO colon carcinoma
ssMCF7 breast cancer
mcf−7aro breast epithelial adenocarcinoma
MCF7 breast epithelial adenocarcinoma
T47D breast ductal carcinoma
BT474 breast cancer
ht−29 colorectal adenocarcinoma
MDA468 breast cancer
unknown lung adenocarcinoma
caco2 colon adenocarcinoma
Calu−3 lung adenocarcinoma
ovarian tumor
colorectal cancer
bronchial epithelia
acute lymphoblastic leukemia
precursor T lymphoblastic leukemia
B cell
thymocyte
lymphocyte
acute myeloid leukemia
acute promyelocytic leukemia
macrophage
hl60 promyelocytic leukemia
hematopoietic stem cell
myelogenous leukemia
CD138+ plasma cell myeloma
mononuclear cell infection
mononuclear cell
blood
leukocyte
chronic myeloid leukemia
T cell
T cell diseased
bronchoalveolar lavage cell
monocyte
tonsil
B−cell lymphoma
lymph node
ts anaplastic large cell lymphoma
k562 myelogenous leukaemia
MOLT4 T cell acute lymphoblastic leukemia
frontal cortex Huntingtons
frontal cortex
brain bipolar disorder
cerebellum Huntingtons
cerebellum
caudate nucleus
brain
caudate nucleus Huntingtons
hypothalamus
hypothalamus
caudate nucleus Huntingtons
brain
cerebellum
cerebellum Huntingtons
brain bipolar disorder
frontal cortex
frontal cortex Huntingtons
MOLT4 T cell acute lymphoblastic leukemia
k562 myelogenous leukaemia
ts anaplastic large cell lymphoma
lymph node
B−cell lymphoma
tonsil
monocyte
bronchoalveolar lavage cell
T cell diseased
T cell
chronic myeloid leukemia
leukocyte
blood
mononuclear cell
mononuclear cell infection
CD138+ plasma cell myeloma
myelogenous leukemia
hematopoietic stem cell
hl60 promyelocytic leukemia
macrophage
acute promyelocytic leukemia
acute myeloid leukemia
lymphocyte
thymocyte
B cell
precursor T lymphoblastic leukemia
acute lymphoblastic leukemia
bronchial epithelia
colorectal cancer
ovarian tumor
Calu−3 lung adenocarcinoma
caco2 colon adenocarcinoma
unknown lung adenocarcinoma
MDA468 breast cancer
ht−29 colorectal adenocarcinoma
BT474 breast cancer
T47D breast ductal carcinoma
MCF7 breast epithelial adenocarcinoma
mcf−7aro breast epithelial adenocarcinoma
ssMCF7 breast cancer
RKO colon carcinoma
HEK293 embryonic kidney
skmel5 melanoma
universal reference
Hep−2 larynx sqamous cell carcinoma
MDA−MB−231 breast cancer
pc3 prostate cancer
HeLa cervical adenocarcinoma transfected
HeLa cervical adenocarcinoma
A549 lung adenocarcinoma
IB3−1 adenovirus transformed bronchial epithelia
unknown lung small cell cancer
INCaP prostate cancer
neuroblastoma
brain tumor
heart disease
heart
skeletal muscle
skeletal muscle muscular dystrophy
skeletal muscle diseased
primary intervertebral disc
smooth muscle
preadipocytes
mesenchymal stem cell
FM9514 human embryonic myoblast
embryonic lung fibroblast
fetal lung fibroblast
embryonic skin fibroblast
uterine tumor
myometrium
placenta basal plate
oral squamous cell carcinoma
umbilical vein endothelial cell
Kaposi sarcoma
embryonal rhabdomyosarcoma
ewings sarcoma
epidermis dermatitis
prostate gland
prostate cancer
hepatocellular carcinoma
thyroid adenocarcinoma
breast cancer
bladder cancer
lung cancer
germ cell tumor
kidney
renal cell carcinoma

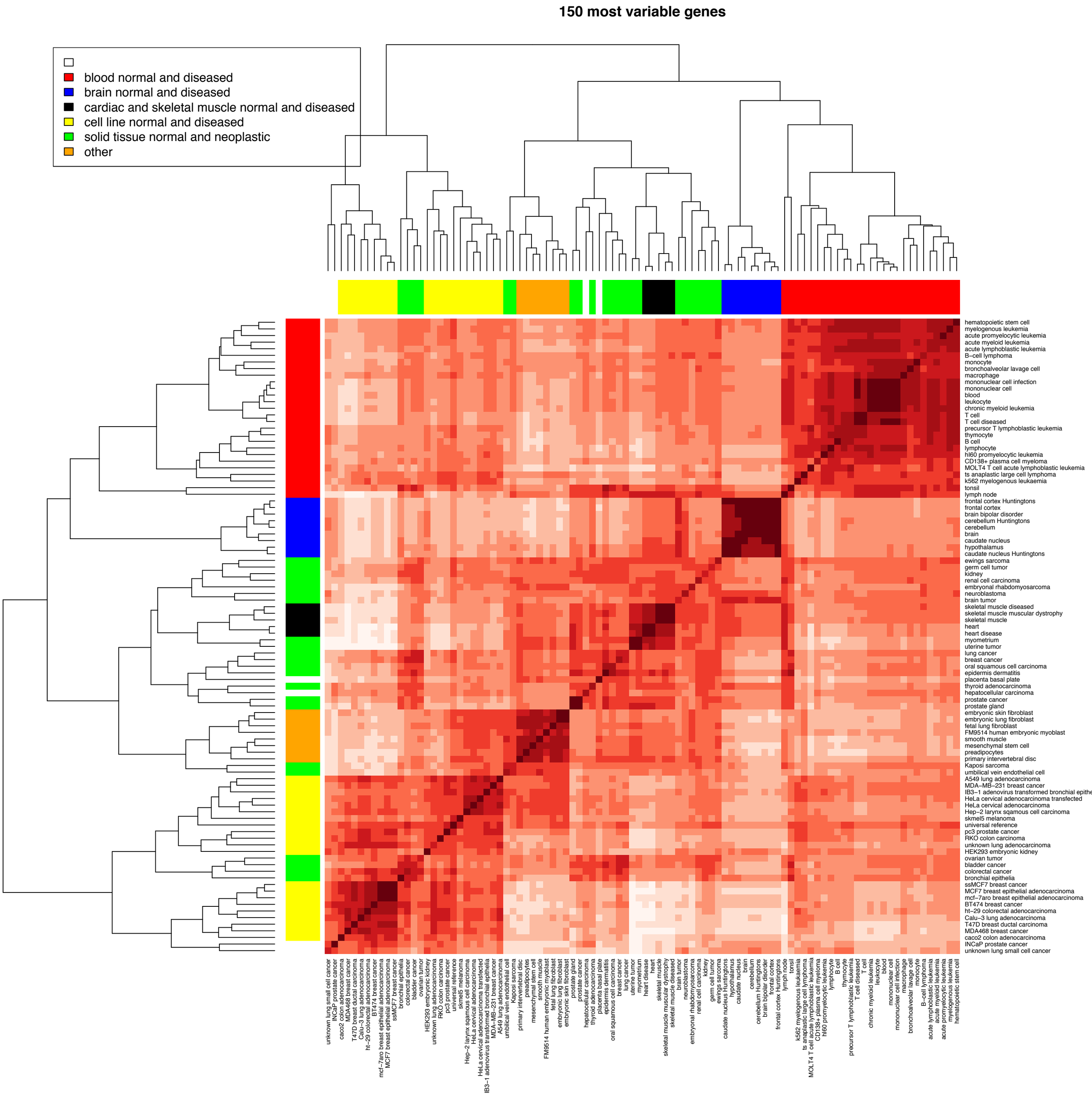

400 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
IB3−1 adenovirus transformed bronchial epithelia
A549 lung adenocarcinoma
HeLa cervical adenocarcinoma transfected
HeLa cervical adenocarcinoma
MDA−MB−231 breast cancer
pc3 prostate cancer
Hep−2 larynx sqamous cell carcinoma
universal reference
skmel5 melanoma
HEK293 embryonic kidney
RKO colon carcinoma
ssMCF7 breast cancer
mcf−7aro breast epithelial adenocarcinoma
MCF7 breast epithelial adenocarcinoma
T47D breast ductal carcinoma
BT474 breast cancer
unknown lung adenocarcinoma
caco2 colon adenocarcinoma
Calu−3 lung adenocarcinoma
ht−29 colorectal adenocarcinoma
MDA468 breast cancer
INCaP prostate cancer
unknown lung small cell cancer
ewings sarcoma
embryonal rhabdomyosarcoma
oral squamous cell carcinoma
placenta basal plate
Kaposi sarcoma
umbilical vein endothelial cell
myometrium
uterine tumor
embryonic skin fibroblast
fetal lung fibroblast
embryonic lung fibroblast
FM9514 human embryonic myoblast
mesenchymal stem cell
preadipocytes
smooth muscle
primary intervertebral disc
skeletal muscle diseased
skeletal muscle muscular dystrophy
skeletal muscle
heart
heart disease
brain tumor
neuroblastoma
hematopoietic stem cell
myelogenous leukemia
acute myeloid leukemia
acute promyelocytic leukemia
macrophage
B−cell lymphoma
CD138+ plasma cell myeloma
precursor T lymphoblastic leukemia
acute lymphoblastic leukemia
thymocyte
B cell
lymphocyte
hl60 promyelocytic leukemia
mononuclear cell infection
mononuclear cell
blood
leukocyte
chronic myeloid leukemia
T cell diseased
T cell
bronchoalveolar lavage cell
monocyte
ts anaplastic large cell lymphoma
k562 myelogenous leukaemia
MOLT4 T cell acute lymphoblastic leukemia
lung cancer
bladder cancer
breast cancer
epidermis dermatitis
renal cell carcinoma
kidney
germ cell tumor
hepatocellular carcinoma
prostate cancer
prostate gland
thyroid adenocarcinoma
ovarian tumor
colorectal cancer
bronchial epithelia
tonsil
lymph node
frontal cortex Huntingtons
frontal cortex
brain bipolar disorder
cerebellum Huntingtons
cerebellum
caudate nucleus
brain
caudate nucleus Huntingtons
hypothalamus
hypothalamus
caudate nucleus Huntingtons
brain
caudate nucleus
cerebellum
cerebellum Huntingtons
brain bipolar disorder
frontal cortex
frontal cortex Huntingtons
lymph node
tonsil
bronchial epithelia
colorectal cancer
ovarian tumor
thyroid adenocarcinoma
prostate gland
prostate cancer
hepatocellular carcinoma
germ cell tumor
kidney
renal cell carcinoma
epidermis dermatitis
breast cancer
bladder cancer
lung cancer
MOLT4 T cell acute lymphoblastic leukemia
k562 myelogenous leukaemia
ts anaplastic large cell lymphoma
monocyte
bronchoalveolar lavage cell
T cell
T cell diseased
chronic myeloid leukemia
leukocyte
blood
mononuclear cell
mononuclear cell infection
hl60 promyelocytic leukemia
lymphocyte
B cell
thymocyte
acute lymphoblastic leukemia
precursor T lymphoblastic leukemia
CD138+ plasma cell myeloma
B−cell lymphoma
macrophage
acute promyelocytic leukemia
acute myeloid leukemia
myelogenous leukemia
hematopoietic stem cell
neuroblastoma
brain tumor
heart disease
heart
skeletal muscle
skeletal muscle muscular dystrophy
skeletal muscle diseased
primary intervertebral disc
smooth muscle
preadipocytes
mesenchymal stem cell
FM9514 human embryonic myoblast
embryonic lung fibroblast
fetal lung fibroblast
embryonic skin fibroblast
uterine tumor
myometrium
umbilical vein endothelial cell
Kaposi sarcoma
placenta basal plate
oral squamous cell carcinoma
embryonal rhabdomyosarcoma
ewings sarcoma
unknown lung small cell cancer
INCaP prostate cancer
MDA468 breast cancer
ht−29 colorectal adenocarcinoma
Calu−3 lung adenocarcinoma
caco2 colon adenocarcinoma
unknown lung adenocarcinoma
BT474 breast cancer
T47D breast ductal carcinoma
MCF7 breast epithelial adenocarcinoma
mcf−7aro breast epithelial adenocarcinoma
ssMCF7 breast cancer
RKO colon carcinoma
HEK293 embryonic kidney
skmel5 melanoma
universal reference
Hep−2 larynx sqamous cell carcinoma
pc3 prostate cancer
MDA−MB−231 breast cancer
HeLa cervical adenocarcinoma
HeLa cervical adenocarcinoma transfected
A549 lung adenocarcinoma
IB3−1 adenovirus transformed bronchial epithelia

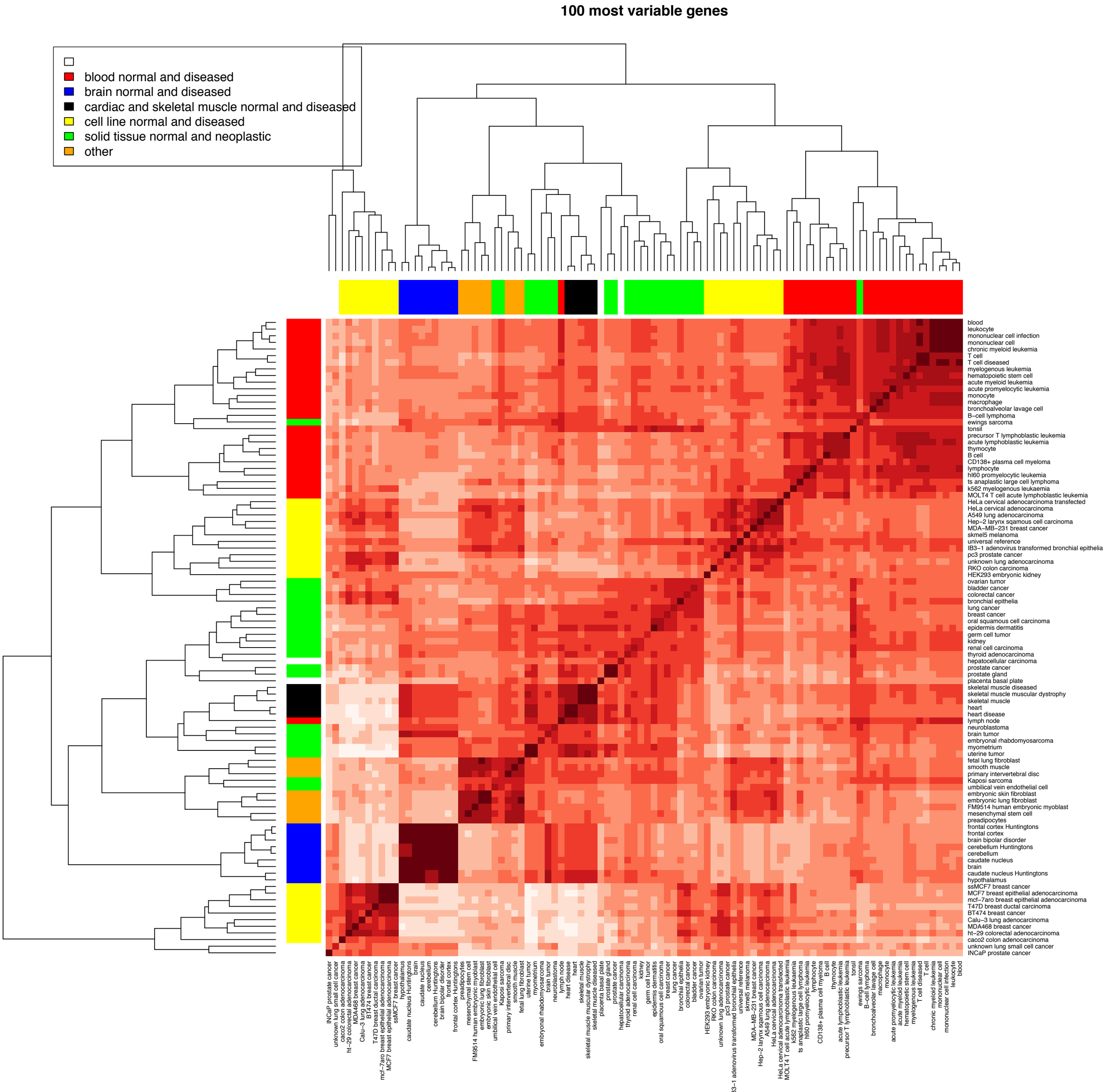

350 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
bladder cancer
lung cancer
oral squamous cell carcinoma
epidermis dermatitis
breast cancer
renal cell carcinoma
kidney
thyroid adenocarcinoma
hepatocellular carcinoma
prostate cancer
prostate gland
ovarian tumor
colorectal cancer
bronchial epithelia
skeletal muscle diseased
skeletal muscle muscular dystrophy
skeletal muscle
heart
heart disease
myometrium
uterine tumor
brain tumor
neuroblastoma
fetal lung fibroblast
embryonic skin fibroblast
embryonic lung fibroblast
mesenchymal stem cell
smooth muscle
FM9514 human embryonic myoblast
primary intervertebral disc
preadipocytes
germ cell tumor
ewings sarcoma
embryonal rhabdomyosarcoma
Kaposi sarcoma
umbilical vein endothelial cell
placenta basal plate
IB3−1 adenovirus transformed bronchial epithelia
A549 lung adenocarcinoma
HeLa cervical adenocarcinoma transfected
HeLa cervical adenocarcinoma
Hep−2 larynx sqamous cell carcinoma
MDA−MB−231 breast cancer
pc3 prostate cancer
universal reference
skmel5 melanoma
HEK293 embryonic kidney
RKO colon carcinoma
ssMCF7 breast cancer
mcf−7aro breast epithelial adenocarcinoma
MCF7 breast epithelial adenocarcinoma
T47D breast ductal carcinoma
BT474 breast cancer
Calu−3 lung adenocarcinoma
ht−29 colorectal adenocarcinoma
MDA468 breast cancer
unknown lung adenocarcinoma
caco2 colon adenocarcinoma
INCaP prostate cancer
unknown lung small cell cancer
precursor T lymphoblastic leukemia
acute lymphoblastic leukemia
thymocyte
B cell
hematopoietic stem cell
myelogenous leukemia
CD138+ plasma cell myeloma
acute myeloid leukemia
acute promyelocytic leukemia
macrophage
lymphocyte
hl60 promyelocytic leukemia
mononuclear cell infection
mononuclear cell
blood
leukocyte
chronic myeloid leukemia
T cell
T cell diseased
bronchoalveolar lavage cell
monocyte
ts anaplastic large cell lymphoma
k562 myelogenous leukaemia
MOLT4 T cell acute lymphoblastic leukemia
tonsil
B−cell lymphoma
lymph node
frontal cortex Huntingtons
frontal cortex
brain bipolar disorder
cerebellum Huntingtons
cerebellum
brain
caudate nucleus
caudate nucleus Huntingtons
hypothalamus


300 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
hematopoietic stem cell
myelogenous leukemia
CD138+ plasma cell myeloma
acute promyelocytic leukemia
acute myeloid leukemia
macrophage
B−cell lymphoma
precursor T lymphoblastic leukemia
acute lymphoblastic leukemia
thymocyte
B cell
lymphocyte
hl60 promyelocytic leukemia
mononuclear cell infection
mononuclear cell
blood
leukocyte
chronic myeloid leukemia
T cell diseased
T cell
bronchoalveolar lavage cell
monocyte
ts anaplastic large cell lymphoma
k562 myelogenous leukaemia
MOLT4 T cell acute lymphoblastic leukemia
tonsil
lymph node
IB3−1 adenovirus transformed bronchial epithelia
A549 lung adenocarcinoma
HeLa cervical adenocarcinoma transfected
HeLa cervical adenocarcinoma
Hep−2 larynx sqamous cell carcinoma
pc3 prostate cancer
MDA−MB−231 breast cancer
universal reference
skmel5 melanoma
HEK293 embryonic kidney
RKO colon carcinoma
ovarian tumor
bladder cancer
colorectal cancer
bronchial epithelia
ssMCF7 breast cancer
mcf−7aro breast epithelial adenocarcinoma
MCF7 breast epithelial adenocarcinoma
T47D breast ductal carcinoma
BT474 breast cancer
unknown lung adenocarcinoma
caco2 colon adenocarcinoma
Calu−3 lung adenocarcinoma
ht−29 colorectal adenocarcinoma
MDA468 breast cancer
INCaP prostate cancer
unknown lung small cell cancer
renal cell carcinoma
kidney
thyroid adenocarcinoma
hepatocellular carcinoma
prostate cancer
prostate gland
lung cancer
breast cancer
oral squamous cell carcinoma
epidermis dermatitis
placenta basal plate
germ cell tumor
ewings sarcoma
embryonal rhabdomyosarcoma
Kaposi sarcoma
umbilical vein endothelial cell
skeletal muscle diseased
skeletal muscle muscular dystrophy
skeletal muscle
heart
heart disease
myometrium
uterine tumor
neuroblastoma
brain tumor
fetal lung fibroblast
embryonic skin fibroblast
embryonic lung fibroblast
FM9514 human embryonic myoblast
mesenchymal stem cell
smooth muscle
preadipocytes
primary intervertebral disc
frontal cortex Huntingtons
frontal cortex
brain bipolar disorder
cerebellum Huntingtons
cerebellum
brain
caudate nucleus
caudate nucleus Huntingtons
hypothalamus

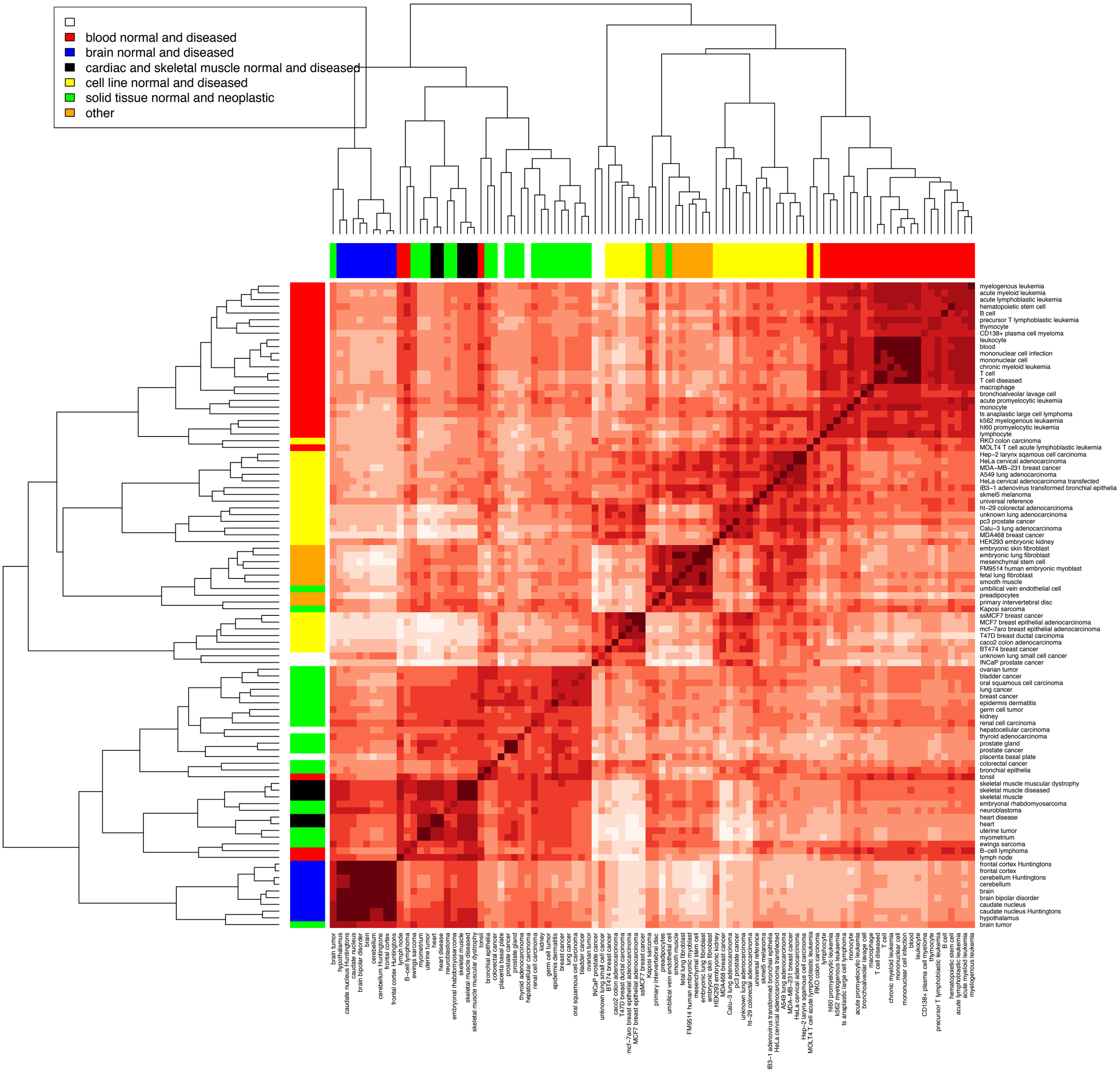

250 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
renal cell carcinoma
kidney
hepatocellular carcinoma
thyroid adenocarcinoma
prostate cancer
prostate gland
lung cancer
breast cancer
oral squamous cell carcinoma
epidermis dermatitis
placenta basal plate
ewings sarcoma
germ cell tumor
embryonal rhabdomyosarcoma
Kaposi sarcoma
umbilical vein endothelial cell
skeletal muscle diseased
skeletal muscle muscular dystrophy
skeletal muscle
heart
heart disease
myometrium
uterine tumor
brain tumor
neuroblastoma
fetal lung fibroblast
embryonic skin fibroblast
embryonic lung fibroblast
FM9514 human embryonic myoblast
smooth muscle
mesenchymal stem cell
preadipocytes
primary intervertebral disc
IB3−1 adenovirus transformed bronchial epithelia
A549 lung adenocarcinoma
HeLa cervical adenocarcinoma transfected
HeLa cervical adenocarcinoma
Hep−2 larynx sqamous cell carcinoma
MDA−MB−231 breast cancer
pc3 prostate cancer
universal reference
skmel5 melanoma
HEK293 embryonic kidney
RKO colon carcinoma
ovarian tumor
bladder cancer
colorectal cancer
bronchial epithelia
ssMCF7 breast cancer
MCF7 breast epithelial adenocarcinoma
mcf−7aro breast epithelial adenocarcinoma
BT474 breast cancer
unknown lung adenocarcinoma
ht−29 colorectal adenocarcinoma
Calu−3 lung adenocarcinoma
caco2 colon adenocarcinoma
T47D breast ductal carcinoma
MDA468 breast cancer
INCaP prostate cancer
unknown lung small cell cancer
acute lymphoblastic leukemia
acute myeloid leukemia
acute promyelocytic leukemia
hematopoietic stem cell
myelogenous leukemia
precursor T lymphoblastic leukemia
thymocyte
B cell
CD138+ plasma cell myeloma
lymphocyte
hl60 promyelocytic leukemia
macrophage
mononuclear cell infection
mononuclear cell
blood
leukocyte
chronic myeloid leukemia
T cell diseased
T cell
bronchoalveolar lavage cell
monocyte
ts anaplastic large cell lymphoma
k562 myelogenous leukaemia
MOLT4 T cell acute lymphoblastic leukemia
tonsil
B−cell lymphoma
lymph node
frontal cortex Huntingtons
frontal cortex
brain bipolar disorder
cerebellum Huntingtons
cerebellum
brain
caudate nucleus
hypothalamus
caudate nucleus Huntingtons

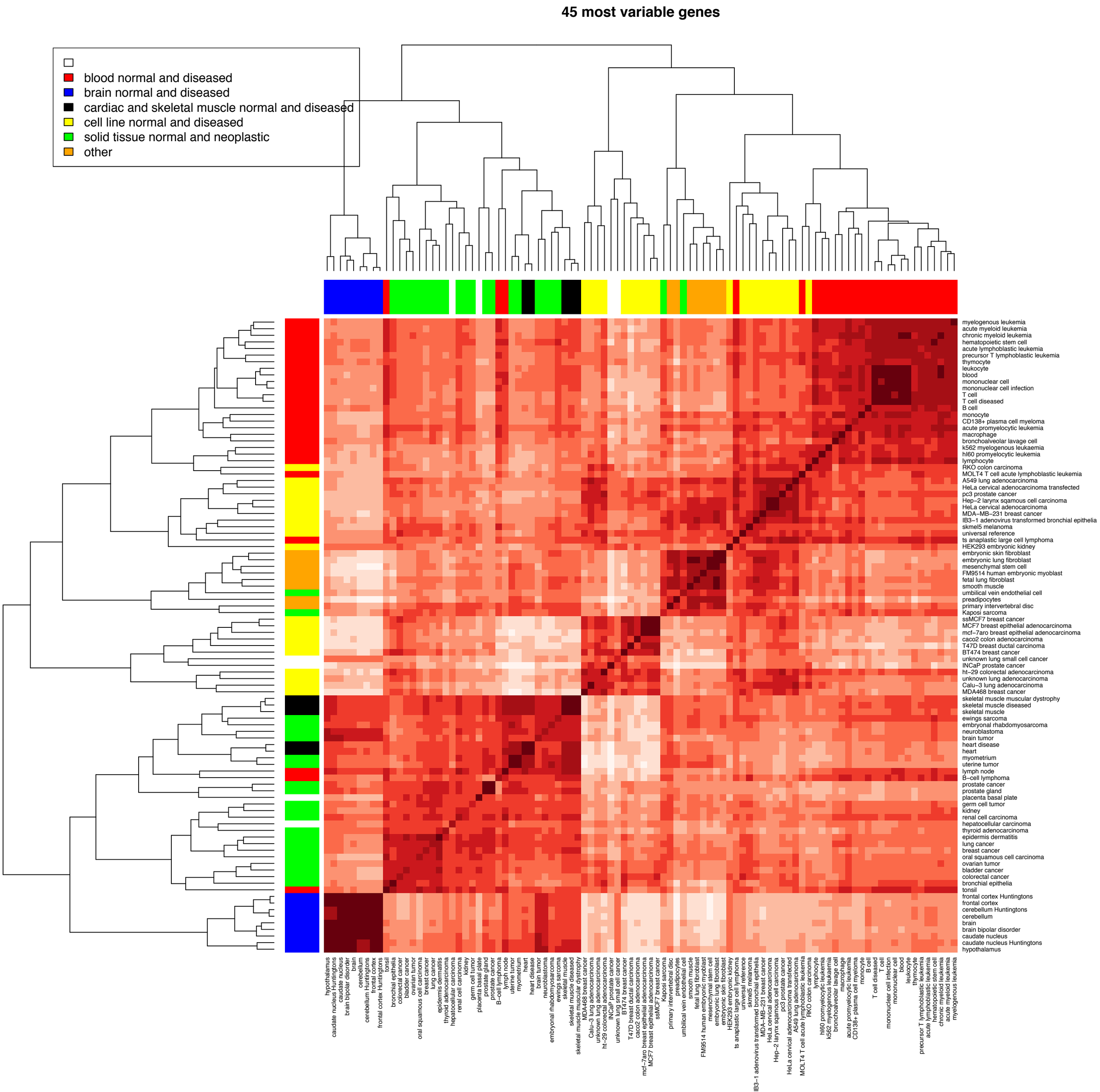

200 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
hematopoietic stem cell
myelogenous leukemia
acute myeloid leukemia
acute lymphoblastic leukemia
acute promyelocytic leukemia
CD138+ plasma cell myeloma
bronchoalveolar lavage cell
monocyte
macrophage
mononuclear cell infection
mononuclear cell
blood
leukocyte
chronic myeloid leukemia
T cell diseased
T cell
precursor T lymphoblastic leukemia
thymocyte
B cell
hl60 promyelocytic leukemia
lymphocyte
MOLT4 T cell acute lymphoblastic leukemia
ts anaplastic large cell lymphoma
k562 myelogenous leukaemia
ewings sarcoma
germ cell tumor
embryonal rhabdomyosarcoma
brain tumor
neuroblastoma
tonsil
B–cell lymphoma
lymph node
frontal cortex Huntingtons
frontal cortex
brain bipolar disorder
cerebellum Huntingtons
cerebellum
caudate nucleus
brain
caudate nucleus Huntingtons
hypothalamus
kidney
renal cell carcinoma
hepatocellular carcinoma
thyroid adenocarcinoma
prostate cancer
prostate gland
lung cancer
breast cancer
oral squamous cell carcinoma
epidermis dermatitis
placenta basal plate
skeletal muscle diseased
skeletal muscle muscular dystrophy
skeletal muscle
heart
heart disease
myometrium
uterine tumor
fetal lung fibroblast
embryonic skin fibroblast
FM9514 human embryonic myoblast
embryonic lung fibroblast
smooth muscle
mesenchymal stem cell
preadipocytes
primary intervertebral disc
Kaposi sarcoma
umbilical vein endothelial cell
HeLa cervical adenocarcinoma transfected
HeLa cervical adenocarcinoma
IB3–1 adenovirus transformed bronchial epithelia
A549 lung adenocarcinoma
MDA–MB–231 breast cancer
Hep–2 larynx sqamous cell carcinoma
universal reference
HEK293 embryonic kidney
skmel5 melanoma
pc3 prostate cancer
RKO colon carcinoma
ovarian tumor
bladder cancer
colorectal cancer
bronchial epithelia
ssMCF7 breast cancer
MCF7 breast epithelial adenocarcinoma
mcf–7aro breast epithelial adenocarcinoma
BT474 breast cancer
unknown lung adenocarcinoma
ht–29 colorectal adenocarcinoma
Calu–3 lung adenocarcinoma
caco2 colon adenocarcinoma
T47D breast ductal carcinoma
MDA468 breast cancer
INCaP prostate cancer
unknown lung small cell cancer

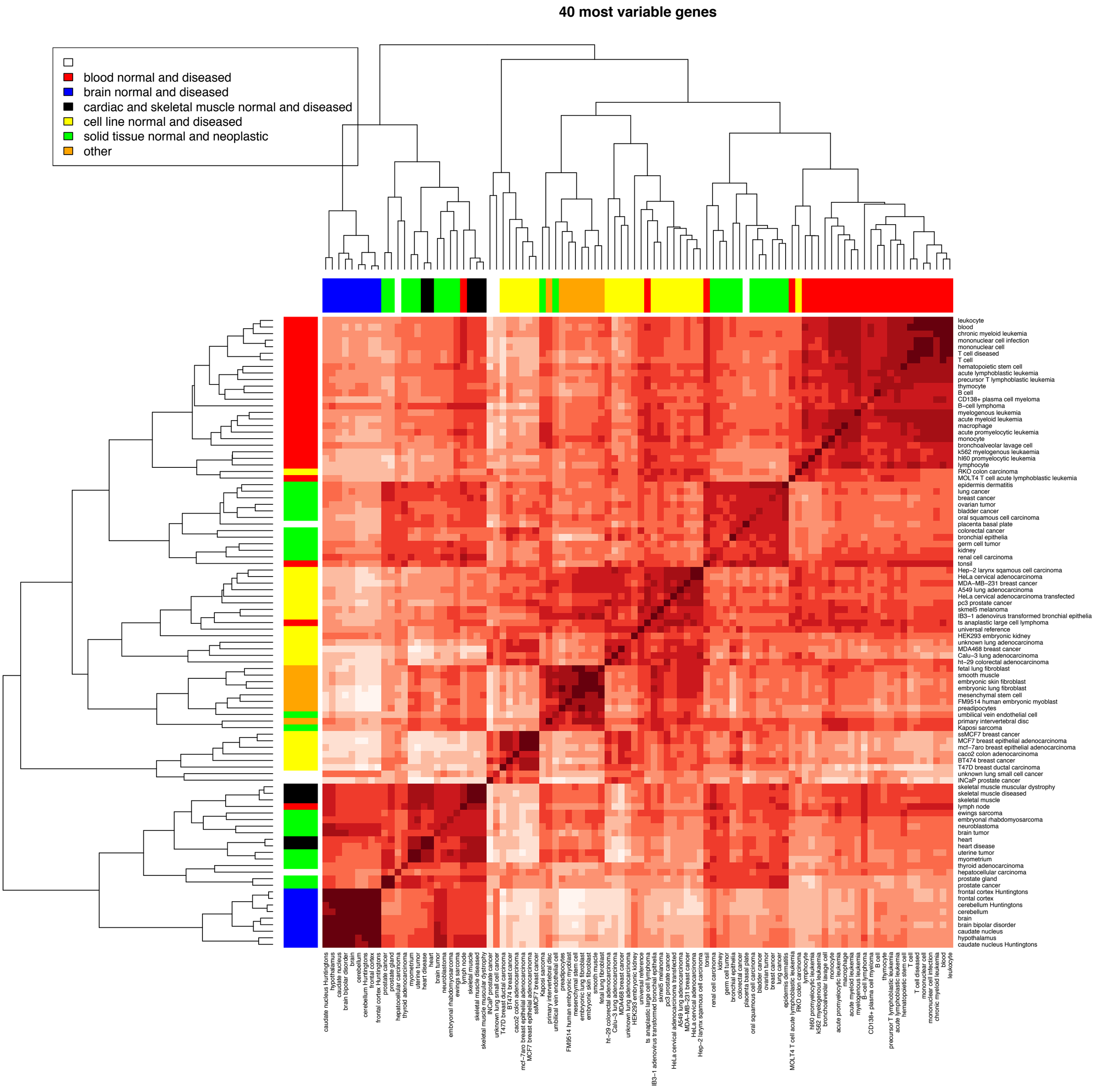

150 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
hematopoietic stem cell
myelogenous leukemia
acute promyelocytic leukemia
acute myeloid leukemia
acute lymphoblastic leukemia
B–cell lymphoma
monocyte
bronchoalveolar lavage cell
macrophage
mononuclear cell infection
mononuclear cell
blood
leukocyte
chronic myeloid leukemia
T cell
T cell diseased
precursor T lymphoblastic leukemia
thymocyte
B cell
lymphocyte
hl60 promyelocytic leukemia
CD138+ plasma cell myeloma
MOLT4 T cell acute lymphoblastic leukemia
ts anaplastic large cell lymphoma
k562 myelogenous leukaemia
tonsil
lymph node
frontal cortex Huntingtons
frontal cortex
brain bipolar disorder
cerebellum Huntingtons
cerebellum
brain
caudate nucleus
hypothalamus
caudate nucleus Huntingtons
ewings sarcoma
germ cell tumor
kidney
renal cell carcinoma
embryonal rhabdomyosarcoma
neuroblastoma
brain tumor
skeletal muscle diseased
skeletal muscle muscular dystrophy
skeletal muscle
heart
heart disease
myometrium
uterine tumor
lung cancer
breast cancer
oral squamous cell carcinoma
epidermis dermatitis
placenta basal plate
thyroid adenocarcinoma
hepatocellular carcinoma
prostate cancer
prostate gland
embryonic skin fibroblast
embryonic lung fibroblast
fetal lung fibroblast
FM9514 human embryonic myoblast
smooth muscle
mesenchymal stem cell
preadipocytes
primary intervertebral disc
Kaposi sarcoma
umbilical vein endothelial cell
A549 lung adenocarcinoma
MDA–MB–231 breast cancer
IB3–1 adenovirus transformed bronchial epithelia
HeLa cervical adenocarcinoma transfected
HeLa cervical adenocarcinoma
Hep–2 larynx sqamous cell carcinoma
skmel5 melanoma
universal reference
pc3 prostate cancer
RKO colon carcinoma
unknown lung adenocarcinoma
HEK293 embryonic kidney
ovarian tumor
bladder cancer
colorectal cancer
bronchial epithelia
ssMCF7 breast cancer
MCF7 breast epithelial adenocarcinoma
mcf–7aro breast epithelial adenocarcinoma
BT474 breast cancer
ht–29 colorectal adenocarcinoma
Calu–3 lung adenocarcinoma
T47D breast ductal carcinoma
MDA468 breast cancer
caco2 colon adenocarcinoma
INCaP prostate cancer
unknown lung small cell cancer

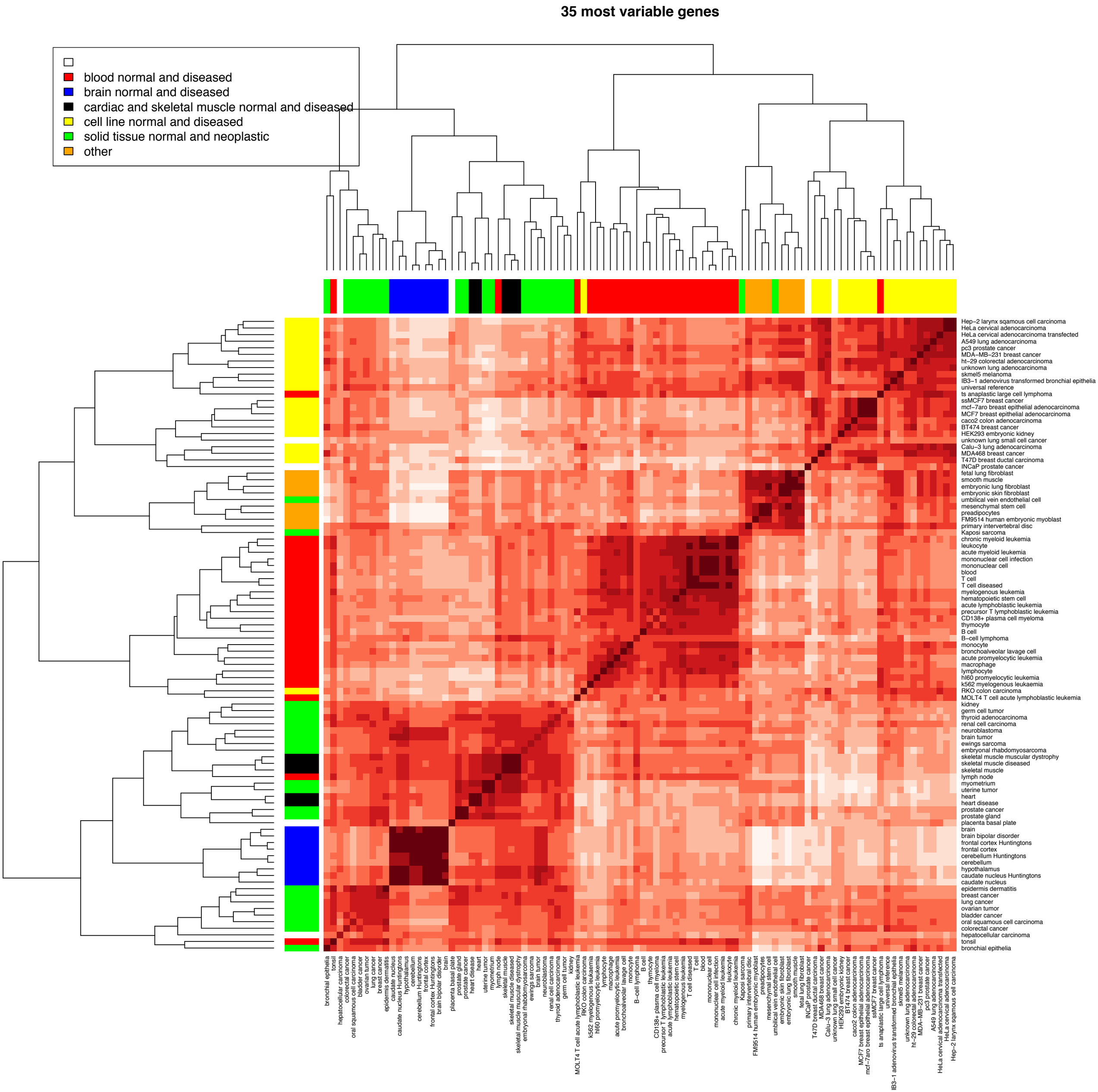

100 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
blood
leukocyte
mononuclear cell infection
mononuclear cell
chronic myeloid leukemia
T cell
T cell diseased
myelogenous leukemia
hematopoietic stem cell
acute myeloid leukemia
acute promyelocytic leukemia
monocyte
macrophage
bronchoalveolar lavage cell
B−cell lymphoma
ewings sarcoma
tonsil
precursor T lymphoblastic leukemia
acute lymphoblastic leukemia
thymocyte
B cell
CD138+ plasma cell myeloma
lymphocyte
hl60 promyelocytic leukemia
ts anaplastic large cell lymphoma
k562 myelogenous leukaemia
MOLT4 T cell acute lymphoblastic leukemia
HeLa cervical adenocarcinoma transfected
HeLa cervical adenocarcinoma
A549 lung adenocarcinoma
Hep−2 larynx sqamous cell carcinoma
MDA−MB−231 breast cancer
skmel5 melanoma
universal reference
IB3−1 adenovirus transformed bronchial epithelia
pc3 prostate cancer
unknown lung adenocarcinoma
RKO colon carcinoma
HEK293 embryonic kidney
ovarian tumor
bladder cancer
colorectal cancer
bronchial epithelia
lung cancer
breast cancer
oral squamous cell carcinoma
epidermis dermatitis
germ cell tumor
kidney
renal cell carcinoma
thyroid adenocarcinoma
hepatocellular carcinoma
prostate cancer
prostate gland
placenta basal plate
skeletal muscle diseased
skeletal muscle muscular dystrophy
skeletal muscle
heart
heart disease
lymph node
neuroblastoma
brain tumor
embryonal rhabdomyosarcoma
myometrium
uterine tumor
fetal lung fibroblast
smooth muscle
primary intervertebral disc
Kaposi sarcoma
umbilical vein endothelial cell
embryonic skin fibroblast
embryonic lung fibroblast
FM9514 human embryonic myoblast
mesenchymal stem cell
preadipocytes
frontal cortex Huntingtons
frontal cortex
brain bipolar disorder
cerebellum Huntingtons
cerebellum
caudate nucleus
brain
caudate nucleus Huntingtons
hypothalamus
ssMCF7 breast cancer
MCF7 breast epithelial adenocarcinoma
mcf−7aro breast epithelial adenocarcinoma
T47D breast ductal carcinoma
BT474 breast cancer
Calu−3 lung adenocarcinoma
MDA468 breast cancer
ht−29 colorectal adenocarcinoma
caco2 colon adenocarcinoma
unknown lung small cell cancer
INCaP prostate cancer

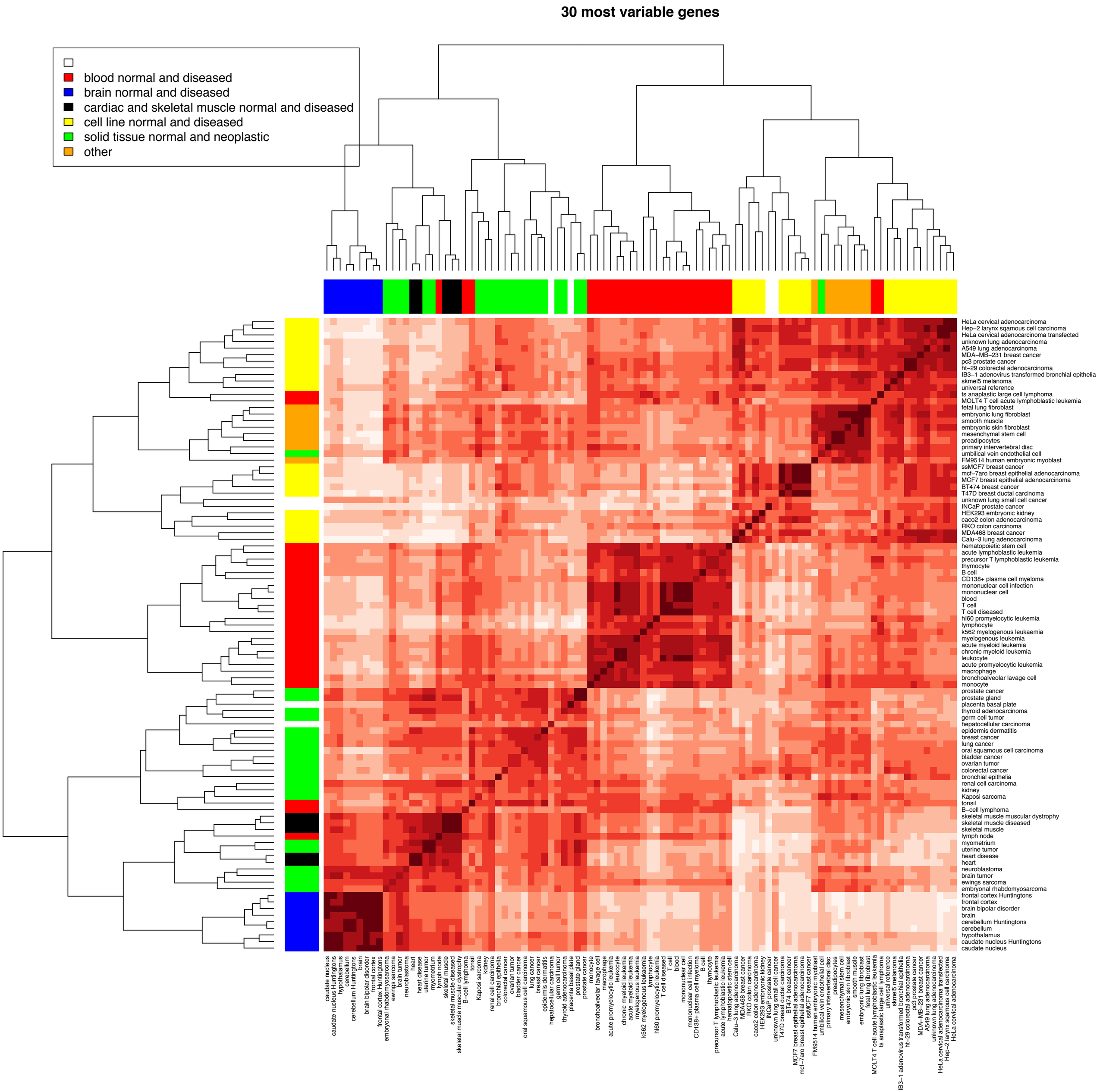
50 most variable genes

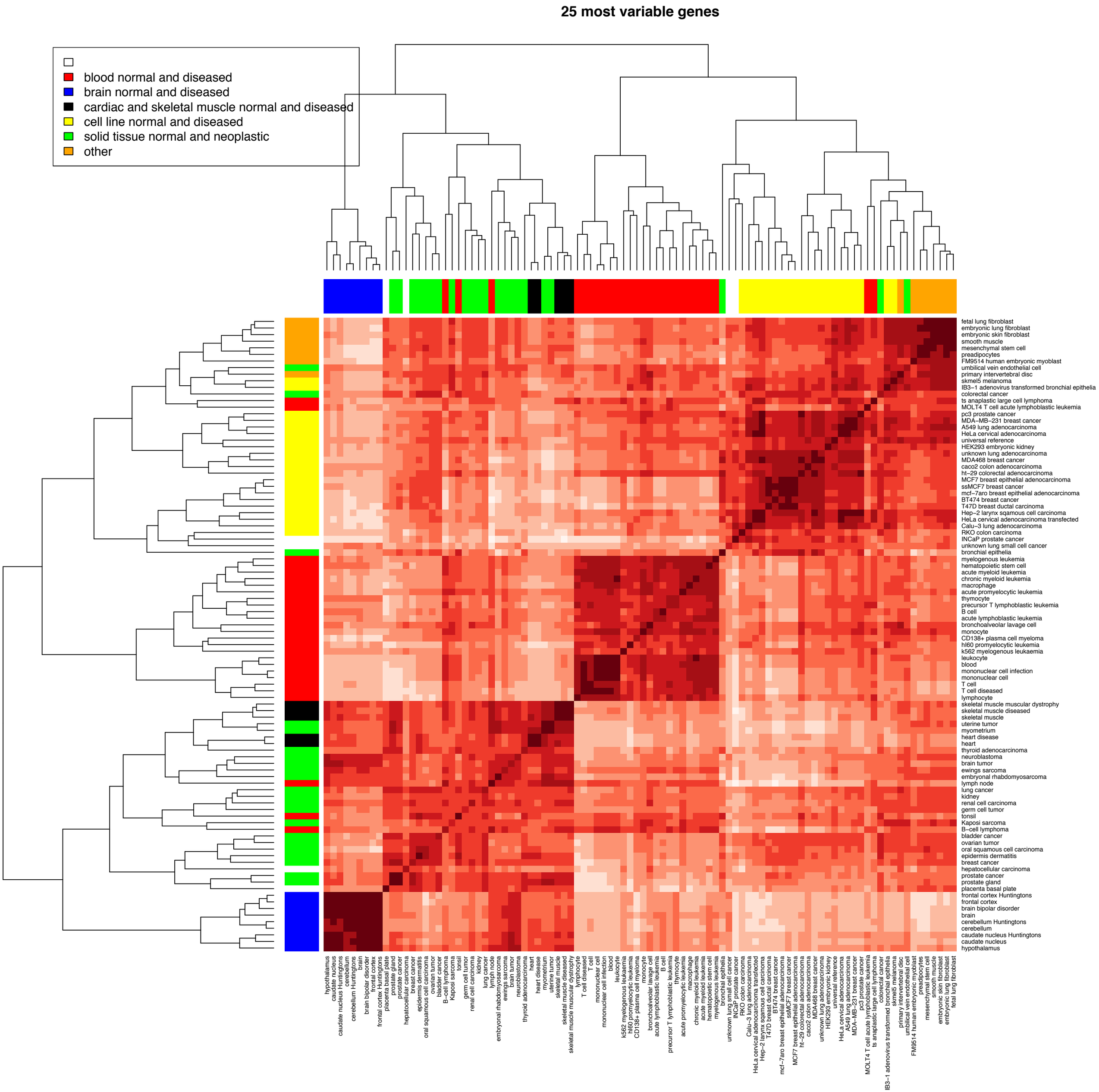


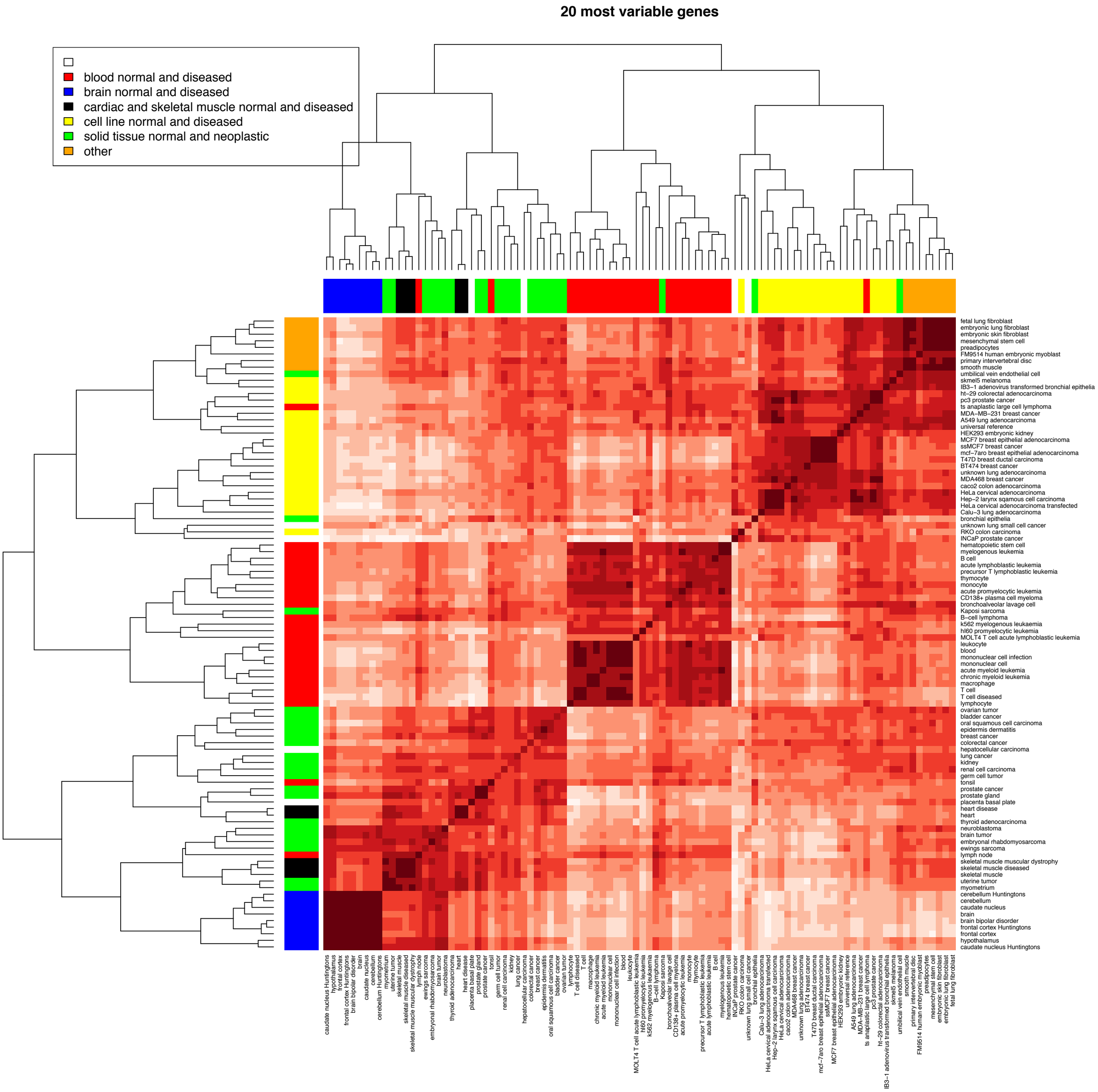

45 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
myelogenous leukemia
acute myeloid leukemia
chronic myeloid leukemia
hematopoietic stem cell
acute lymphoblastic leukemia
precursor T lymphoblastic leukemia
thymocyte
leukocyte
blood
mononuclear cell
mononuclear cell infection
T cell
T cell diseased
B cell
monocyte
CD138+ plasma cell myeloma
acute promyelocytic leukemia
macrophage
bronchoalveolar lavage cell
k562 myelogenous leukaemia
hl60 promyelocytic leukemia
lymphocyte
RKO colon carcinoma
MOLT4 T cell acute lymphoblastic leukemia
A549 lung adenocarcinoma
HeLa cervical adenocarcinoma transfected
pc3 prostate cancer
Hep−2 larynx sqamous cell carcinoma
HeLa cervical adenocarcinoma
MDA−MB−231 breast cancer
IB3−1 adenovirus transformed bronchial epithelia
skmel5 melanoma
universal reference
ts anaplastic large cell lymphoma
HEK293 embryonic kidney
embryonic skin fibroblast
embryonic lung fibroblast
mesenchymal stem cell
FM9514 human embryonic myoblast
fetal lung fibroblast
smooth muscle
umbilical vein endothelial cell
preadipocytes
primary intervertebral disc
Kaposi sarcoma
ssMCF7 breast cancer
MCF7 breast epithelial adenocarcinoma
mcf−7aro breast epithelial adenocarcinoma
caco2 colon adenocarcinoma
T47D breast ductal carcinoma
BT474 breast cancer
unknown lung small cell cancer
INCaP prostate cancer
ht−29 colorectal adenocarcinoma
unknown lung adenocarcinoma
Calu−3 lung adenocarcinoma
MDA468 breast cancer
skeletal muscle muscular dystrophy
skeletal muscle diseased
skeletal muscle
ewings sarcoma
embryonal rhabdomyosarcoma
neuroblastoma
brain tumor
heart disease
heart
myometrium
uterine tumor
lymph node
B−cell lymphoma
prostate cancer
prostate gland
placenta basal plate
germ cell tumor
kidney
renal cell carcinoma
hepatocellular carcinoma
thyroid adenocarcinoma
epidermis dermatitis
lung cancer
breast cancer
oral squamous cell carcinoma
ovarian tumor
bladder cancer
colorectal cancer
bronchial epithelia
tonsil
frontal cortex Huntingtons
frontal cortex
cerebellum Huntingtons
cerebellum
brain
brain bipolar disorder
caudate nucleus
caudate nucleus Huntingtons
hypothalamus

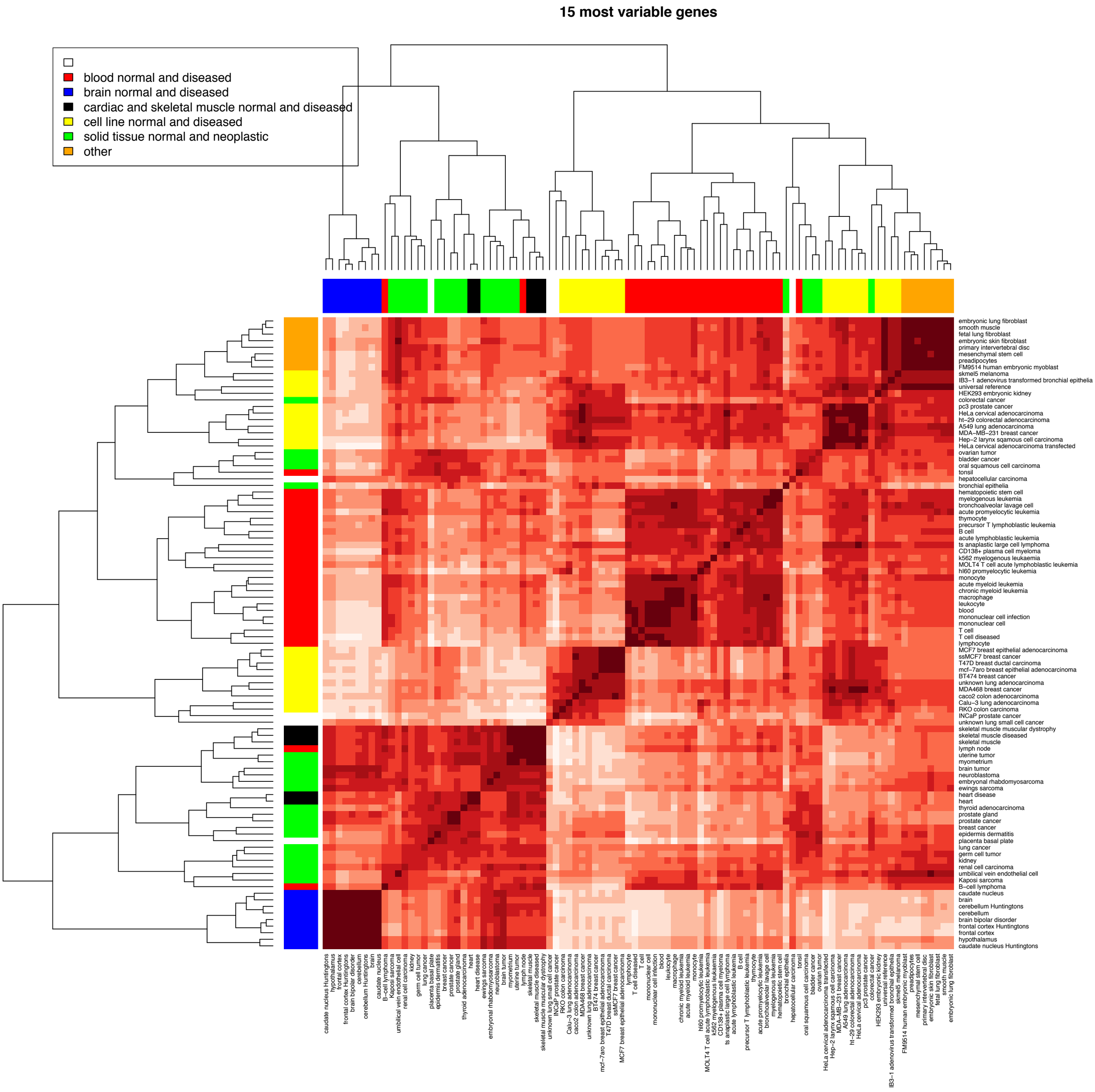

40 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
leukocyte
blood
chronic myeloid leukemia
mononuclear cell infection
mononuclear cell
T cell diseased
T cell
hematopoietic stem cell
acute lymphoblastic leukemia
precursor T lymphoblastic leukemia
thymocyte
B cell
CD138+ plasma cell myeloma
B−cell lymphoma
myelogenous leukemia
acute myeloid leukemia
macrophage
acute promyelocytic leukemia
monocyte
bronchoalveolar lavage cell
k562 myelogenous leukaemia
hl60 promyelocytic leukemia
lymphocyte
RKO colon carcinoma
MOLT4 T cell acute lymphoblastic leukemia
epidermis dermatitis
lung cancer
breast cancer
ovarian tumor
bladder cancer
oral squamous cell carcinoma
placenta basal plate
colorectal cancer
bronchial epithelia
germ cell tumor
kidney
renal cell carcinoma
tonsil
Hep−2 larynx sqamous cell carcinoma
HeLa cervical adenocarcinoma
MDA−MB−231 breast cancer
A549 lung adenocarcinoma
HeLa cervical adenocarcinoma transfected
pc3 prostate cancer
skmel5 melanoma
IB3−1 adenovirus transformed bronchial epithelia
ts anaplastic large cell lymphoma
universal reference
HEK293 embryonic kidney
unknown lung adenocarcinoma
MDA468 breast cancer
Calu−3 lung adenocarcinoma
ht−29 colorectal adenocarcinoma
fetal lung fibroblast
smooth muscle
embryonic skin fibroblast
embryonic lung fibroblast
mesenchymal stem cell
FM9514 human embryonic myoblast
preadipocytes
umbilical vein endothelial cell
primary intervertebral disc
Kaposi sarcoma
ssMCF7 breast cancer
MCF7 breast epithelial adenocarcinoma
mcf−7aro breast epithelial adenocarcinoma
caco2 colon adenocarcinoma
BT474 breast cancer
T47D breast ductal carcinoma
unknown lung small cell cancer
INCaP prostate cancer
skeletal muscle muscular dystrophy
skeletal muscle diseased
skeletal muscle
lymph node
ewings sarcoma
embryonal rhabdomyosarcoma
neuroblastoma
brain tumor
heart
heart disease
uterine tumor
myometrium
thyroid adenocarcinoma
hepatocellular carcinoma
prostate gland
prostate cancer
frontal cortex Huntingtons
frontal cortex
cerebellum Huntingtons
cerebellum
brain
brain bipolar disorder
caudate nucleus
hypothalamus
caudate nucleus Huntingtons
caudate nucleus Huntingtons
hypothalamus
caudate nucleus
brain bipolar disorder
brain
cerebellum
cerebellum Huntingtons
frontal cortex Huntingtons
prostate cancer
prostate gland
hepatocellular carcinoma
thyroid adenocarcinoma
myometrium
uterine tumor
heart disease
heart
brain tumor
neuroblastoma
embryonal rhabdomyosarcoma
ewings sarcoma
lymph node
skeletal muscle
skeletal muscle diseased
skeletal muscle muscular dystrophy
INCaP prostate cancer
unknown lung small cell cancer
T47D breast ductal carcinoma
BT474 breast cancer
caco2 colon adenocarcinoma
mcf−7aro breast epithelial adenocarcinoma
MCF7 breast epithelial adenocarcinoma
ssMCF7 breast cancer
Kaposi sarcoma
primary intervertebral disc
umbilical vein endothelial cell
preadipocytes
FM9514 human embryonic myoblast
mesenchymal stem cell
embryonic lung fibroblast
embryonic skin fibroblast
smooth muscle
fetal lung fibroblast
ht−29 colorectal adenocarcinoma
Calu−3 lung adenocarcinoma
MDA468 breast cancer
unknown lung adenocarcinoma
HEK293 embryonic kidney
universal reference
ts anaplastic large cell lymphoma
IB3−1 adenovirus transformed bronchial epithelia
skmel5 melanoma
pc3 prostate cancer
HeLa cervical adenocarcinoma transfected
A549 lung adenocarcinoma
MDA−MB−231 breast cancer
HeLa cervical adenocarcinoma
Hep−2 larynx sqamous cell carcinoma
tonsil
renal cell carcinoma
kidney
germ cell tumor
bronchial epithelia
colorectal cancer
placenta basal plate
oral squamous cell carcinoma
bladder cancer
ovarian tumor
breast cancer
lung cancer
epidermis dermatitis
MOLT4 T cell acute lymphoblastic leukemia
RKO colon carcinoma
lymphocyte
hl60 promyelocytic leukemia
k562 myelogenous leukaemia
bronchoalveolar lavage cell
monocyte
acute promyelocytic leukemia
macrophage
acute myeloid leukemia
myelogenous leukemia
B−cell lymphoma
CD138+ plasma cell myeloma
B cell
thymocyte
precursor T lymphoblastic leukemia
acute lymphoblastic leukemia
hematopoietic stem cell
T cell
T cell diseased
mononuclear cell
mononuclear cell infection
chronic myeloid leukemia
blood
leukocyte

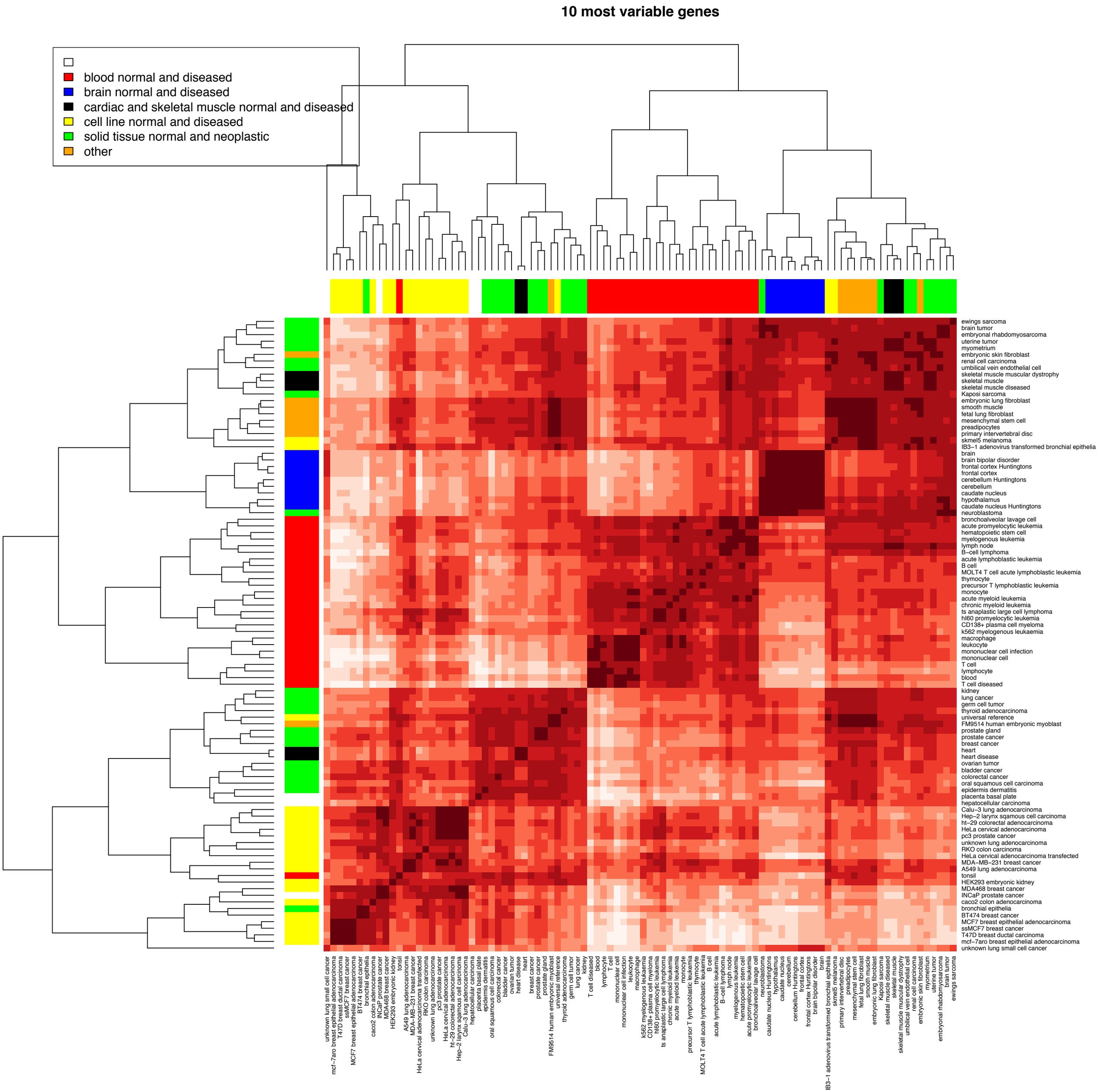

35 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
Hep−2 larynx sqamous cell carcinoma
HeLa cervical adenocarcinoma
HeLa cervical adenocarcinoma transfected
A549 lung adenocarcinoma
pc3 prostate cancer
MDA−MB−231 breast cancer
ht−29 colorectal adenocarcinoma
unknown lung adenocarcinoma
skmel5 melanoma
IB3−1 adenovirus transformed bronchial epithelia
universal reference
ts anaplastic large cell lymphoma
ssMCF7 breast cancer
mcf−7aro breast epithelial adenocarcinoma
MCF7 breast epithelial adenocarcinoma
caco2 colon adenocarcinoma
BT474 breast cancer
HEK293 embryonic kidney
unknown lung small cell cancer
Calu−3 lung adenocarcinoma
MDA468 breast cancer
T47D breast ductal carcinoma
INCaP prostate cancer
fetal lung fibroblast
smooth muscle
embryonic lung fibroblast
embryonic skin fibroblast
umbilical vein endothelial cell
mesenchymal stem cell
preadipocytes
FM9514 human embryonic myoblast
primary intervertebral disc
Kaposi sarcoma
chronic myeloid leukemia
leukocyte
acute myeloid leukemia
mononuclear cell infection
mononuclear cell
blood
T cell
T cell diseased
myelogenous leukemia
hematopoietic stem cell
acute lymphoblastic leukemia
precursor T lymphoblastic leukemia
CD138+ plasma cell myeloma
thymocyte
B cell
B−cell lymphoma
monocyte
bronchoalveolar lavage cell
acute promyelocytic leukemia
macrophage
lymphocyte
hl60 promyelocytic leukemia
k562 myelogenous leukaemia
RKO colon carcinoma
MOLT4 T cell acute lymphoblastic leukemia
kidney
germ cell tumor
thyroid adenocarcinoma
renal cell carcinoma
neuroblastoma
brain tumor
ewings sarcoma
embryonal rhabdomyosarcoma
skeletal muscle muscular dystrophy
skeletal muscle diseased
skeletal muscle
lymph node
myometrium
uterine tumor
heart
heart disease
prostate cancer
prostate gland
placenta basal plate
brain
brain bipolar disorder
frontal cortex Huntingtons
frontal cortex
cerebellum Huntingtons
cerebellum
hypothalamus
caudate nucleus Huntingtons
caudate nucleus
epidermis dermatitis
breast cancer
lung cancer
ovarian tumor
bladder cancer
oral squamous cell carcinoma
colorectal cancer
hepatocellular carcinoma
tonsil
bronchial epithelia


30 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
HeLa cervical adenocarcinoma
Hep−2 larynx sqamous cell carcinoma
HeLa cervical adenocarcinoma transfected
unknown lung adenocarcinoma
A549 lung adenocarcinoma
MDA−MB−231 breast cancer
pc3 prostate cancer
ht−29 colorectal adenocarcinoma
IB3−1 adenovirus transformed bronchial epithelia
skmel5 melanoma
universal reference
ts anaplastic large cell lymphoma
MOLT4 T cell acute lymphoblastic leukemia
fetal lung fibroblast
embryonic lung fibroblast
smooth muscle
embryonic skin fibroblast
mesenchymal stem cell
preadipocytes
primary intervertebral disc
umbilical vein endothelial cell
FM9514 human embryonic myoblast
ssMCF7 breast cancer
mcf−7aro breast epithelial adenocarcinoma
MCF7 breast epithelial adenocarcinoma
BT474 breast cancer
T47D breast ductal carcinoma
unknown lung small cell cancer
INCaP prostate cancer
HEK293 embryonic kidney
caco2 colon adenocarcinoma
RKO colon carcinoma
MDA468 breast cancer
Calu−3 lung adenocarcinoma
hematopoietic stem cell
acute lymphoblastic leukemia
precursor T lymphoblastic leukemia
thymocyte
B cell
CD138+ plasma cell myeloma
mononuclear cell infection
mononuclear cell
blood
T cell
T cell diseased
hl60 promyelocytic leukemia
lymphocyte
k562 myelogenous leukaemia
myelogenous leukemia
acute myeloid leukemia
chronic myeloid leukemia
leukocyte
acute promyelocytic leukemia
macrophage
bronchoalveolar lavage cell
monocyte
prostate cancer
prostate gland
placenta basal plate
thyroid adenocarcinoma
germ cell tumor
hepatocellular carcinoma
epidermis dermatitis
breast cancer
lung cancer
oral squamous cell carcinoma
bladder cancer
ovarian tumor
colorectal cancer
bronchial epithelia
renal cell carcinoma
kidney
Kaposi sarcoma
tonsil
B−cell lymphoma
skeletal muscle muscular dystrophy
skeletal muscle diseased
skeletal muscle
lymph node
myometrium
uterine tumor
heart disease
heart
neuroblastoma
brain tumor
ewings sarcoma
embryonal rhabdomyosarcoma
frontal cortex Huntingtons
frontal cortex
brain bipolar disorder
brain
cerebellum Huntingtons
cerebellum
hypothalamus
caudate nucleus Huntingtons
caudate nucleus

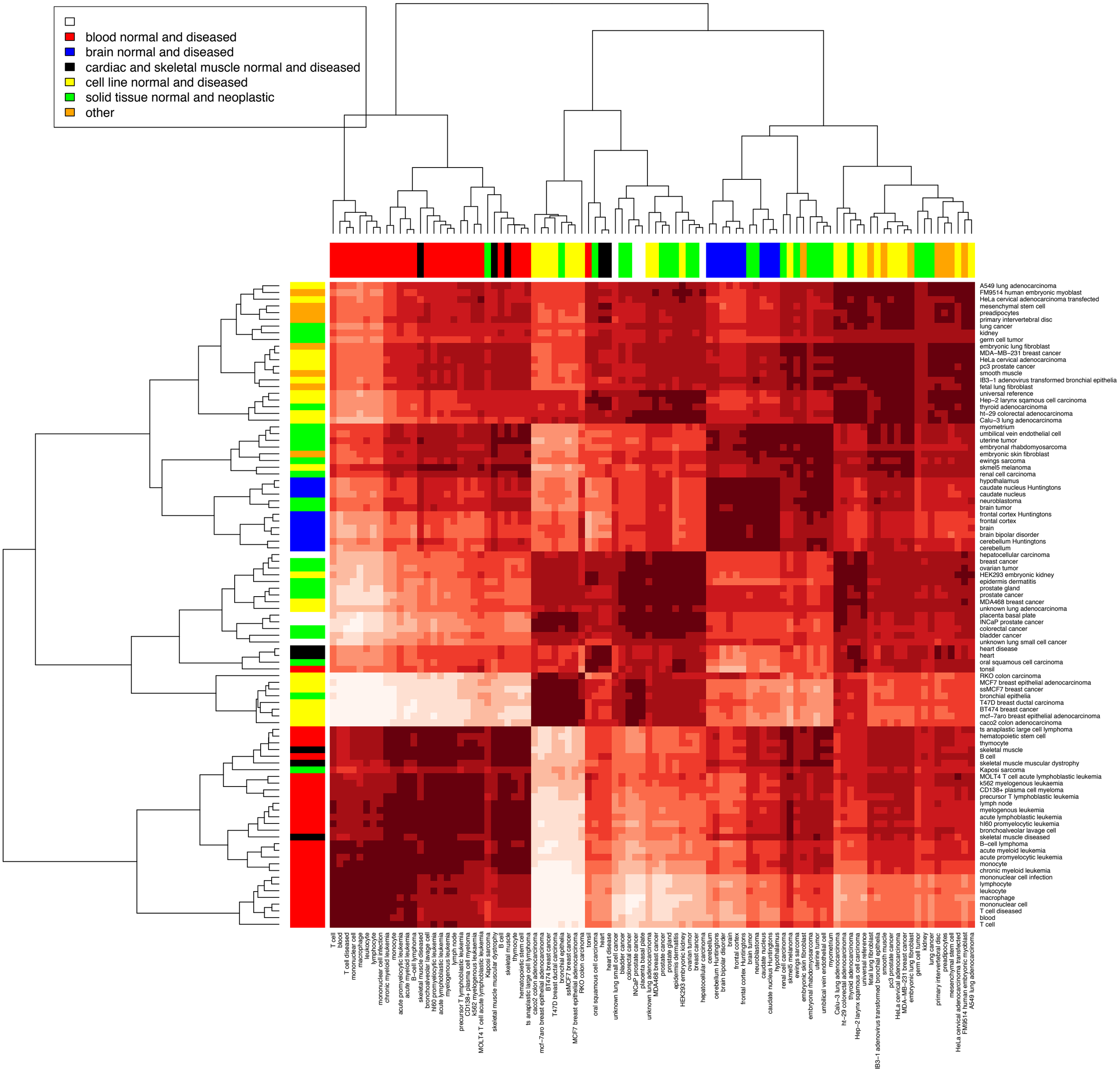

25 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
fetal lung fibroblast
embryonic lung fibroblast
embryonic skin fibroblast
smooth muscle
mesenchymal stem cell
preadipocytes
FM9514 human embryonic myoblast
umbilical vein endothelial cell
primary intervertebral disc
skmel5 melanoma
IB3−1 adenovirus transformed bronchial epithelia
colorectal cancer
ts anaplastic large cell lymphoma
MOLT4 T cell acute lymphoblastic leukemia
pc3 prostate cancer
MDA−MB−231 breast cancer
A549 lung adenocarcinoma
HeLa cervical adenocarcinoma
universal reference
HEK293 embryonic kidney
unknown lung adenocarcinoma
MDA468 breast cancer
caco2 colon adenocarcinoma
ht−29 colorectal adenocarcinoma
MCF7 breast epithelial adenocarcinoma
ssMCF7 breast cancer
mcf−7aro breast epithelial adenocarcinoma
BT474 breast cancer
T47D breast ductal carcinoma
Hep−2 larynx sqamous cell carcinoma
HeLa cervical adenocarcinoma transfected
Calu−3 lung adenocarcinoma
RKO colon carcinoma
INCaP prostate cancer
unknown lung small cell cancer
bronchial epithelia
myelogenous leukemia
hematopoietic stem cell
acute myeloid leukemia
chronic myeloid leukemia
macrophage
acute promyelocytic leukemia
thymocyte
precursor T lymphoblastic leukemia
B cell
acute lymphoblastic leukemia
bronchoalveolar lavage cell
monocyte
CD138+ plasma cell myeloma
hl60 promyelocytic leukemia
k562 myelogenous leukaemia
leukocyte
blood
mononuclear cell infection
mononuclear cell
T cell
T cell diseased
lymphocyte
skeletal muscle muscular dystrophy
skeletal muscle diseased
skeletal muscle
uterine tumor
myometrium
heart disease
heart
thyroid adenocarcinoma
neuroblastoma
brain tumor
ewings sarcoma
embryonal rhabdomyosarcoma
lymph node
lung cancer
kidney
renal cell carcinoma
germ cell tumor
tonsil
Kaposi sarcoma
B−cell lymphoma
bladder cancer
ovarian tumor
oral squamous cell carcinoma
epidermis dermatitis
breast cancer
hepatocellular carcinoma
prostate cancer
prostate gland
placenta basal plate
frontal cortex Huntingtons
frontal cortex
brain bipolar disorder
brain
cerebellum Huntingtons
cerebellum
caudate nucleus Huntingtons
caudate nucleus
hypothalamus


20 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
fetal lung fibroblast
embryonic lung fibroblast
embryonic skin fibroblast
mesenchymal stem cell
preadipocytes
FM9514 human embryonic myoblast
primary intervertebral disc
smooth muscle
umbilical vein endothelial cell
skmel5 melanoma
IB3−1 adenovirus transformed bronchial epithelia
ht−29 colorectal adenocarcinoma
pc3 prostate cancer
ts anaplastic large cell lymphoma
MDA−MB−231 breast cancer
A549 lung adenocarcinoma
universal reference
HEK293 embryonic kidney
MCF7 breast epithelial adenocarcinoma
ssMCF7 breast cancer
mcf−7aro breast epithelial adenocarcinoma
T47D breast ductal carcinoma
BT474 breast cancer
unknown lung adenocarcinoma
MDA468 breast cancer
caco2 colon adenocarcinoma
HeLa cervical adenocarcinoma
Hep−2 larynx sqamous cell carcinoma
HeLa cervical adenocarcinoma transfected
Calu−3 lung adenocarcinoma
bronchial epithelia
unknown lung small cell cancer
RKO colon carcinoma
INCaP prostate cancer
hematopoietic stem cell
myelogenous leukemia
B cell
acute lymphoblastic leukemia
precursor T lymphoblastic leukemia
thymocyte
monocyte
acute promyelocytic leukemia
CD138+ plasma cell myeloma
bronchoalveolar lavage cell
Kaposi sarcoma
B−cell lymphoma
k562 myelogenous leukaemia
hl60 promyelocytic leukemia
MOLT4 T cell acute lymphoblastic leukemia
leukocyte
blood
mononuclear cell infection
mononuclear cell
acute myeloid leukemia
chronic myeloid leukemia
macrophage
T cell
T cell diseased
lymphocyte
ovarian tumor
bladder cancer
oral squamous cell carcinoma
epidermis dermatitis
breast cancer
colorectal cancer
hepatocellular carcinoma
lung cancer
kidney
renal cell carcinoma
germ cell tumor
tonsil
prostate cancer
prostate gland
placenta basal plate
heart disease
heart
thyroid adenocarcinoma
neuroblastoma
brain tumor
embryonal rhabdomyosarcoma
ewings sarcoma
lymph node
skeletal muscle muscular dystrophy
skeletal muscle diseased
skeletal muscle
uterine tumor
myometrium
cerebellum Huntingtons
cerebellum
caudate nucleus
brain
brain bipolar disorder
frontal cortex Huntingtons
frontal cortex
hypothalamus
caudate nucleus Huntingtons


15 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
embryonic lung fibroblast
smooth muscle
fetal lung fibroblast
embryonic skin fibroblast
primary intervertebral disc
mesenchymal stem cell
preadipocytes
FM9514 human embryonic myoblast
skmel5 melanoma
IB3−1 adenovirus transformed bronchial epithelia
universal reference
HEK293 embryonic kidney
colorectal cancer
pc3 prostate cancer
HeLa cervical adenocarcinoma
ht−29 colorectal adenocarcinoma
A549 lung adenocarcinoma
MDA−MB−231 breast cancer
Hep−2 larynx sqamous cell carcinoma
HeLa cervical adenocarcinoma transfected
ovarian tumor
bladder cancer
oral squamous cell carcinoma
tonsil
hepatocellular carcinoma
bronchial epithelia
hematopoietic stem cell
myelogenous leukemia
bronchoalveolar lavage cell
acute promyelocytic leukemia
thymocyte
precursor T lymphoblastic leukemia
B cell
acute lymphoblastic leukemia
ts anaplastic large cell lymphoma
CD138+ plasma cell myeloma
k562 myelogenous leukaemia
MOLT4 T cell acute lymphoblastic leukemia
hl60 promyelocytic leukemia
monocyte
acute myeloid leukemia
chronic myeloid leukemia
macrophage
leukocyte
blood
mononuclear cell infection
mononuclear cell
T cell
T cell diseased
lymphocyte
MCF7 breast epithelial adenocarcinoma
ssMCF7 breast cancer
T47D breast ductal carcinoma
mcf−7aro breast epithelial adenocarcinoma
BT474 breast cancer
unknown lung adenocarcinoma
MDA468 breast cancer
caco2 colon adenocarcinoma
Calu−3 lung adenocarcinoma
RKO colon carcinoma
INCaP prostate cancer
unknown lung small cell cancer
skeletal muscle muscular dystrophy
skeletal muscle diseased
skeletal muscle
lymph node
uterine tumor
myometrium
brain tumor
neuroblastoma
embryonal rhabdomyosarcoma
ewings sarcoma
heart disease
heart
thyroid adenocarcinoma
prostate gland
prostate cancer
breast cancer
epidermis dermatitis
placenta basal plate
lung cancer
germ cell tumor
kidney
renal cell carcinoma
umbilical vein endothelial cell
Kaposi sarcoma
B−cell lymphoma
caudate nucleus
brain
cerebellum Huntingtons
cerebellum
brain bipolar disorder
frontal cortex Huntingtons
frontal cortex
hypothalamus
caudate nucleus Huntingtons

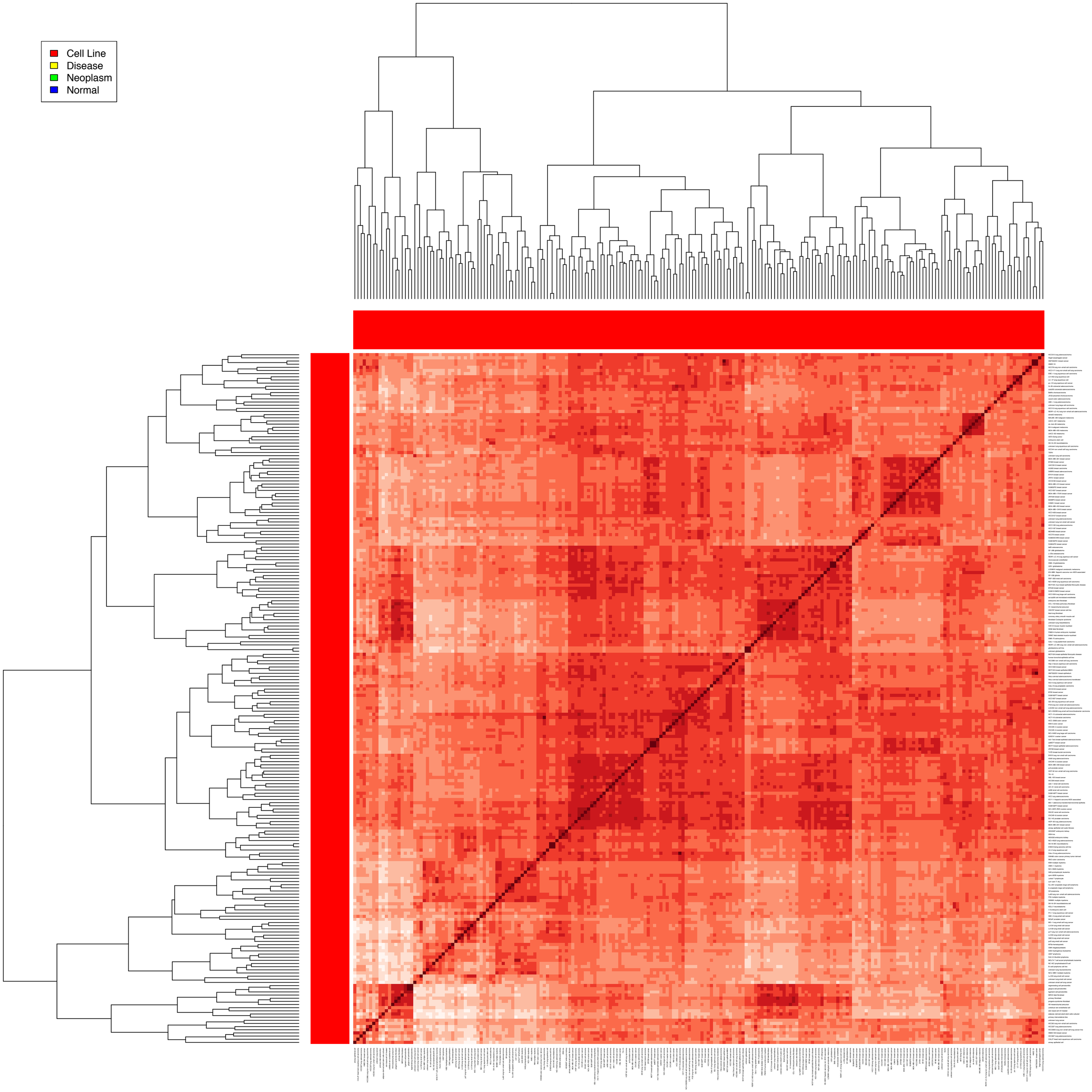

10 most variable genes
blood normal and diseased
brain normal and diseased
cardiac and skeletal muscle normal and diseased
cell line normal and diseased
solid tissue normal and neoplastic
other
ewings sarcoma
brain tumor
embryonal rhabdomyosarcoma
uterine tumor
myometrium
embryonic skin fibroblast
renal cell carcinoma
umbilical vein endothelial cell
skeletal muscle muscular dystrophy
skeletal muscle
skeletal muscle diseased
Kaposi sarcoma
embryonic lung fibroblast
smooth muscle
fetal lung fibroblast
mesenchymal stem cell
preadipocytes
primary intervertebral disc
skmel5 melanoma
IB3−1 adenovirus transformed bronchial epithelia
brain
brain bipolar disorder
frontal cortex Huntingtons
frontal cortex
cerebellum Huntingtons
cerebellum
caudate nucleus
hypothalamus
caudate nucleus Huntingtons
neuroblastoma
bronchoalveolar lavage cell
acute promyelocytic leukemia
hematopoietic stem cell
myelogenous leukemia
lymph node
B−cell lymphoma
acute lymphoblastic leukemia
B cell
MOLT4 T cell acute lymphoblastic leukemia
thymocyte
precursor T lymphoblastic leukemia
monocyte
acute myeloid leukemia
chronic myeloid leukemia
ts anaplastic large cell lymphoma
hl60 promyelocytic leukemia
CD138+ plasma cell myeloma
k562 myelogenous leukaemia
macrophage
leukocyte
mononuclear cell infection
mononuclear cell
T cell
lymphocyte
blood
T cell diseased
kidney
lung cancer
germ cell tumor
thyroid adenocarcinoma
universal reference
FM9514 human embryonic myoblast
prostate gland
prostate cancer
breast cancer
heart
heart disease
ovarian tumor
bladder cancer
colorectal cancer
oral squamous cell carcinoma
epidermis dermatitis
placenta basal plate
hepatocellular carcinoma
Calu−3 lung adenocarcinoma
Hep−2 larynx sqamous cell carcinoma
ht−29 colorectal adenocarcinoma
HeLa cervical adenocarcinoma
pc3 prostate cancer
unknown lung adenocarcinoma
RKO colon carcinoma
HeLa cervical adenocarcinoma transfected
MDA−MB−231 breast cancer
A549 lung adenocarcinoma
tonsil
HEK293 embryonic kidney
MDA468 breast cancer
INCaP prostate cancer
caco2 colon adenocarcinoma
bronchial epithelia
BT474 breast cancer
MCF7 breast epithelial adenocarcinoma
ssMCF7 breast cancer
T47D breast ductal carcinoma
mcf−7aro breast epithelial adenocarcinoma
unknown lung small cell cancer

5 most variable genes



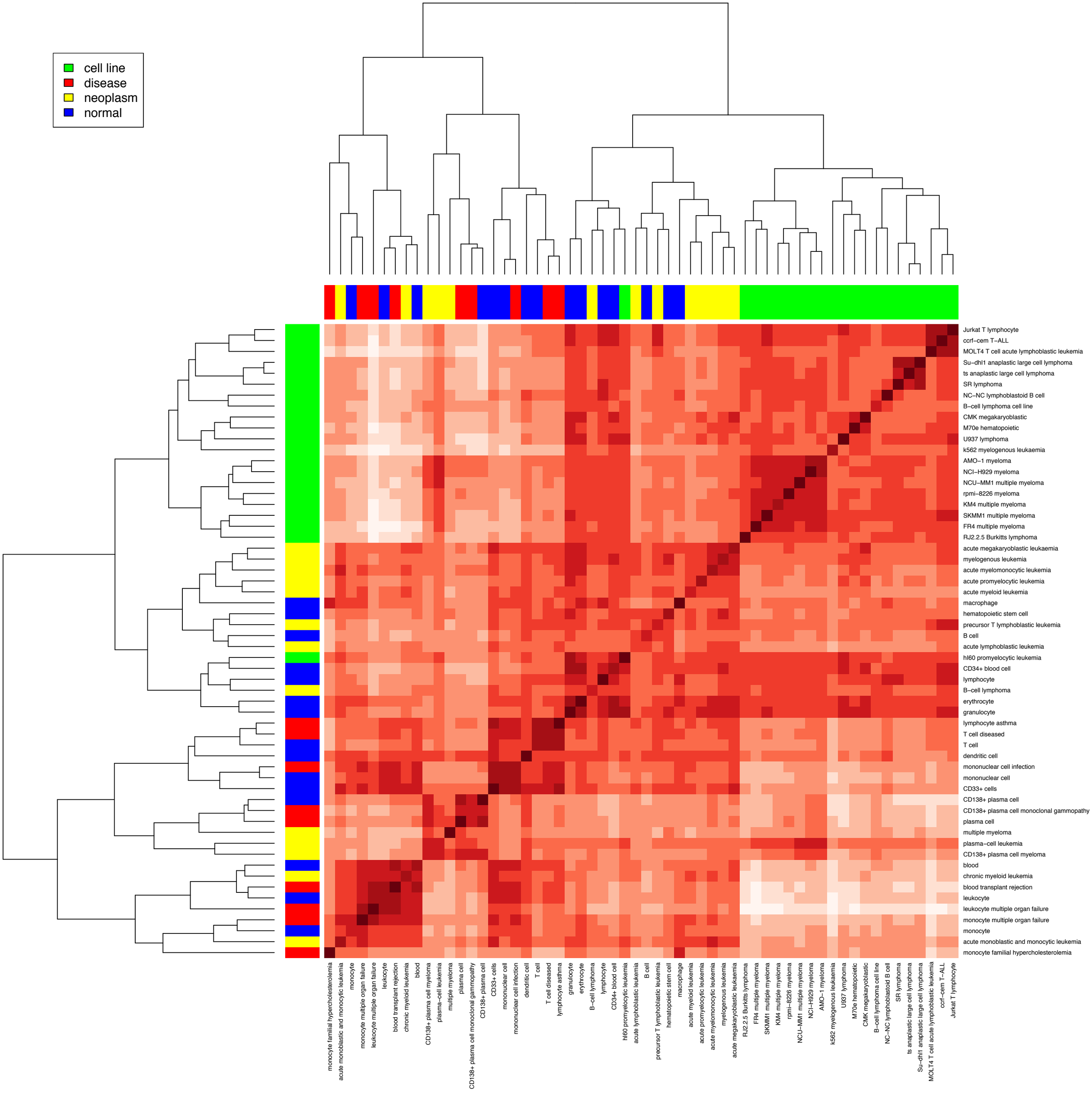
Supplementary Figure 4e:

Sample clustering based on 1000, 950, 900, …, 50, 45, … ,5 most variable probesets in the 96 biological groups. The four colors at the left and top of the heat map depict 6 clusters identified on self-vs-self heat-map of 96 biological groups in supplementary Figure 4a.


Cell Line
Disease
Neoplasm
Normal

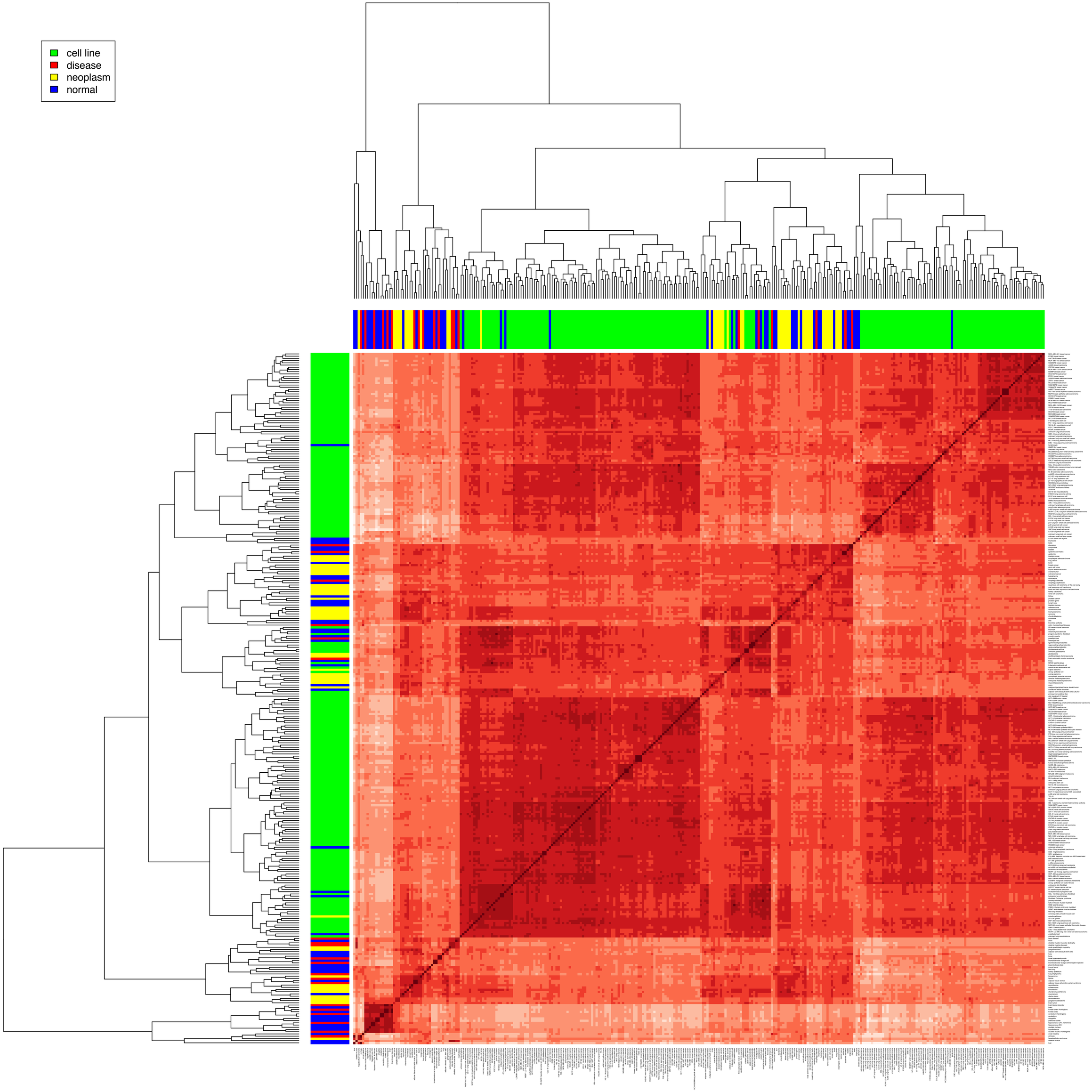
Supplementary Figure 4f:

Heat map comparison of biological groups of cell lines.


cell line
disease
neoplasm
normal
Jurkat T lymphocyte
ccrf−cem T−ALL
MOLT4 T cell acute lymphoblastic leukemia
Su−dhl1 anaplastic large cell lymphoma
ts anaplastic large cell lymphoma
SR lymphoma
NC−NC lymphoblastoid B cell
B−cell lymphoma cell line
CMK megakaryoblastic
M70e hematopoietic
U937 lymphoma
k562 myelogenous leukaemia
AMO−1 myeloma
NCI−H929 myeloma
NCU−MM1 multiple myeloma
rpmi−8226 myeloma
KM4 multiple myeloma
SKMM1 multiple myeloma
FR4 multiple myeloma
RJ2.2.5 Burkitts lymphoma
acute megakaryoblastic leukaemia
myelogenous leukemia
acute myelomonocytic leukemia
acute promyelocytic leukemia
acute myeloid leukemia
macrophage
hematopoietic stem cell
precursor T lymphoblastic leukemia
B cell
acute lymphoblastic leukemia
hl60 promyelocytic leukemia
CD34+ blood cell
lymphocyte
B−cell lymphoma
erythrocyte
granulocyte
lymphocyte asthma
T cell diseased
T cell
dendritic cell
mononuclear cell infection
mononuclear cell
CD33+ cells
CD138+ plasma cell
CD138+ plasma cell monoclonal gammopathy
plasma cell
multiple myeloma
plasma−cell leukemia
CD138+ plasma cell myeloma
blood
chronic myeloid leukemia
blood transplant rejection
leukocyte
leukocyte multiple organ failure
monocyte multiple organ failure
monocyte
acute monoblastic and monocytic leukemia
monocyte familial hypercholesterolemia
monocyte familial hypercholesterolemia
acute monoblastic and monocytic leukemia
monocyte
monocyte multiple organ failure
leukocyte multiple organ failure
leukocyte
blood transplant rejection
chronic myeloid leukemia
blood
CD138+ plasma cell myeloma
plasma−cell leukemia
multiple myeloma
plasma cell
CD138+ plasma cell monoclonal gammopathy
CD138+ plasma cell
CD33+ cells
mononuclear cell
mononuclear cell infection
dendritic cell
T cell
T cell diseased
lymphocyte asthma
granulocyte
erythrocyte
B−cell lymphoma
lymphocyte
CD34+ blood cell
hl60 promyelocytic leukemia
acute lymphoblastic leukemia
B cell
precursor T lymphoblastic leukemia
hematopoietic stem cell
macrophage
acute myeloid leukemia
acute promyelocytic leukemia
acute myelomonocytic leukemia
myelogenous leukemia
acute megakaryoblastic leukaemia
RJ2.2.5 Burkitts lymphoma
FR4 multiple myeloma
SKMM1 multiple myeloma
KM4 multiple myeloma
rpmi−8226 myeloma
NCU−MM1 multiple myeloma
NCI−H929 myeloma
AMO−1 myeloma
k562 myelogenous leukaemia
U937 lymphoma
M70e hematopoietic
CMK megakaryoblastic
B−cell lymphoma cell line
NC−NC lymphoblastoid B cell
SR lymphoma
ts anaplastic large cell lymphoma
Su−dhl1 anaplastic large cell lymphoma
MOLT4 T cell acute lymphoblastic leukemia
ccrf−cem T−ALL
Jurkat T lymphocyte

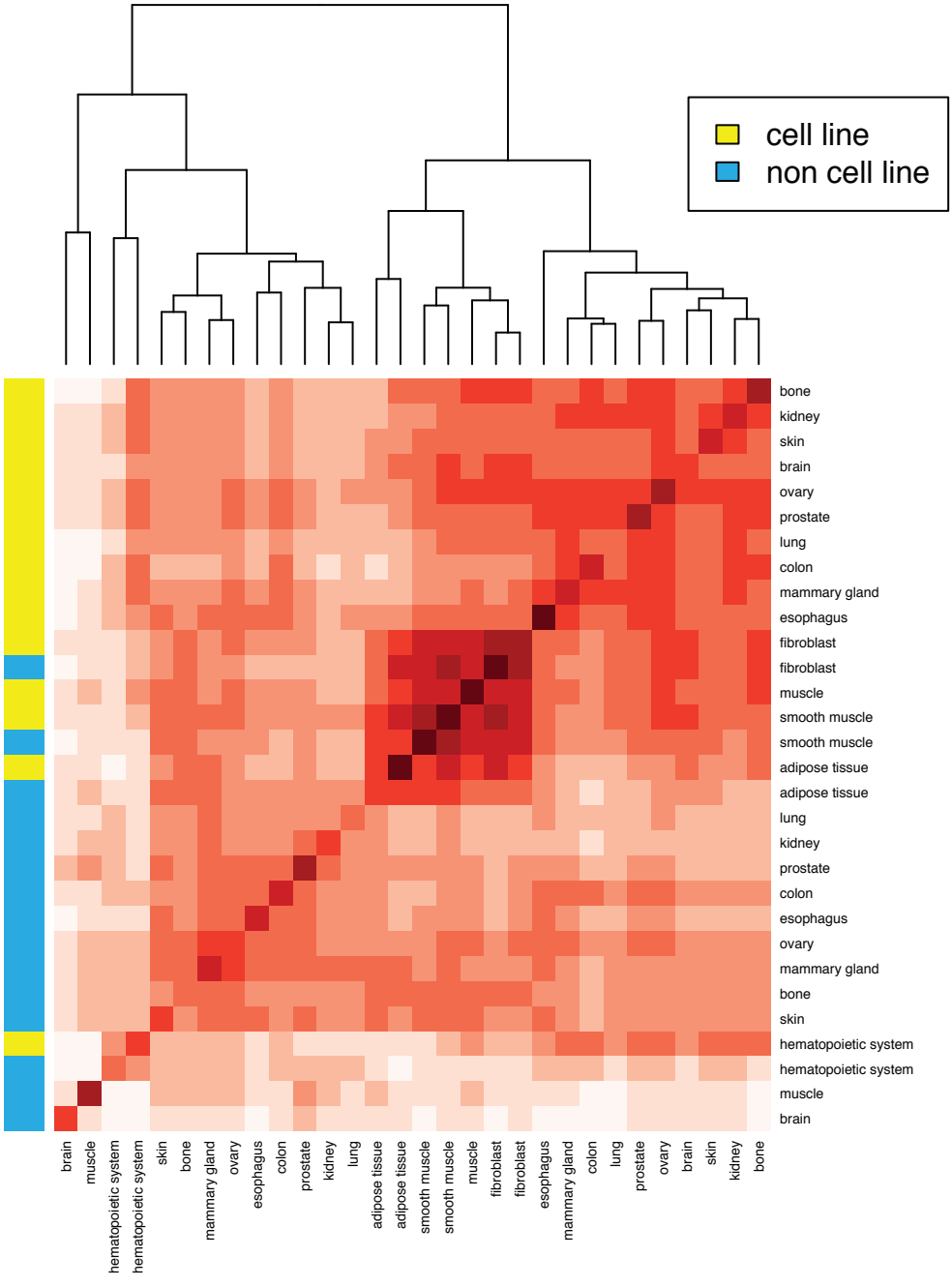
Supplementary Figure 4g:

Heat map comparison of biological groups of hematopoietic system.


cell line
disease
neoplasm
normal

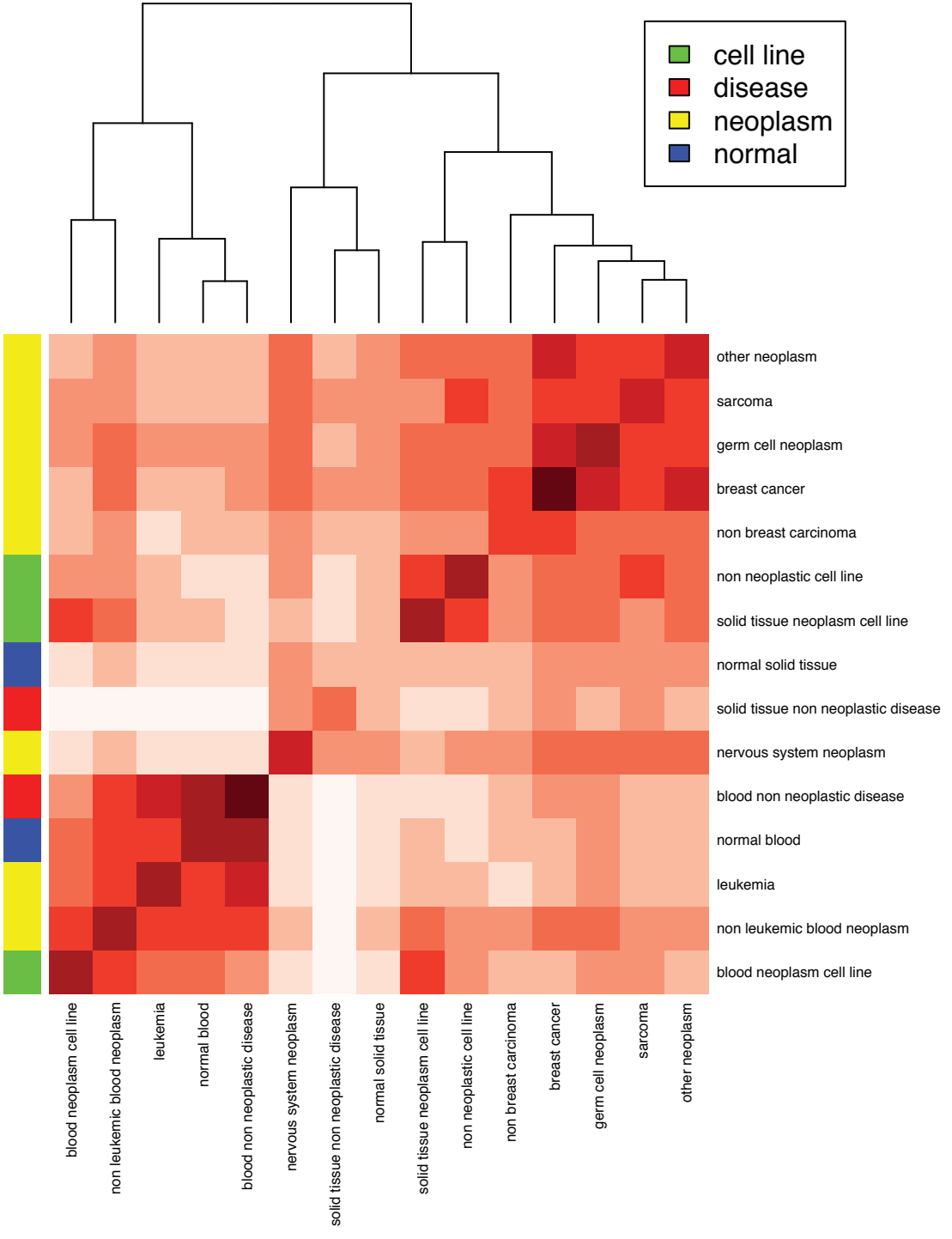
Supplementary Figure 4h:

Heat map comparison of biological groups of non-hematopoietic system.


cell line
non cell line
bone
kidney
skin
brain
ovary
prostate
lung
colon
mammary gland
esophagus
fibroblast
fibroblast
muscle
smooth muscle
smooth muscle
adipose tissue
adipose tissue
lung
kidney
prostate
colon
esophagus
ovary
mammary gland
bone
skin
hematopoietic system
hematopoietic system
muscle
brain
brain
muscle
hematopoietic system
hematopoietic system
skin
bone
mammary gland
ovary
esophagus
colon
prostate
kidney
lung
adipose tissue
adipose tissue
smooth muscle
smooth muscle
muscle
fibroblast
fibroblast
esophagus
mammary gland
colon
lung
prostate
ovary
brain
skin
kidney
bone

Supplementary Figure 4i:

15 Tissues of origin (as in Supplementary Figure 1e) each containing both cell lines and the respective tissues of origin, cell lines samples are binned separately from non cell line based samples. Except for the incompletely differentiated/connective tissue cluster consisting of fibroblasts, smooth muscle and adipose tissue, the solid tissue based cell lines are distinct in their gene expression from the primary tissue bases sample groups.

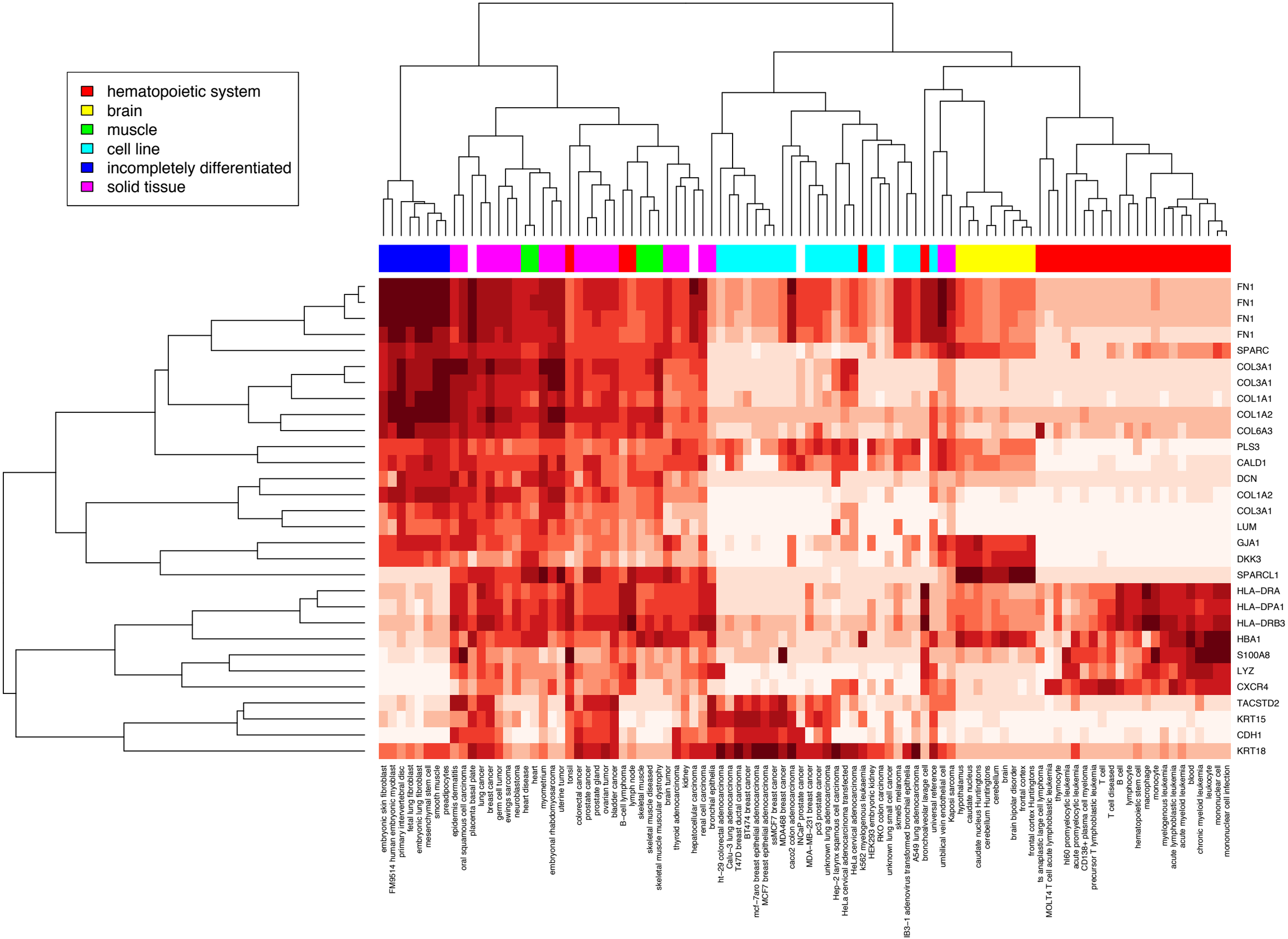

cell line
disease
neoplasm
normal
other neoplasm
sarcoma
germ cell neoplasm
breast cancer
non breast carcinoma
non neoplastic cell line
solid tissue neoplasm cell line
normal solid tissue
solid tissue non neoplastic disease
nervous system neoplasm
blood non neoplastic disease
normal blood
leukemia
non leukemic blood neoplasm
blood neoplasm cell line
blood neoplasm cell line
non leukemic blood neoplasm
leukemia
normal blood
blood non neoplastic disease
nervous system neoplasm
solid tissue non neoplastic disease
normal solid tissue
solid tissue neoplasm cell line
non neoplastic cell line
non breast carcinoma
breast cancer
germ cell neoplasm
sarcoma
other neoplasm

Supplementary Figure 4j:

15 meta-groups (as in Supplementary Figure 1d). Note that although the two largest clusters are blood based samples vs. the rest, the two solid cell line groups cluster together and they exhibit high similarity to blood based cell lines.

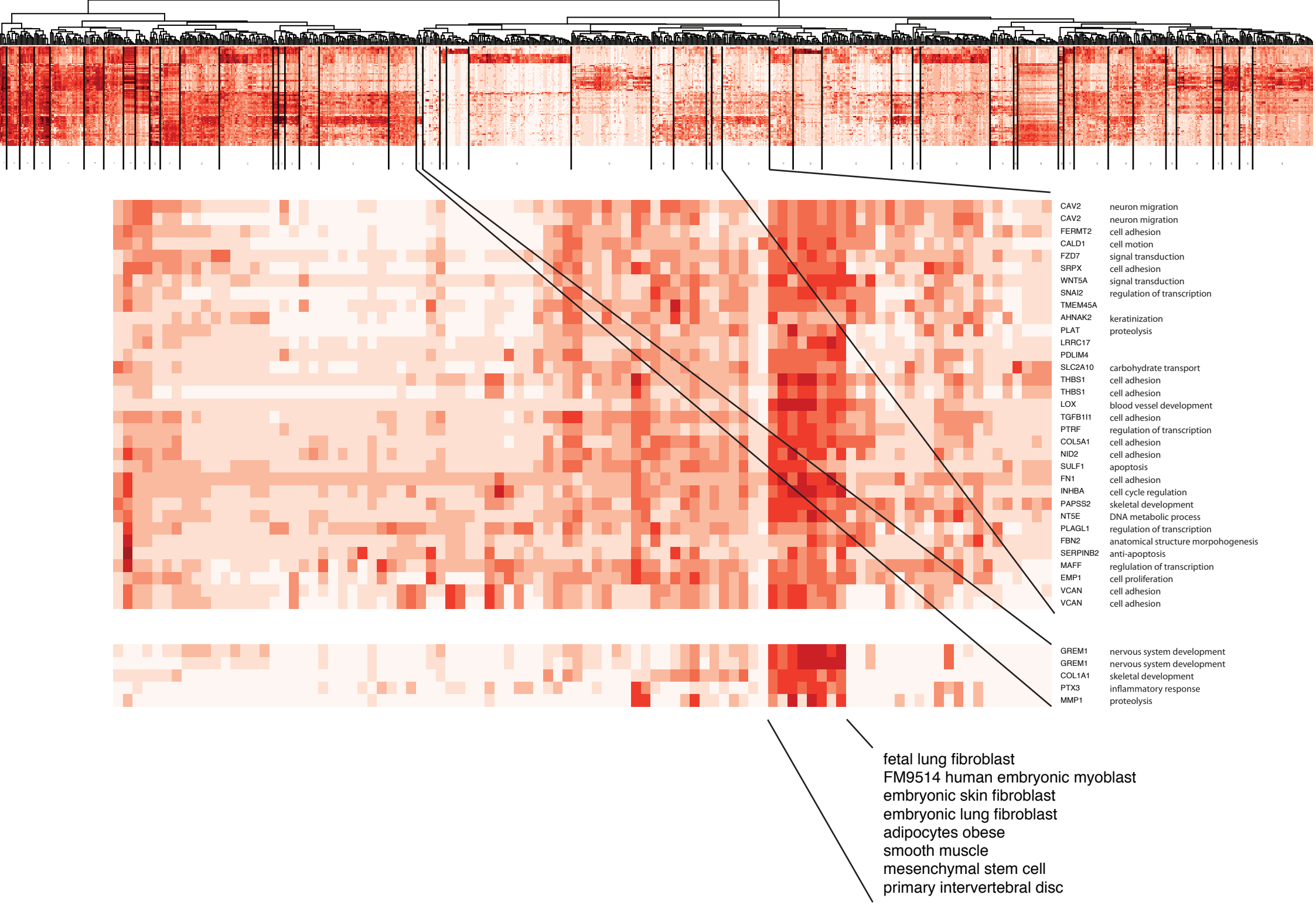
Supplementary Figure 5a:

Gene cluster heat-map. Mean gene expression in the 96 larger groups for the 1000 most variable probesets. By visual inspection we identified 50 gene clusters. Each cluster typically contains genes that are expressed in one or more of the 6 sample clusters in Supplementary Figure 4a.


hematopoietic system
brain
muscle
cell line
incompletely differentiated
solid tissue
FN1
FN1
FN1
FN1
SPARC
COL3A1
COL3A1
COL1A1
COL1A2
COL6A3
PLS3
CALD1
DCN
COL1A2
COL3A1
LUM
GJA1
DKK3
SPARCL1
HLA−DRA
HLA−DPA1
HLA−DRB3
HBA1
S100A8
LYZ
CXCR4
TACSTD2
KRT15
CDH1
KRT18
embryonic skin fibroblast
FM9514 human embryonic myoblast
primary intervertebral disc
fetal lung fibroblast
embryonic lung fibroblast
mesenchymal stem cell
smooth muscle
preadipocytes
epidermis dermatitis
oral squamous cell carcinoma
placenta basal plate
lung cancer
breast cancer
germ cell tumor
ewings sarcoma
neuroblastoma
heart disease
heart
myometrium
embryonal rhabdomyosarcoma
uterine tumor
tonsil
colorectal cancer
prostate cancer
prostate gland
ovarian tumor
bladder cancer
B−cell lymphoma
lymph node
skeletal muscle
skeletal muscle diseased
skeletal muscle muscular dystrophy
brain tumor
thyroid adenocarcinoma
kidney
hepatocellular carcinoma
renal cell carcinoma
bronchial epithelia
ht−29 colorectal adenocarcinoma
Calu−3 lung adenocarcinoma
T47D breast ductal carcinoma
BT474 breast cancer
mcf−7aro breast epithelial adenocarcinoma
MCF7 breast epithelial adenocarcinoma
ssMCF7 breast cancer
MDA468 breast cancer
caco2 colon adenocarcinoma
INCaP prostate cancer
MDA−MB−231 breast cancer
pc3 prostate cancer
unknown lung adenocarcinoma
Hep−2 larynx squamous cell carcinoma
HeLa cervical adenocarcinoma transfected
HeLa cervical adenocarcinoma
k562 myelogenous leukaemia
HEK293 embryonic kidney
RKO colon carcinoma
unknown lung small cell cancer
skmel5 melanoma
IB3−1 adenovirus transformed bronchial epithelia
A549 lung adenocarcinoma
bronchoalveolar lavage cell
universal reference
umbilical vein endothelial cell
Kaposi sarcoma
hypothalamus
caudate nucleus
caudate nucleus Huntingtons
cerebellum Huntingtons
cerebellum
brain
brain bipolar disorder
frontal cortex
frontal cortex Huntingtons
ts anaplastic large cell lymphoma
MOLT4 T cell acute lymphoblastic leukemia
thymocyte
hl60 promyelocytic leukemia
acute promyelocytic leukemia
CD138+ plasma cell myeloma
precursor T lymphoblastic leukemia
T cell
T cell diseased
B cell
lymphocyte
hematopoietic stem cell
macrophage
monocyte
myelogenous leukemia
acute lymphoblastic leukemia
acute myeloid leukemia
blood
chronic myeloid leukemia
leukocyte
mononuclear cell
mononuclear cell infection

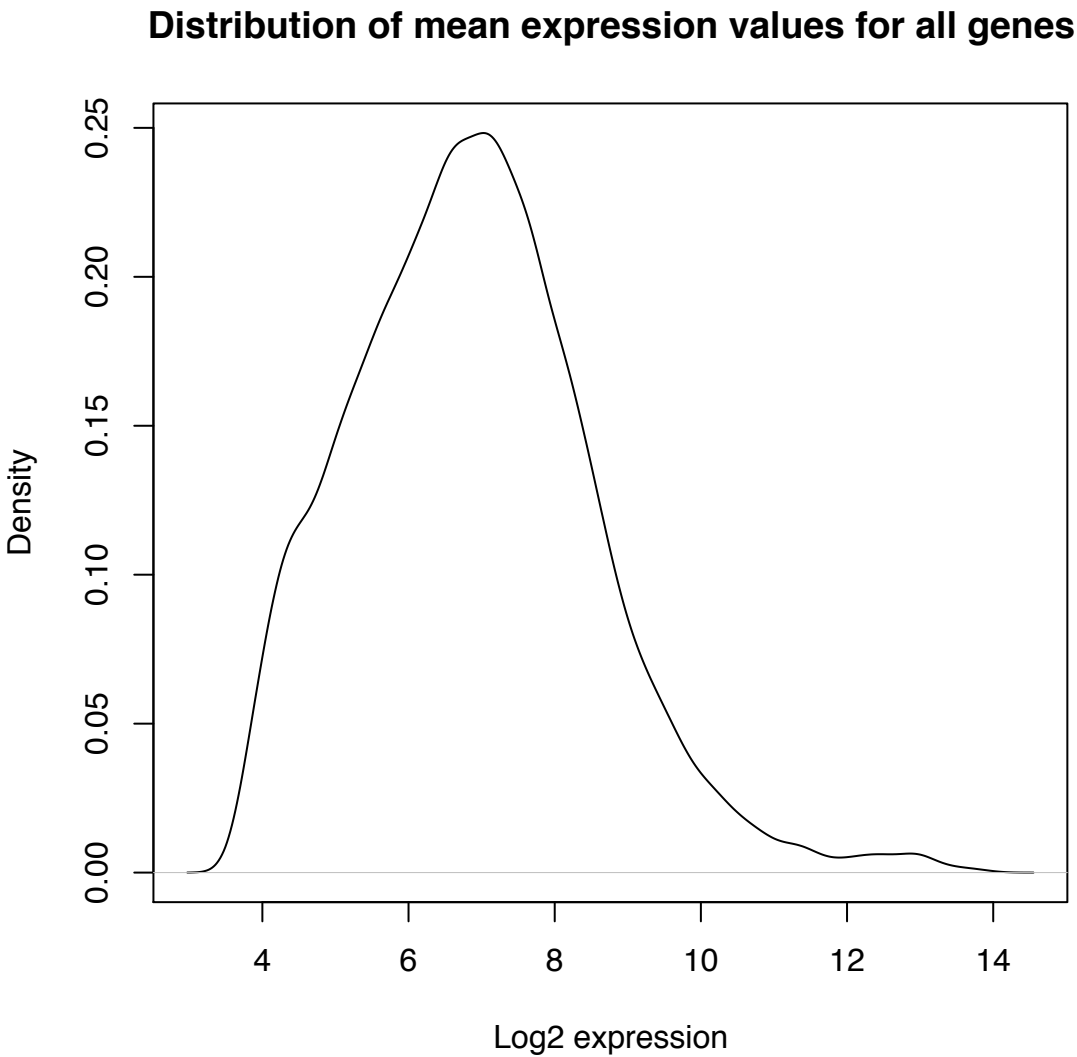
Supplementary Figure 5b:

The clustering of the 96 larger groups based on 30 most variable probesets. The samples are colored according to the 6 original sample clusters based on all data. We can observe that the samples from the 6 original clusters as defined by expression profiles of all genes are still closely together, in particular the brain and the incompletely differentiated cell clusters are intact. The 30 most variable genes include some well-known marker genes such as FN1 (fibronectin), SPARC, several collagen genes, PLS3 (actin-binding protein), CALD1 (calmodium), several HLA genes, HBA (human alpha globin), LYZ (lysozyme), CDH1 (cadherin), and KRT18 (keratin).


CAV2 neuron migration
CAV2 neuron migration
FERMT2 cell adhesion
CALD1 cell motion
FZD7 signal transduction
SRPX cell adhesion
WNT5A signal transduction
SNAI2 regulation of transcription
TMEM45A
AHNAK2 keratinization
PLAT proteolysis
LRRC17
PDLIM4
SLC2A10 carbohydrate transport
THBS1 cell adhesion
THBS1 cell adhesion
LOX blood vessel development
TGFB1I1 cell adhesion
PTRF regulation of transcription
COL5A1 cell adhesion
NID2 cell adhesion
SULF1 apoptosis
FN1 cell adhesion
INHBA cell cycle regulation
PAPSS2 skeletal development
NT5E DNA metabolic process
PLAGL1 regulation of transcription
FBN2 anatomical structure morpohogenesis
SERPINB2 anti-apoptosis
MAFF reglulation of transcription
EMP1 cell proliferation
VCAN cell adhesion
VCAN cell adhesion
GREM1 nervous system development
GREM1 nervous system development
COL1A1 skeletal development
PTX3 inflammatory response
MMP1 proteolysis
fetal lung fibroblast
FM9514 human embryonic myoblast
embryonic skin fibroblast
embryonic lung fibroblast
adipocytes obese
smooth muscle
mesenchymal stem cell
primary intervertebral disc

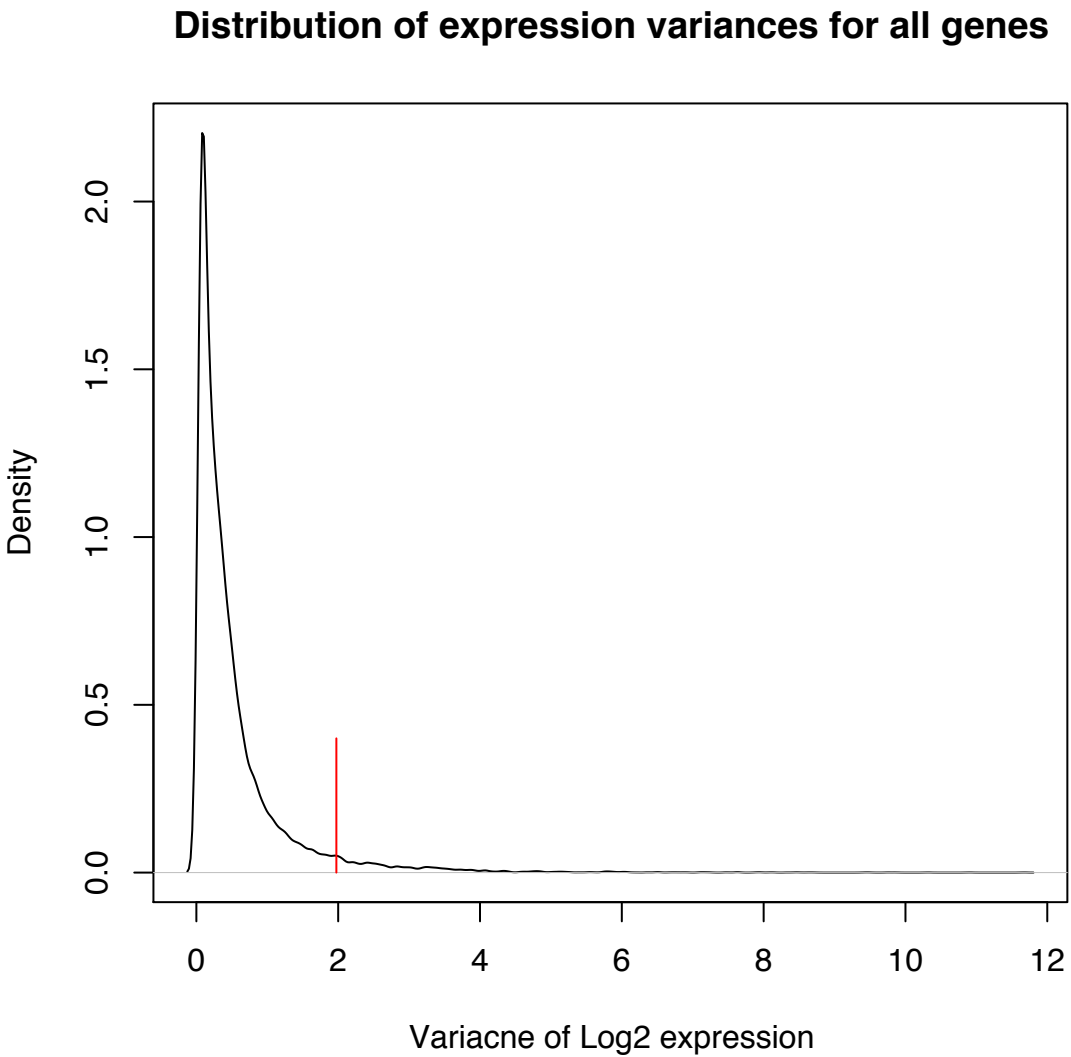
Supplementary Figure 5c:

Gene cluster heat-map. Mean gene expression in the 96 larger groups for the 1000 most variable probesets. By visual inspection we identified 50 gene clusters. Each cluster typically contains genes that are expressed in one or more of the 6 sample clusters in Supplementary Figure 4a. We zoom into the two gene clusters that map onto the incompletely differentiated cell type and connective tissue sample cluster. These genes are overrepresented for GO terms related to cell adhesion, developmental processes and sub-cellular location - extracellular region.


Distribution of mean expression values for all genes
Density
0.00
0.05
0.10
0.15
0.20
0.25
4
6
8
10
12
14
Log2 expression

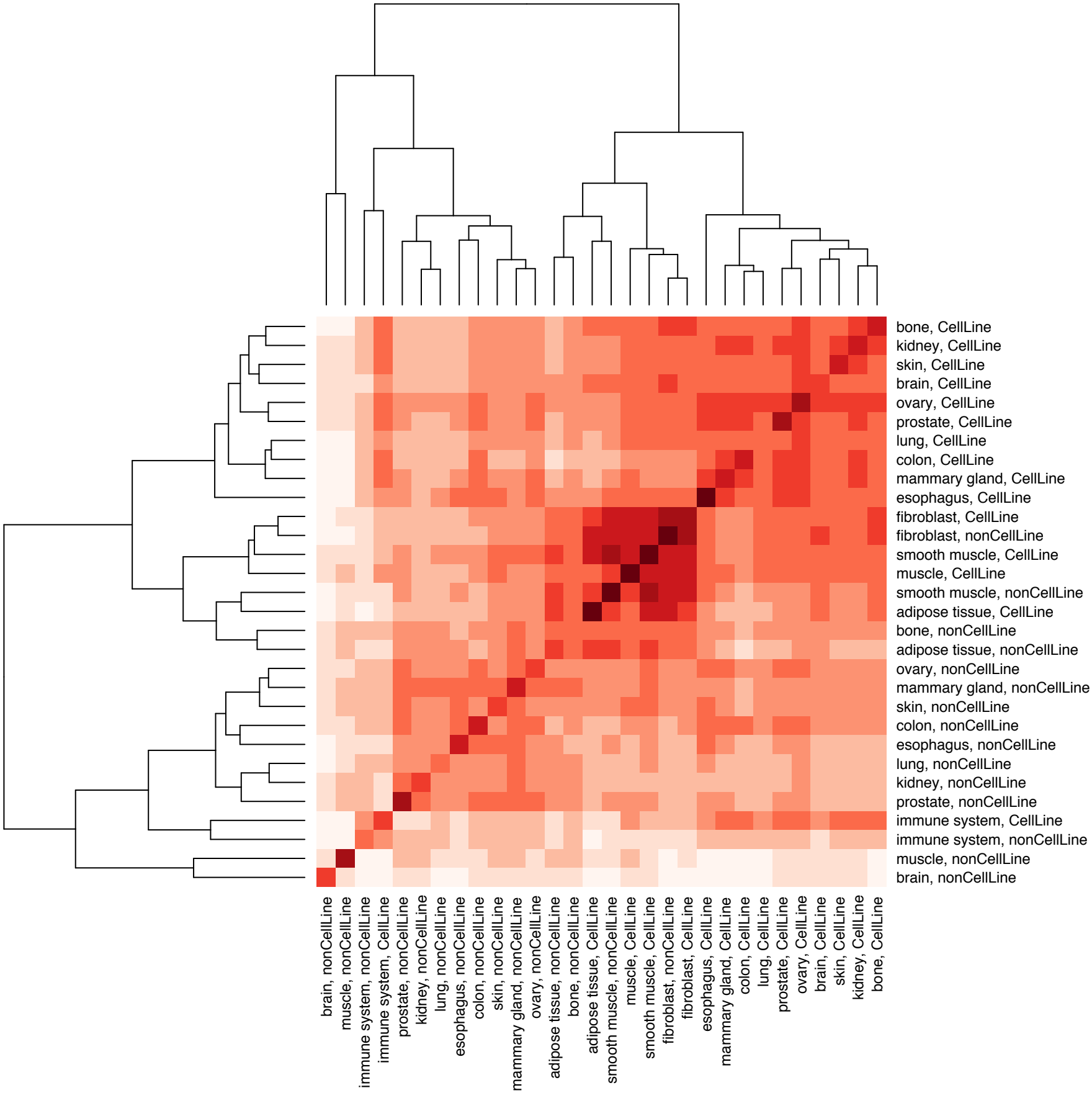
Supplementary Figure 6a:

Distribution of the mean expressions of the probesets across samples.


Distribution of expression variances for all genes
Density
0.0
0.5
1.0
1.5
2.0
0
2
4
6
8
10
12
Variacne of Log2 expression

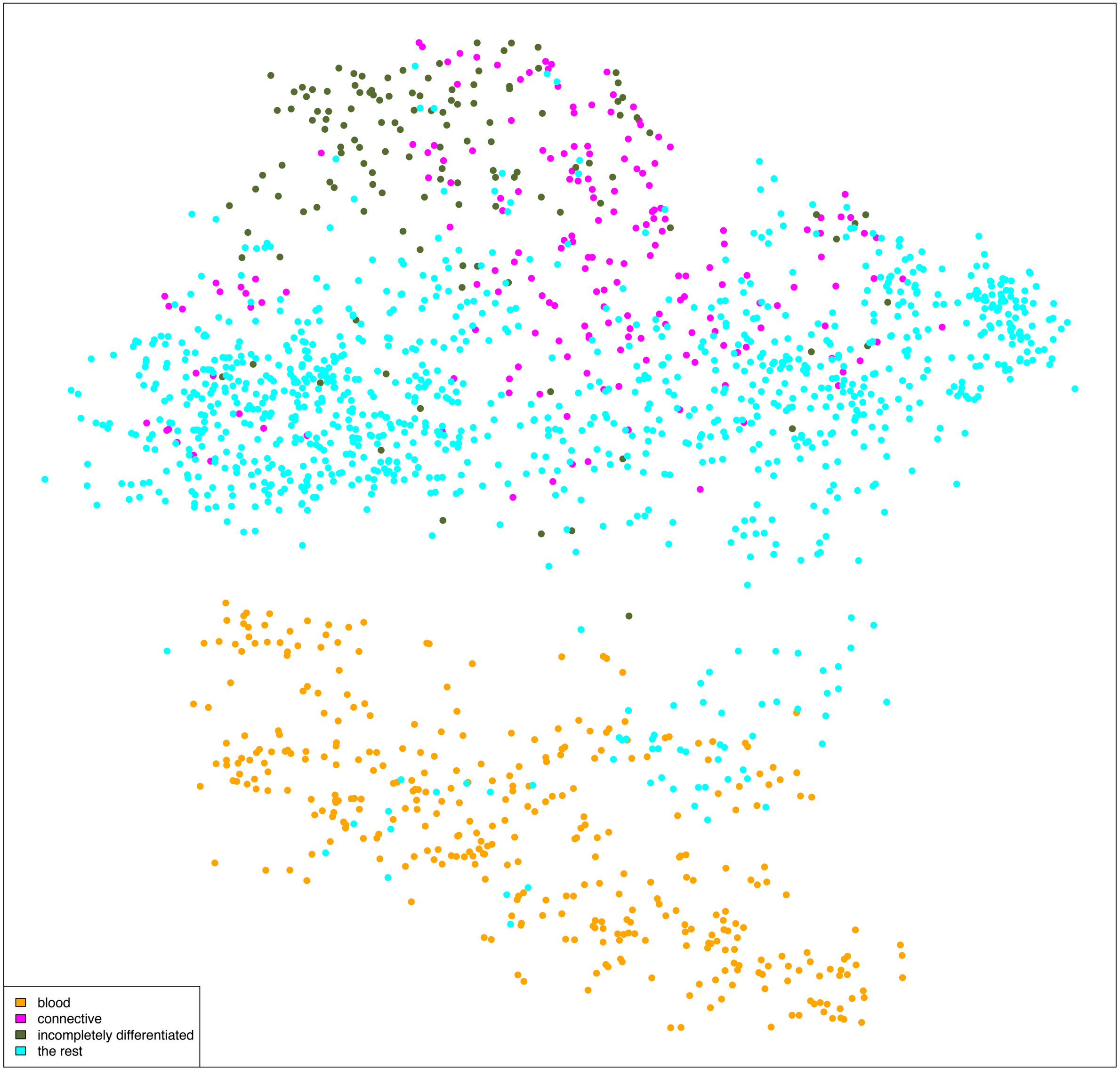
Supplementary Figure 6b:

Distribution of variations in the probeset expression across samples. The red line indicates the cutoff for the 1000 most variable probesets.


bone, CellLine
kidney, CellLine
skin, CellLine
brain, CellLine
ovary, CellLine
prostate, CellLine
lung, CellLine
colon, CellLine
mammary gland, CellLine
esophagus, CellLine
fibroblast, CellLine
fibroblast, nonCellLine
smooth muscle, CellLine
muscle, CellLine
smooth muscle, nonCellLine
adipose tissue, CellLine
bone, nonCellLine
adipose tissue, nonCellLine
ovary, nonCellLine
mammary gland, nonCellLine
skin, nonCellLine
colon, nonCellLine
esophagus, nonCellLine
lung, nonCellLine
kidney, nonCellLine
prostate, nonCellLine
immune system, CellLine
immune system, nonCellLine
muscle, nonCellLine
brain, nonCellLine
brain, nonCellLine
muscle, nonCellLine
immune system, nonCellLine
prostate, nonCellLine
kidney, nonCellLine
lung, nonCellLine
esophagus, nonCellLine
colon, nonCellLine
skin, nonCellLine
mammary gland, nonCellLine
ovary, nonCellLine
adipose tissue, nonCellLine
bone, nonCellLine
adipose tissue, CellLine
smooth muscle, nonCellLine
muscle, CellLine
smooth muscle, CellLine
fibroblast, nonCellLine
fibroblast, CellLine
esophagus, CellLine
mammary gland, CellLine
colon, CellLine
lung, CellLine
prostate, CellLine
ovary, CellLine
brain, CellLine
skin, CellLine
kidney, CellLine
bone, CellLine

Supplementary Figure 7:

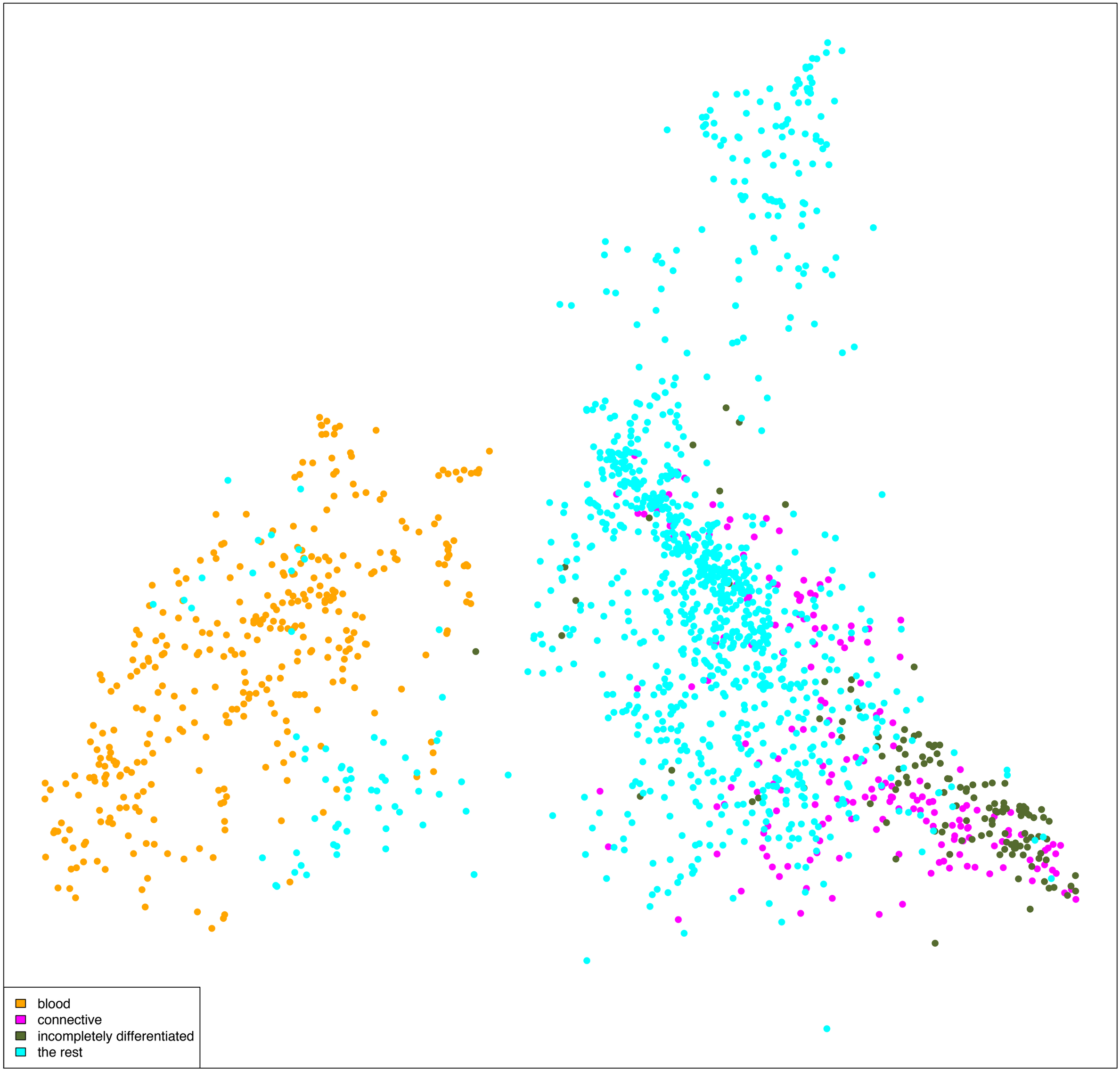
Heat map comparison of 15 tissues of origin after removal of 1410 cell line specific probesets.


blood
connective
incompletely differentiated
the rest

Supplementary Figure 8a:

Principal component analysis on sampled data. Each dot represents one of the 5372 samples in a multidimensional gene expression space projected on the 'principal' plane formed by pairs of first 5 principal components. The dots are colored (semi-transparently) according to the biological group the sample belongs to (see figure legends).

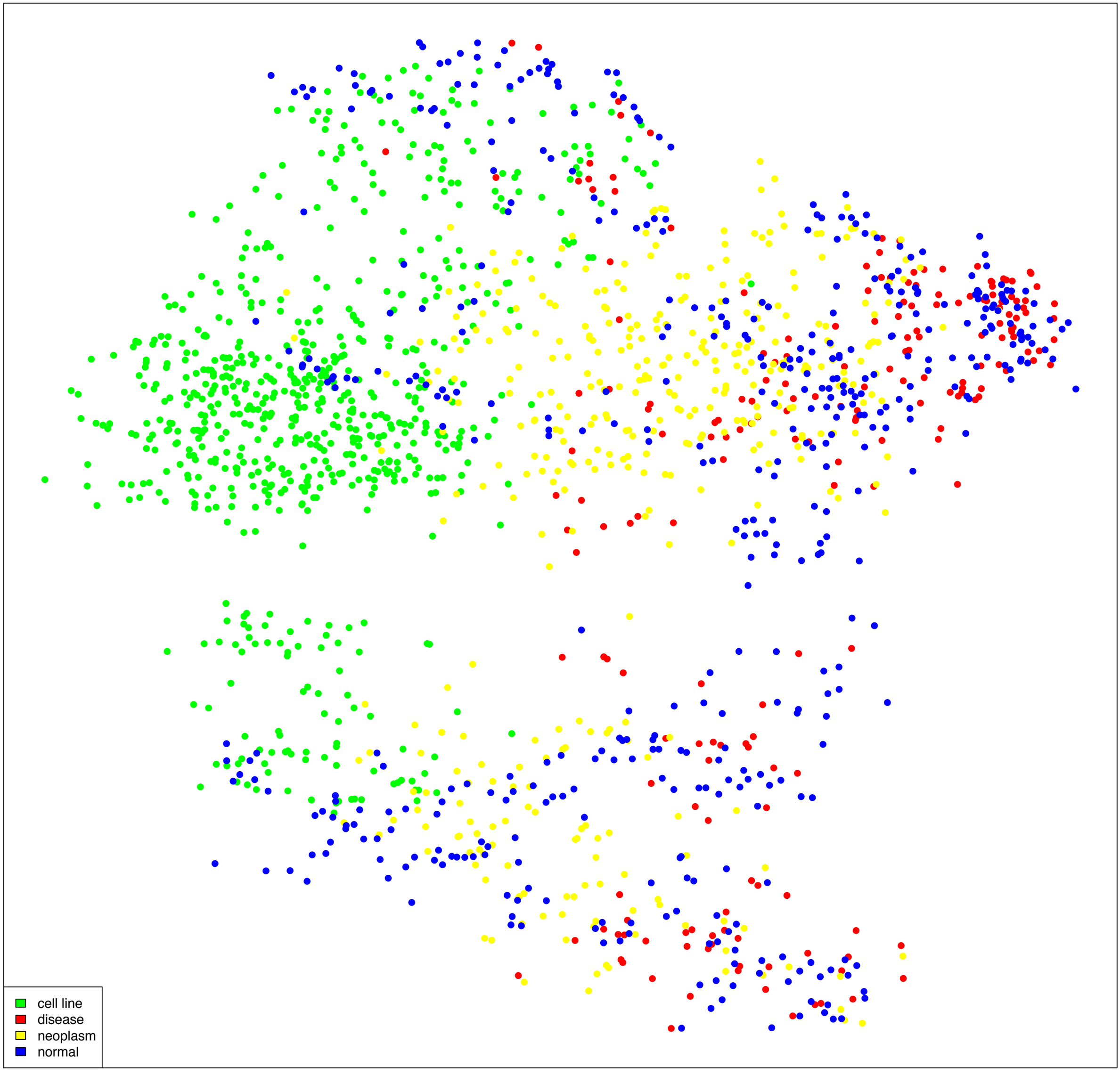
Samples of blood (orange), connective tissues (magenta), incompletely differentiated cells (green) and the rest (cyan) colored on the 1st and 2nd principal component plane.


blood
connective
incompletely differentiated
the rest

Supplementary Figure 8a23:

Principal component analysis on sampled data. Each dot represents one of the 5372 samples in a multidimensional gene expression space projected on the 'principal' plane formed by pairs of first 5 principal components. The dots are colored (semi-transparently) according to the biological group the sample belongs to (see figure legends).

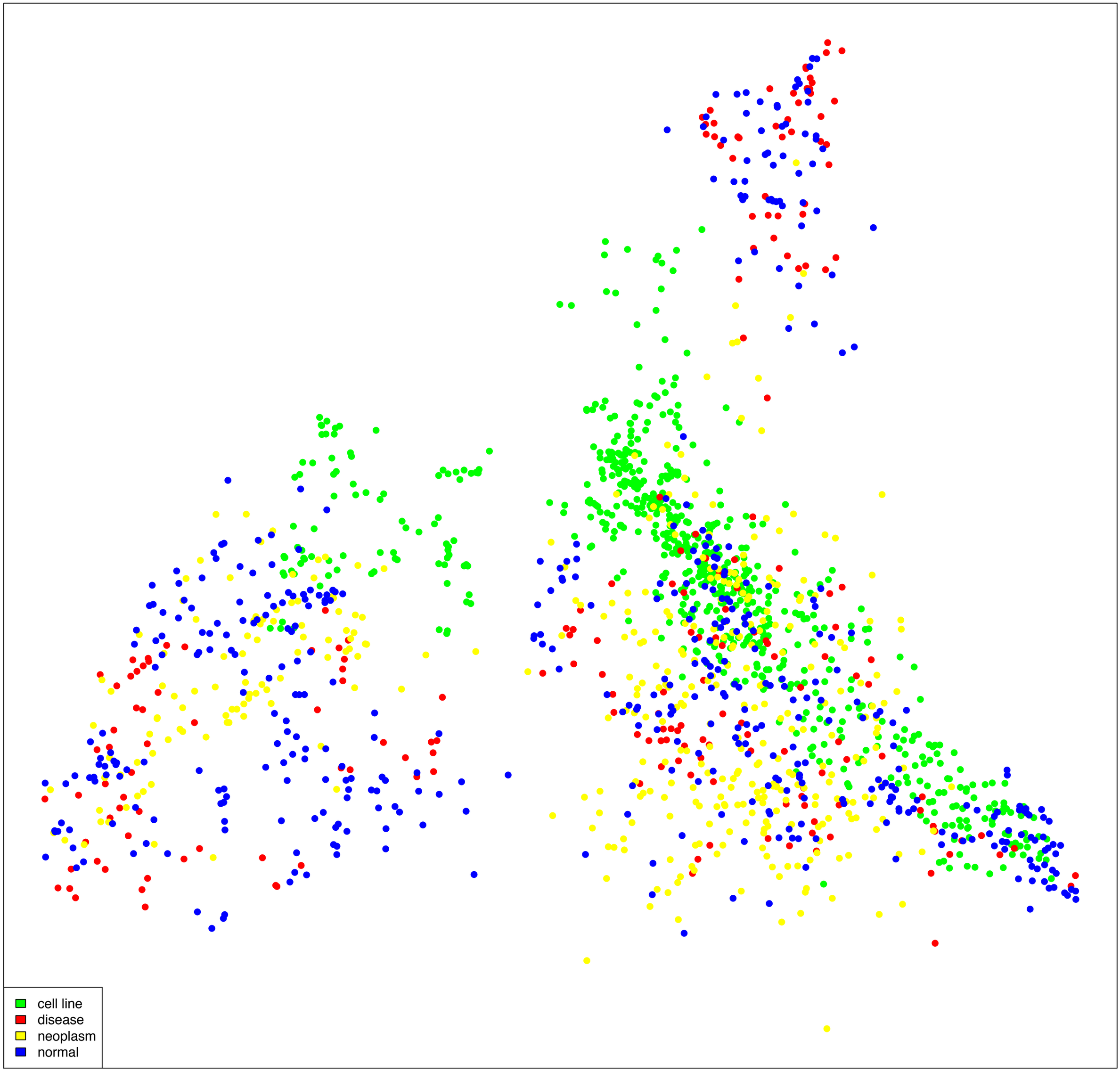
Samples of blood (orange), connective tissues (magenta), incompletely differentiated cells (green) and the rest (cyan) colored on the 2nd and 3rd principal component plane.


cell line
disease
neoplasm
normal

Supplementary Figure 8b:

Principal component analysis on sampled data. Each dot represents one of the 5372 samples in a multidimensional gene expression space projected on the 'principal' plane formed by pairs of first 5 principal components. The dots are colored (semi-transparently) according to the biological group the sample belongs to (see figure legends).

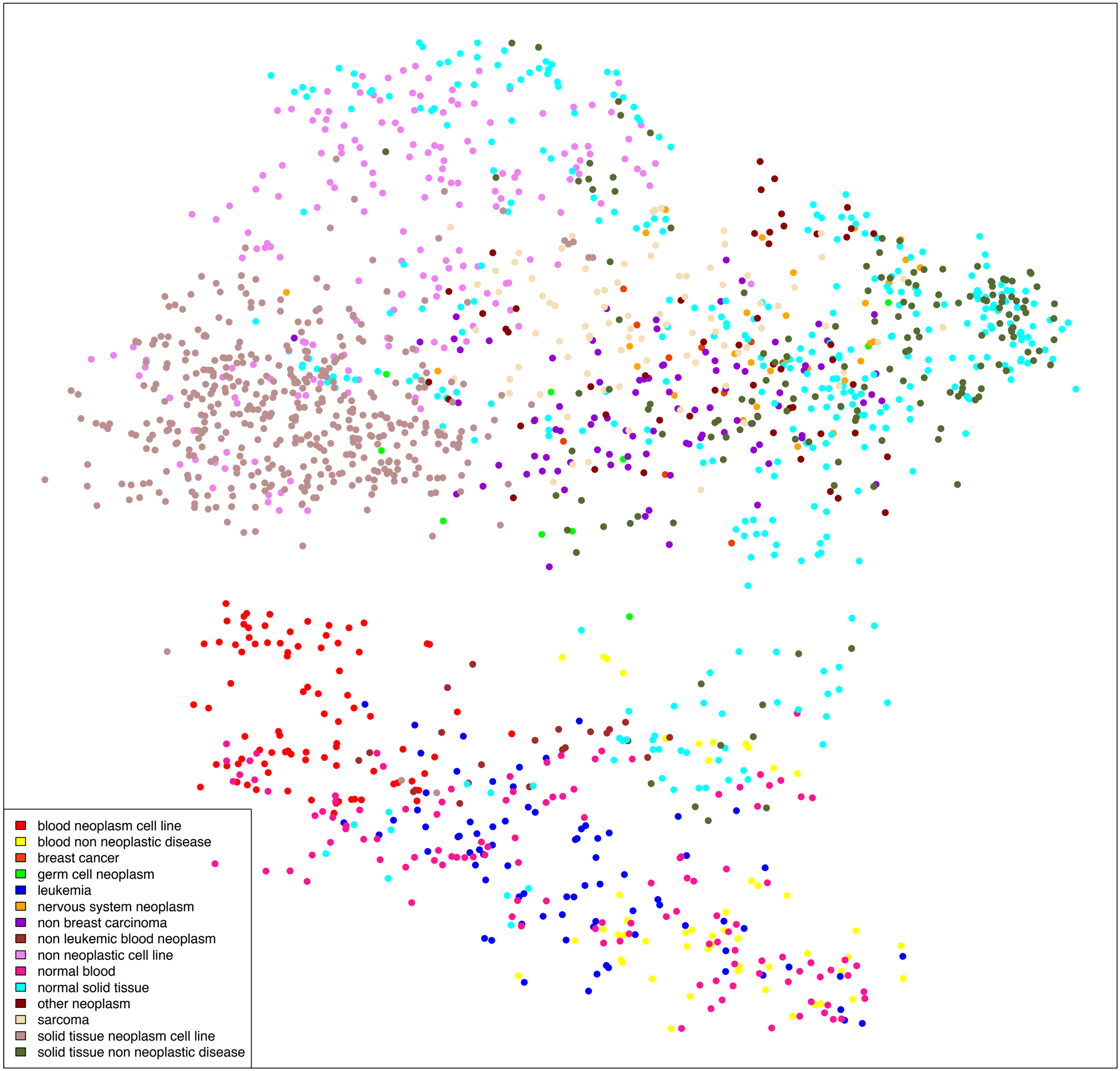
Samples colored by 4 meta-groups (see legend) on the 1st and 2nd principal component plane.


cell line
disease
neoplasm
normal

Supplementary Figure 8b23:

Principal component analysis on sampled data. Each dot represents one of the 5372 samples in a multidimensional gene expression space projected on the 'principal' plane formed by pairs of first 5 principal components. The dots are colored (semi-transparently) according to the biological group the sample belongs to (see figure legends).

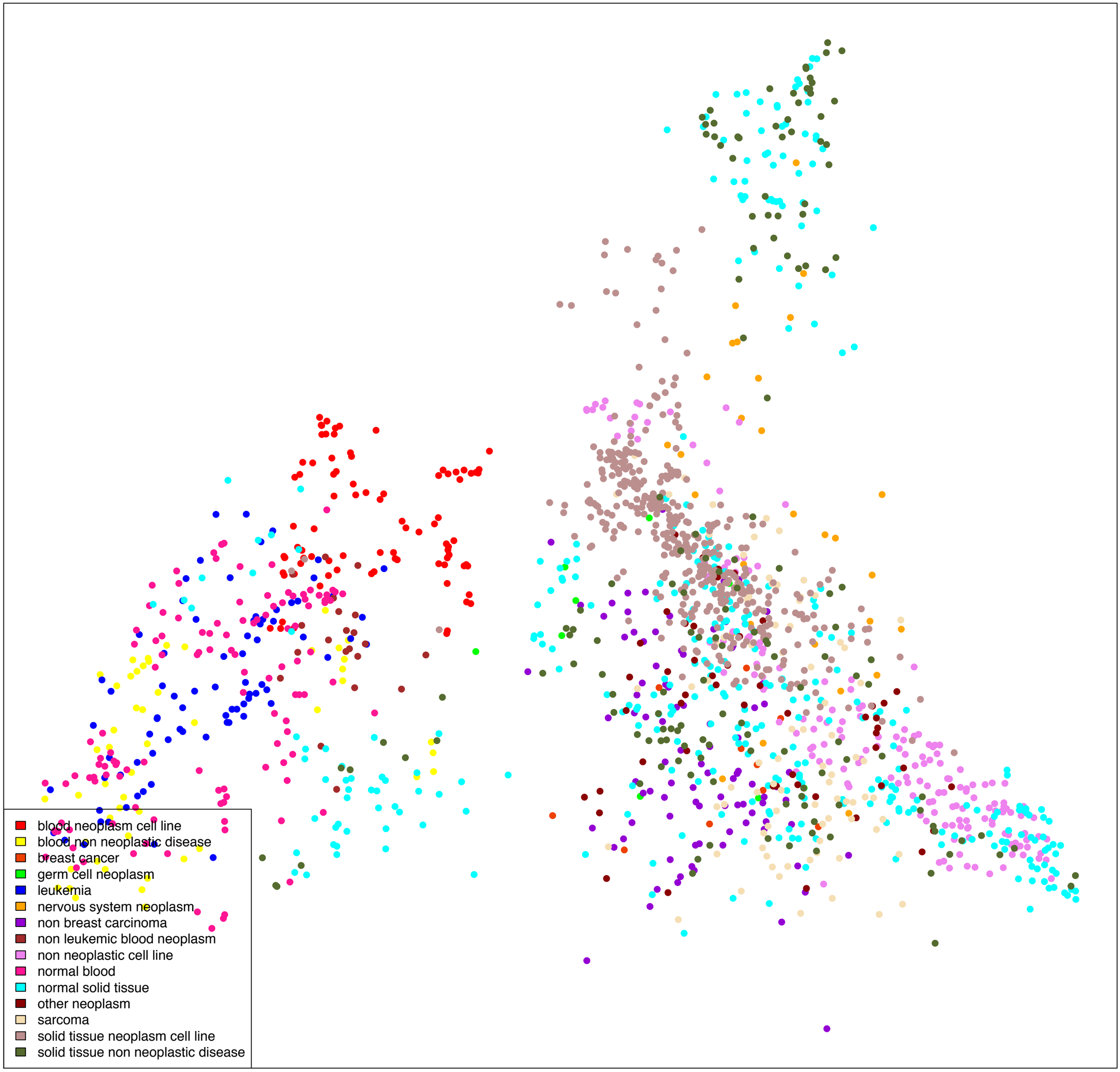
Samples colored by 4 meta-groups (see legend) on the 2nd and 3rd principal component plane.


blood neoplasm cell line
blood non neoplastic disease
breast cancer
germ cell neoplasm
leukemia
nervous system neoplasm
non breast carcinoma
non leukemic blood neoplasm
non neoplastic cell line
normal blood
normal solid tissue
other neoplasm
sarcoma
solid tissue neoplasm cell line
solid tissue non neoplastic disease

Supplementary Figure 8c:

Principal component analysis on sampled data. Each dot represents one of the 5372 samples in a multidimensional gene expression space projected on the 'principal' plane formed by pairs of first 5 principal components. The dots are colored (semi-transparently) according to the biological group the sample belongs to (see figure legends).

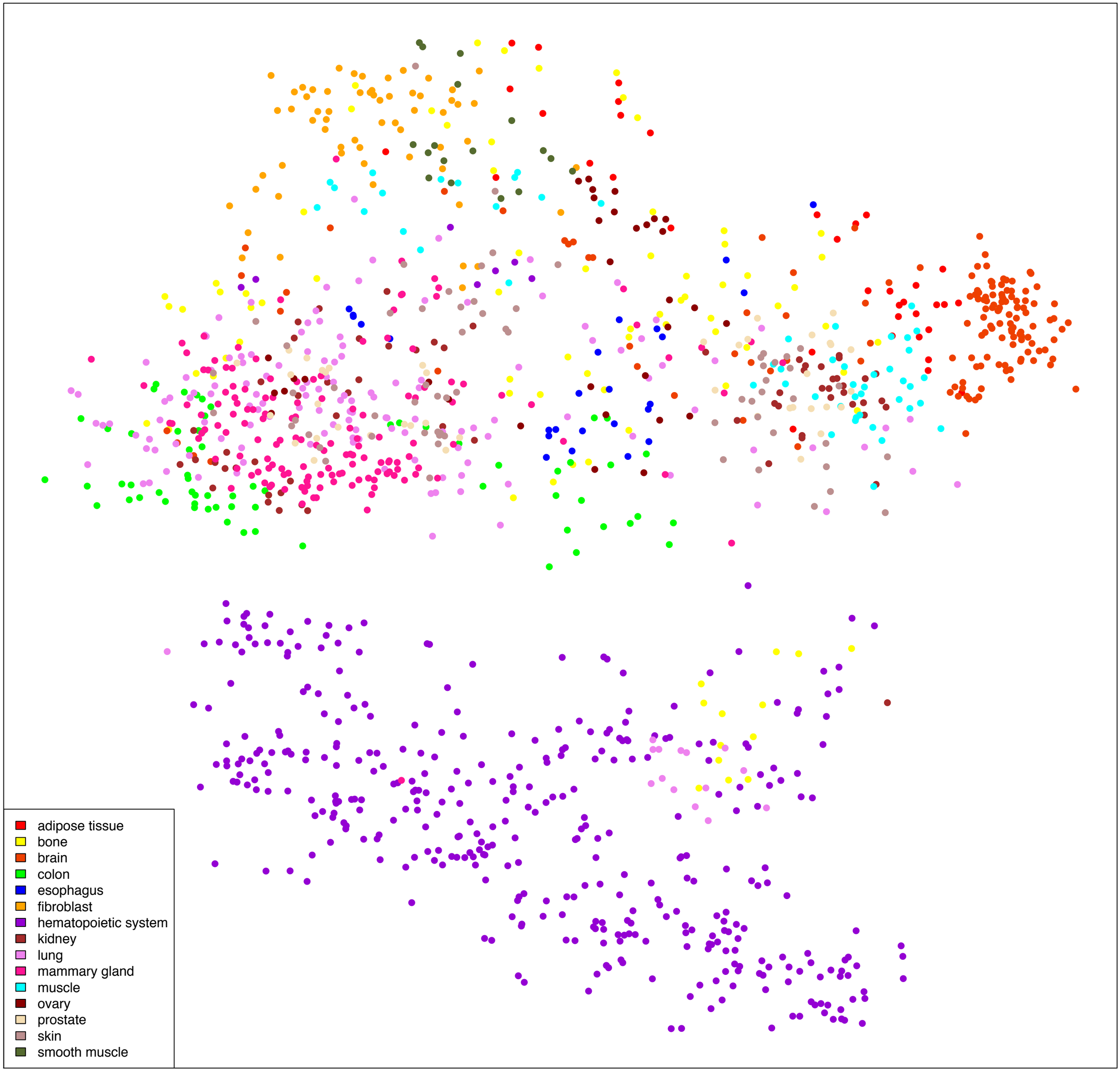
Samples colored by 15 meta-groups (see legend) on the 1st and 2nd principal component plane.


blood neoplasm cell line
blood non neoplastic disease
breast cancer
germ cell neoplasm
leukemia
nervous system neoplasm
non breast carcinoma
non leukemic blood neoplasm
non neoplastic cell line
normal blood
normal solid tissue
other neoplasm
sarcoma
solid tissue neoplasm cell line
solid tissue non neoplastic disease

Supplementary Figure 8c23:

Principal component analysis on sampled data. Each dot represents one of the 5372 samples in a multidimensional gene expression space projected on the 'principal' plane formed by pairs of first 5 principal components. The dots are colored (semi-transparently) according to the biological group the sample belongs to (see figure legends).

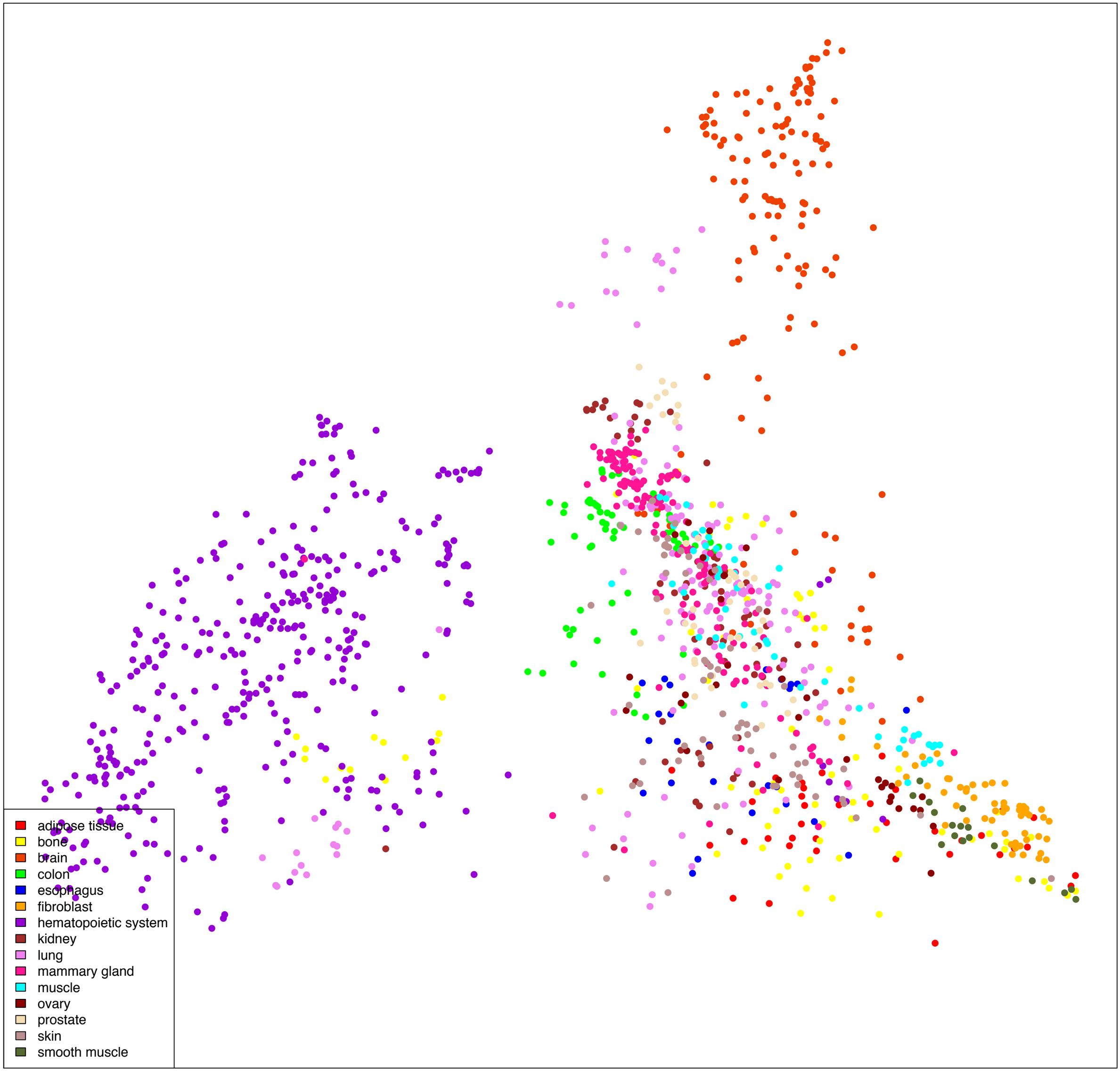
Samples colored by 15 meta-groups (see legend) on the 2nd and 3rd principal component plane.


adipose tissue
bone
brain
colon
esophagus
fibroblast
hematopoietic system
kidney
lung
mammary gland
muscle
ovary
prostate
skin
smooth muscle

Supplementary Figure 8d:

Principal component analysis on sampled data. Each dot represents one of the 5372 samples in a multidimensional gene expression space projected on the 'principal' plane formed by pairs of first 5 principal components. The dots are colored (semi-transparently) according to the biological group the sample belongs to (see figure legends).

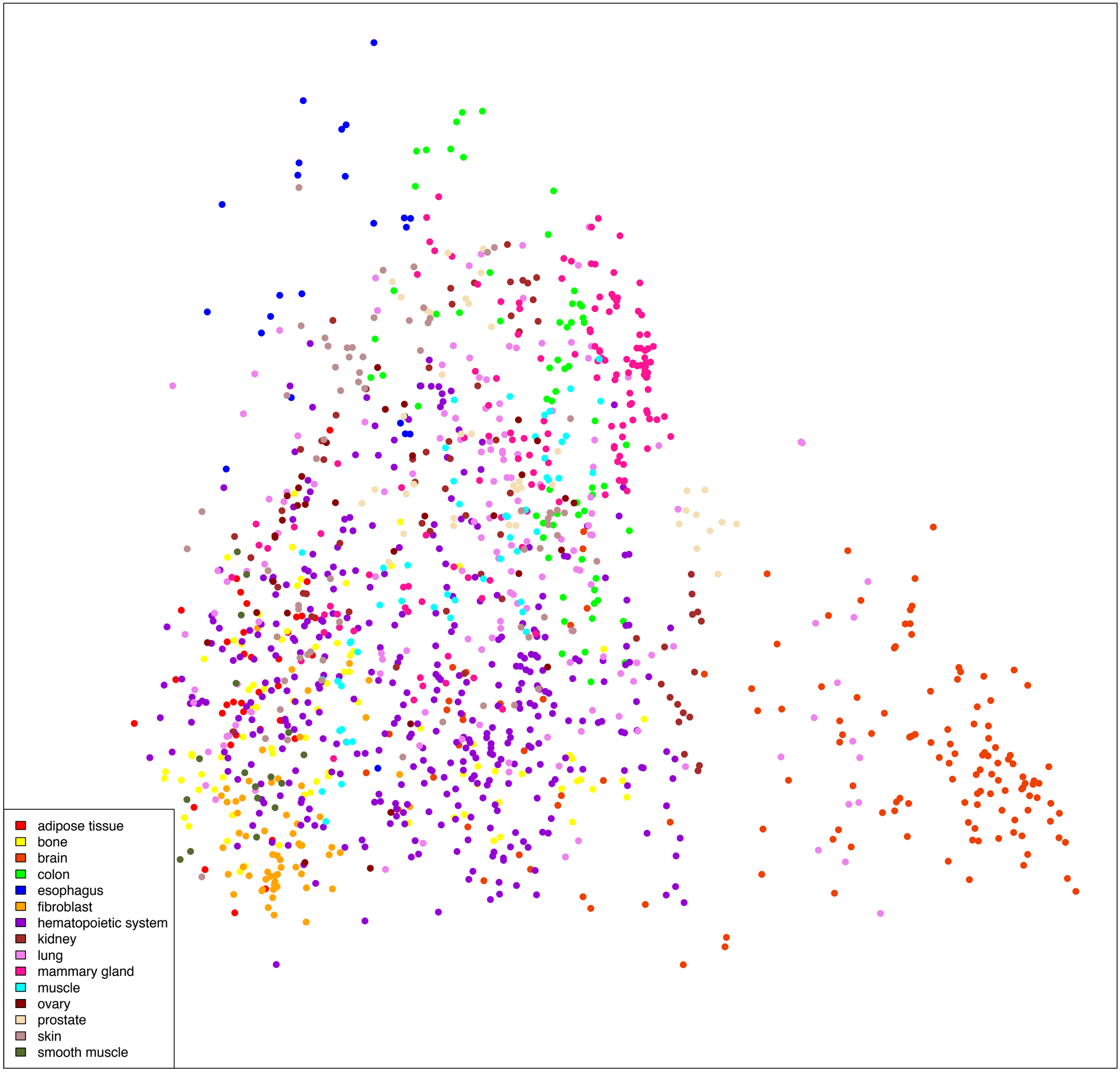
Samples colored by 15 groups of tissues of origin (see legend) on the 1st and 2nd principal component plane.


adipose tissue
bone
brain
colon
esophagus
fibroblast
hematopoietic system
kidney
lung
mammary gland
muscle
ovary
prostate
skin
smooth muscle

Supplementary Figure 8d23:

Principal component analysis on sampled data. Each dot represents one of the 5372 samples in a multidimensional gene expression space projected on the 'principal' plane formed by pairs of first 5 principal components. The dots are colored (semi-transparently) according to the biological group the sample belongs to (see figure legends).

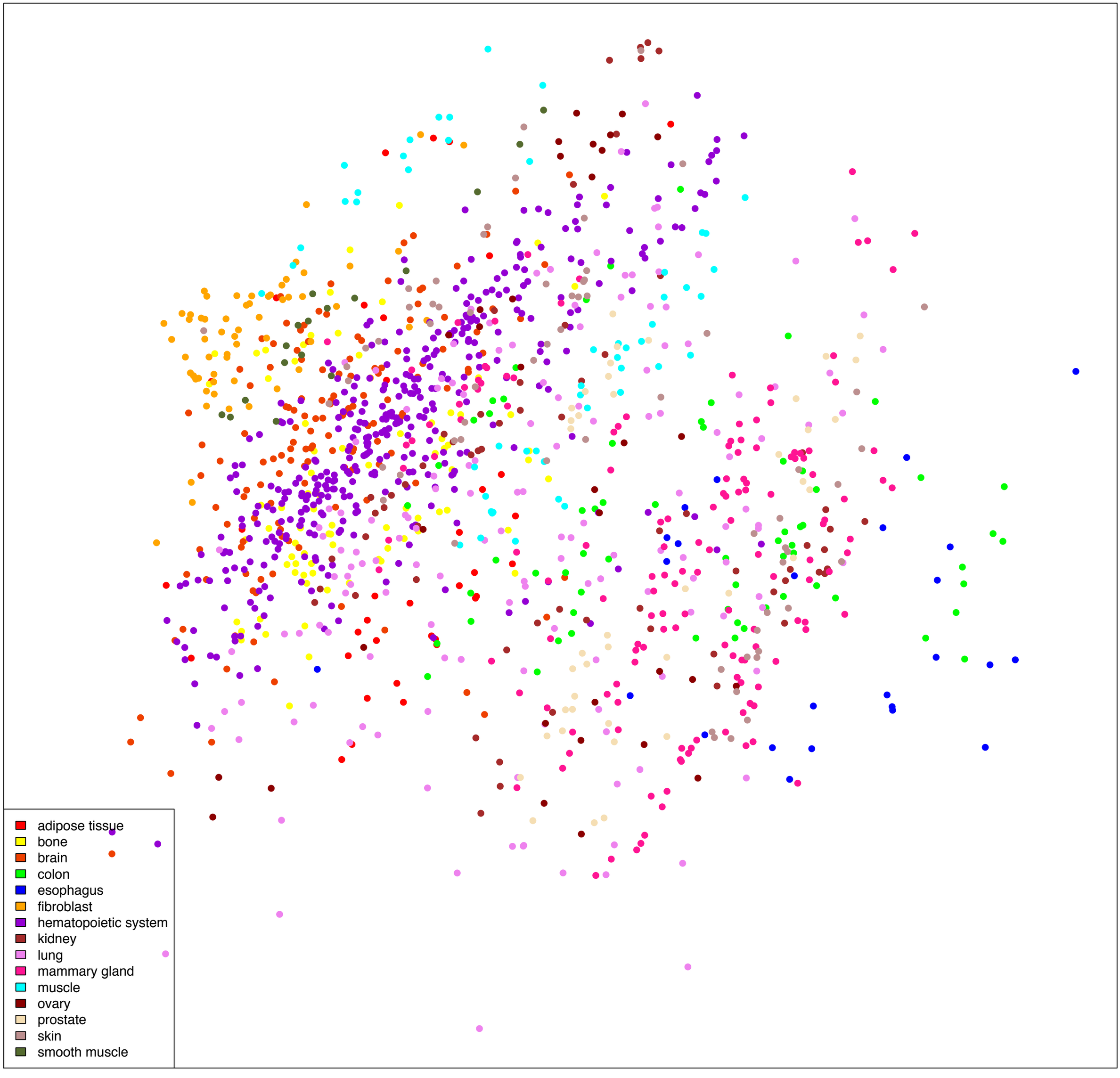
Samples colored by 15 groups of tissues of origin (see legend) on the 2nd and 3rd principal component plane.


adipose tissue
bone
brain
colon
esophagus
fibroblast
hematopoietic system
kidney
lung
mammary gland
muscle
ovary
prostate
skin
smooth muscle

Supplementary Figure 8d34:

Principal component analysis on sampled data. Each dot represents one of the 5372 samples in a multidimensional gene expression space projected on the 'principal' plane formed by pairs of first 5 principal components. The dots are colored (semi-transparently) according to the biological group the sample belongs to (see figure legends).

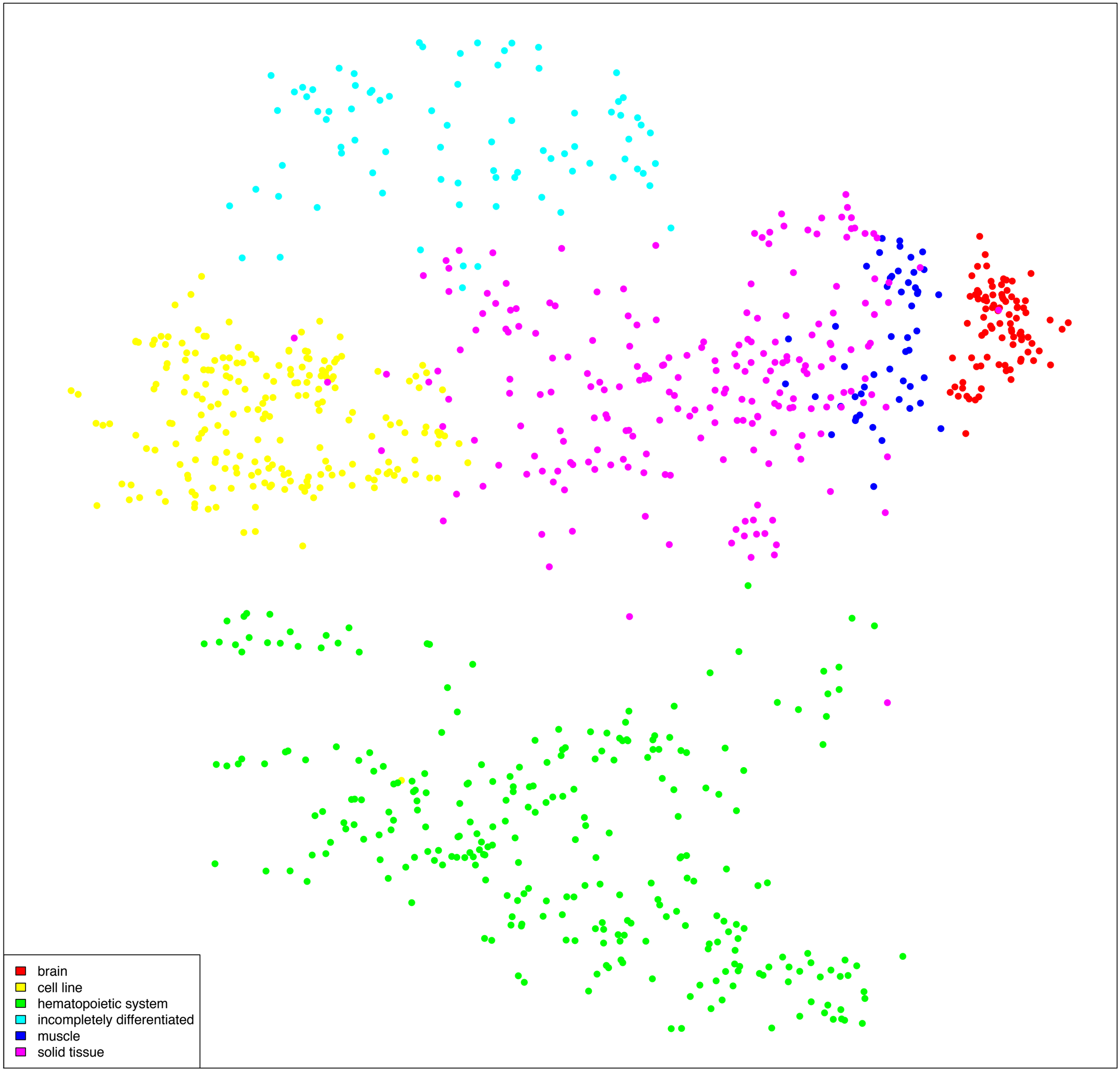
Samples colored by 15 groups of tissues of origin (see legend) on the 3rd and 4th principal component plane.


adipose tissue
bone
brain
colon
esophagus
fibroblast
hematopoietic system
kidney
lung
mammary gland
muscle
ovary
prostate
skin
smooth muscle

Supplementary Figure 8d45:

Principal component analysis on sampled data. Each dot represents one of the 5372 samples in a multidimensional gene expression space projected on the 'principal' plane formed by pairs of first 5 principal components. The dots are colored (semi-transparently) according to the biological group the sample belongs to (see figure legends).

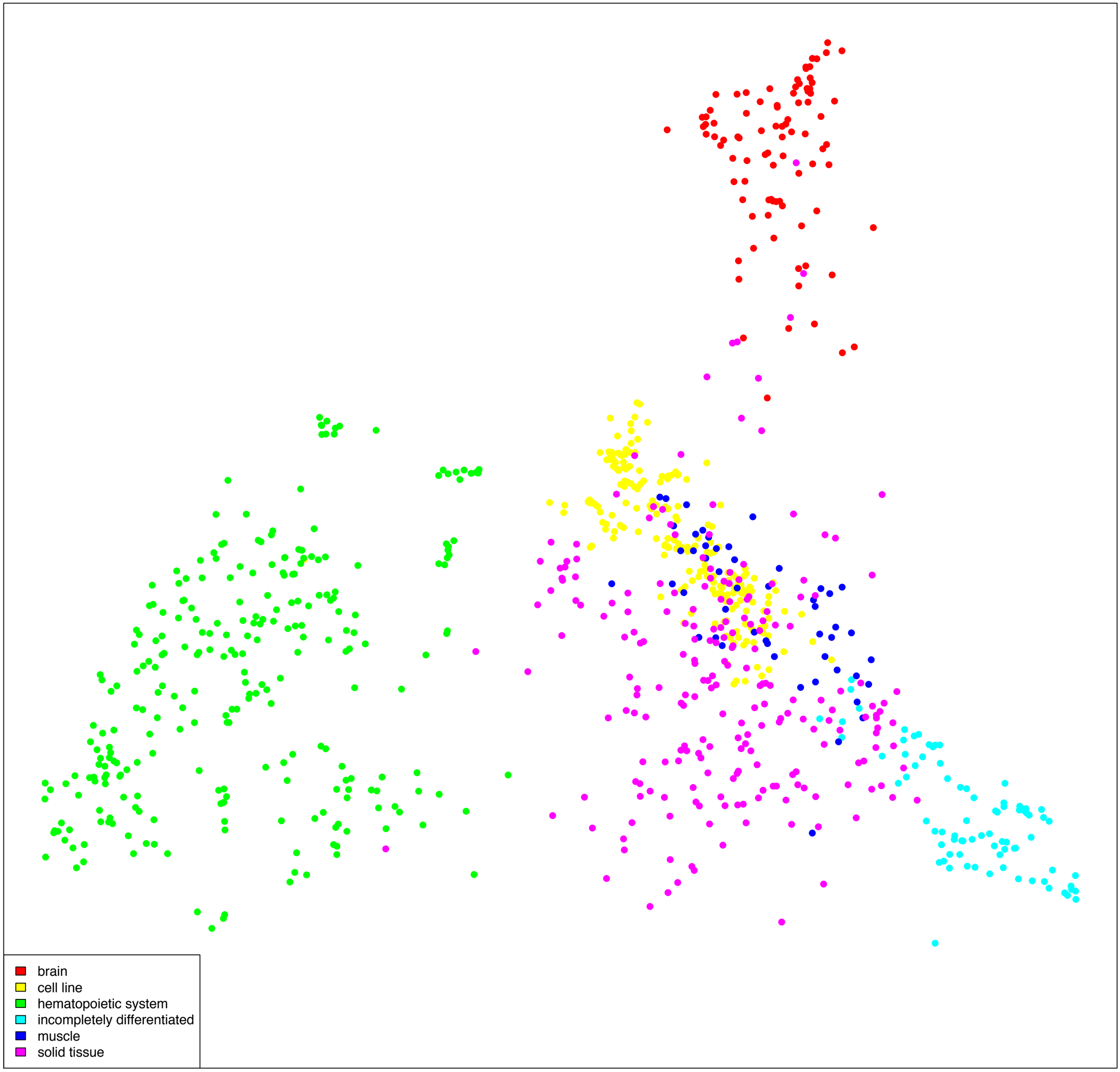
Samples colored by 15 groups of tissues of origin (see legend) on the 4th and 5th principal component plane. One can observe, that the 5th component mammary gland cell line and primary tissue samples seem to be brought closer together. This is true also for several other tissue types, however we were not able to prove this statistically.


brain
cell line
hematopoietic system
incompletely differentiated
muscle
solid tissue

Supplementary Figure 8e:

Principal component analysis on sampled data. Each dot represents one of the 5372 samples in a multidimensional gene expression space projected on the 'principal' plane formed by pairs of first 5 principal components. The dots are colored (semi-transparently) according to the biological group the sample belongs to (see figure legends).

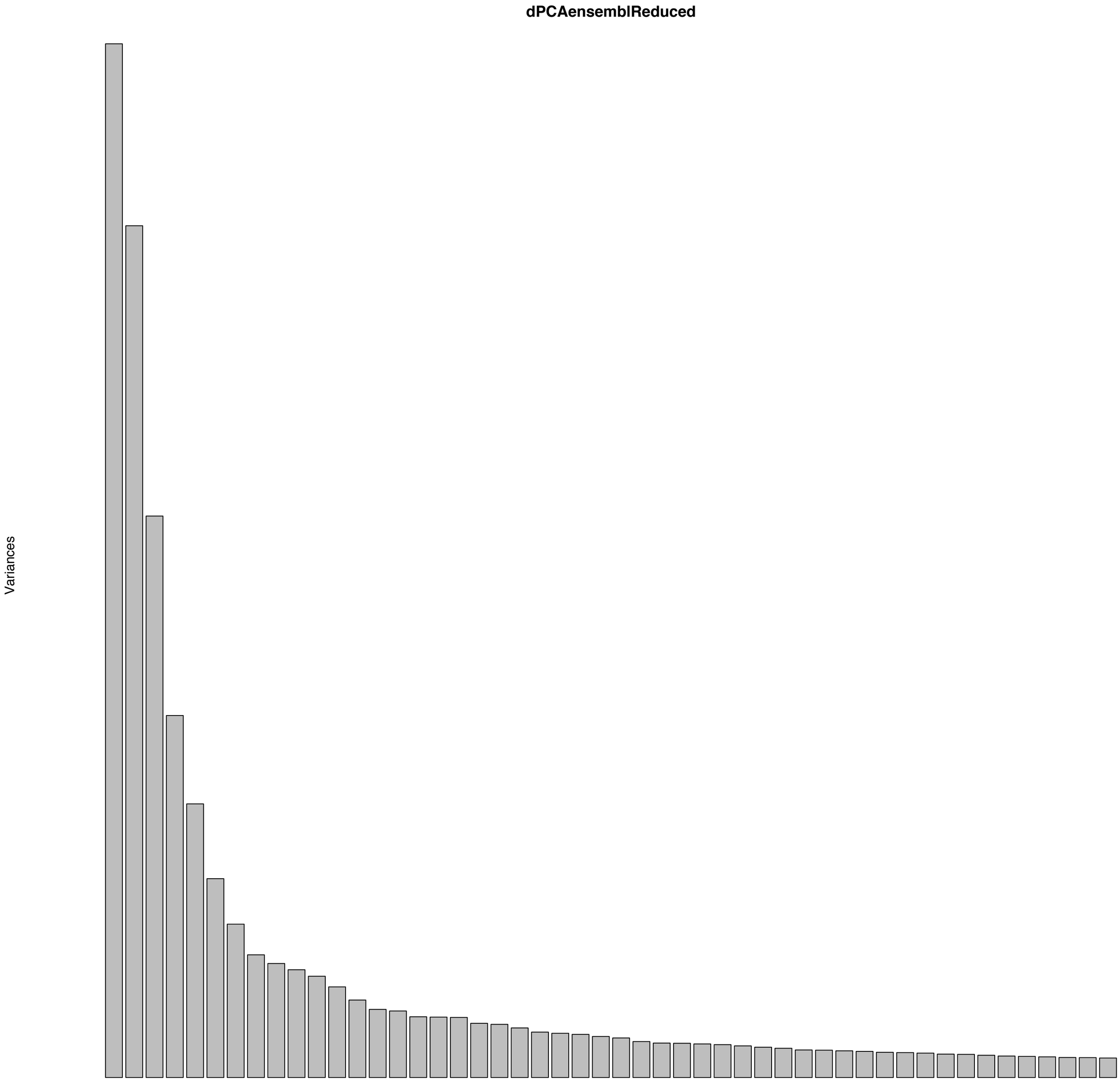
Samples colored by 6 clusters (see legend) identified on self-vs-self heat-map of 96 biological groups. The samples are visualized on the 1st and 2nd principal component plane.


brain
cell line
hematopoietic system
incompletely differentiated
muscle
solid tissue

Supplementary Figure 8e23:

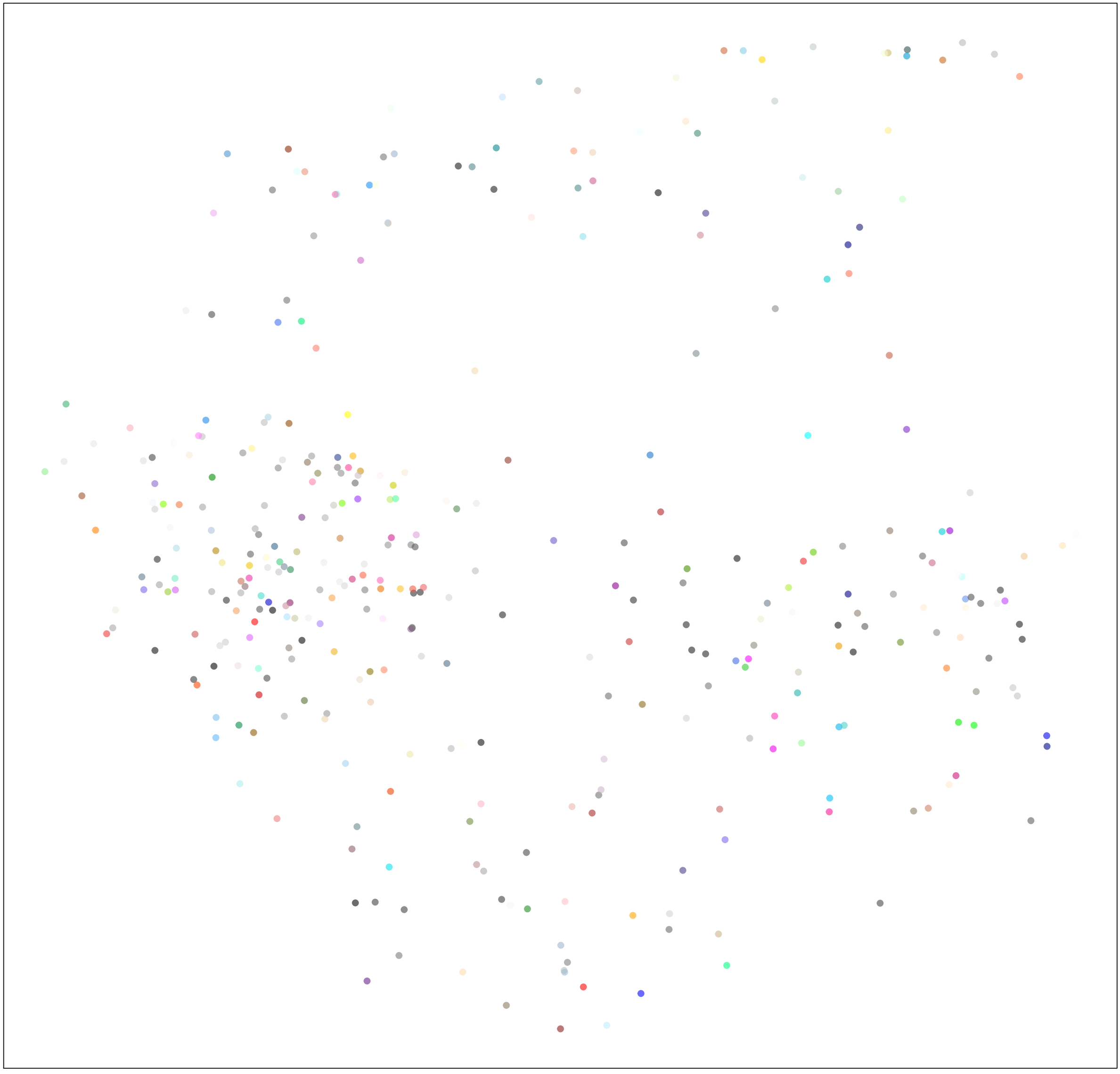
Principal component analysis on sampled data. Each dot represents one of the 5372 samples in a multidimensional gene expression space projected on the 'principal' plane formed by pairs of first 5 principal components. The dots are colored (semi-transparently) according to the biological group the sample belongs to (see figure legends).

Samples colored by 6 clusters (see legend) identified on self-vs-self heat-map of 96 biological groups. The samples are visualized on the 2nd and 3rd principal component plane.


dPCAensemblReduced
Variances

Supplementary Figure 8f:

Principal component analysis on sampled data. The variances captured by first 50 principal components.



Supplementary Figure 9:

Principal component analysis. Centroids of the 369 biological groups.

Methods

1. Data collection

We collected the data from ArrayExpress [1] and GEO [2] using Perl scripts to find and download Affymetrix HG-U133A array based experiments, including CEL files, in both databases. Files from two of the experiments with wrong array designs and corrupted files were rejected while the rest – 1925 from ArrayExpress and 7079 from GEO were retained for further analysis. GEO SOFT format files were processed automatically as described below. Annotations for the ArrayExpress experiments were used as submitted.

2. Duplicate files and data quality checks

We identified and removed 736 duplicate CEL files in the dataset. The remaining 8268 CEL files were subject to quality control using Bioconductor [3]. For all CEL files we used following quality measures: *scaling factor* (sfs), *average background* (avBg), *percentage of present calls* (pps), *RNA degradation from whole array* (RNAdeg), *Normalized Unscaled Standard Errors* (NUSE), and *Relative Log Expression* (RLE). We computed the last two measures in randomly selected subsets of ~1800 CEL files as the whole set was too big to be processed in R at once. Given the size of the random subsets, the probe level models and their parameters in each subset approximate the probe level models and their parameters computed from the whole set reasonably. We based the cut-offs for the quality control on recommendations of Bolstad et al [4] by adjusting these for the distribution of the values for each of the quality measures over the CEL files. The final cut-offs chosen for the acceptable quality CEL files were: $-2 < sfs < 2$, $5.2 < avBg < 7.2$, $RNAdeg < 4.5$, $Pps > 35$, $NUSE < 1.025$, and $-0.15 < RLE < -0.15$. Low quality indicates a CEL file is non-compatible with the other CEL files in the tested set, not necessarily that it is of low quality in an absolute sense. After the application of quality checks we removed further 2896 arrays, leaving 5372 unique comparable CEL files.

3. Array annotation

The HG-U133A array probesets were annotated via probeset sequence mapping on Ensembl (5) release 52 using the biomarRt Bioconductor package [5]. Of the 22283 Affymetrix probesets, 18,609 map to 14245 Ensembl genes.

4. Sample annotation

The data in ArrayExpress were annotated during submissions using the MGED Ontology (MO) [6] categories: *cell line, cell type, organism part, disease sate*, and *developmental stage;* these annotations were retained. Typically each sample has annotations for one or two of these categories. The free text annotations in GEO were converted to the MO categories using the text mining tool Whatizit [7], with a custom dictionary developed for this purpose; this step was followed by manual curation. For each sample a *compound* annotation was created by combining all available annotation values. The combining was done manually on case by case basis, generally following a simple set of rules:

a) If the sample source was a cell line, the cell line annotation was used together with the annotation of disease state, or if not available, with the annotation of organism part or cell type. In this study we do not include sample treatment in the annotation.

b) If the sample source was not a cell line, the sample group was defined by disease state annotation, or if not available, by cell type or organism part annotation.

c) In many cases the disease state annotation described the organism part or cell type (e.g., terms like 'muscular dystrophy' defines the 'organism part' along with

the 'disease state'). In these cases no organism part or cell type annotation was used. Unless the disease state was explicitly annotated, the samples were assumed to be normal.

d) Cell type if given was used in preference over organism part annotation.

The combination of these annotations produced 369 different terms of compound annotations, each defining a biologically similar group of samples. Each set of samples annotated by one of these 369 terms is referred to as a *biological group*.

Of the 369 biological groups, 96 contain 10 or more samples each, representing a total of 4634 samples, 738 samples were represented in 263 smaller groups. We re-grouped the biological groups into 4 and 15 different meta-classes, 15 classes of 'tissues of origin' and into several data driven meta-classes. The 'tissues of origin' meta-class was composed manually by first identifying for each sample its tissue of origin. Only these tissues of origin classes (15) were selected which contained samples from cell lines and non-cell lines.

5. Data normalization

The 5372 CEL files were RMA normalized using the Bioconductor package *affy* [8]. The resulting expression matrix contained 22283 probeset rows and 5372 columns. We call the rows in the normalized gene expression matrix gene expression profiles, and columns sample expression profiles.

6. Expression profiles and similarity computations and clustering

The Pearson correlation coefficient was used as the similarity measure between expression profiles. All correlation coefficients between sample expression profiles turned out to be positive. Average linkage distance [9] was used to define similarity between groups as follows. For each pair of samples from the two biological groups the correlation was calculated and averaged across all pairs. Thus each sample group was represented by a vector of 96 mean correlations to all the groups.

For clustering and heat map visualization we used the heatmap.2 function in R. The clustering of the input correlation distance matrices was carried out internally by the heatmap.2 function with the default parameters. These included usage of Euclidean distance as a distance measure and complete linkage for the agglomeration method.

7. Principal Component Analysis and dataset reduction by re-sampling

The Principal Component Analysis (PCA) analysis was performed in R. PCA was applied to the normalized data matrix in three different ways – first, using the complete dataset, second, using randomly sampled data, where the biological groups that had more than 10 samples were each represented by 10 randomly selected samples, and, third, representing each of the 369 groups by its centroid (i.e., for each gene replacing its expression value by the mean across all the samples in the group). The re-sampled dataset was constructed by limiting the number of replicates in all biological groups with more than 10 replicates to 10 by sampling without replacement. All three different PCA variations yielded the same result except that first and second principal components were swapped between the first and the last two analysis versions.

8. Computation of differentially expressed genes

To identify genes differentially expressed in specific biological groups we performed supervised analysis using the Bioconductor package *limma* [10] with a bootstrapping procedure in R. We identified genes specific to cell lines, neoplasms, non-neoplastic

diseases and more specific groups such as leukemia, breast cancer, tissues of origin, as well as to the six clusters revealed by unsupervised analysis. For each grouping one-way ANOVA were set up in limma with multiple group-wise comparisons to the global mean. The sample numbers in different group are imbalanced with, for instance, some diseases and tissues highly over-represented. To account for this, we used a re-sampling strategy, generating 50 sampled datasets, stratified by group membership and applying the differential expression analysis to each dataset. Probesets with Benjamini-Hochberg adjusted p-values of less than 0.01, with log fold change greater than 0.5, and identified consistently in each of the 50 trials were considered to be significant and were tested for Gene Ontology enrichment. All groups of genes identified for each of the sample groups together with the GO terms enriched in these groups are reported in supplementary Table 1.

9. GO over-representation analysis

We computed the Gene Ontology (GO) class over-representations using Fisher's exact test for whether a certain GO class appears unexpectedly often in a given gene list, compared to a random subset of the same size. q-values (9) were generated from the resulting p-values.

10. Investigation of the effect of cell line specific genes

To investigate if the sample expression profiles of cell lines could be brought closer to the profiles of the corresponding non cell line samples, we used the list of cell line specific genes (see above) and sample expression profiles with labels in the 'tissues of origin' meta-class. We created a similarity heatmap for the tissues of origin meta-class where cell line and non-cell line samples of each tissue of origin were split into separate groups and the cell line specific probesets from all samples were removed (supplementary Figure 7).

References

1. **Parkinson, H. et al. ArrayExpress update--from an archive of functional genomics experiments to the atlas of gene expression. *Nucleic Acids Res* 37, D868-72 (2009).**
2. **Barrett, T. et al. NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res* 37, D885-90 (2009).**
3. **Gentleman, R., Carey, V., Dudoit, S., Irizarry, R., Huber, W. (ed.) *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, 397-420 (Springer, New York, 2005).**
4. **Bolstad, B.M., Irizarry, R.A., Astrand, M. & Speed, T.P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19, 185-93 (2003).**
5. **Durinck, S. et al. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* 21, 3439-40 (2005).**
6. **Whetzel, P.L. et al. The MGED Ontology: a resource for semantics-based description of microarray experiments. *Bioinformatics* 22, 866-73 (2006).**
7. **Rebholz-Schuhmann, D., Arregui, M., Gaudan, M., Kirsch, H., and Jimeno, A. (2007). Text processing through Web services: Calling Whatizit. *Bioinformatics* 24, 296-298 (2007).**
8. **Irizarry, R.A., Gautier,L., Bolstad, B.M. and Miller, C. with contributions from Astrand, M., Cope, L.M., Gentleman, R., Gentry, J., Halling, C., Huber, W.,**

MacDonald, J., & Rubinstein, B.I.P., Workman, C. and Zhang, J. affy: Methods for Affymetrix Oligonucleotide Arrays. R package version 1.12.2. (2006).

9. Everitt, B. *Cluster Analysis*, (Heinemann, London, 1974).

10. Smyth, G.K. Linear Models for microarray data. in *Bioinformatics and Computational Biology Solutions using R and Bioconductor* (ed. Gentleman, R., Carey, V., Dudoit, S., Irizarry, R., and Huber, W.) 397-420 (Springer, New York, 2005).