

MAX-DELBRÜCK-CENTER FOR MOLECULAR MEDICINE

Spectrum Scale for Life Science Workload

Dr. Alf Wachsmann
Head of Scientific Computing

IBM Spectrum Scale User Group Meeting
Nov 12, 2017

Helmholtz Association

Largest research organization in Germany

- 18 national research centers (6 with focus on **Health Research**)
- More than 38,000 staff
- Budget: more than € 4 bn

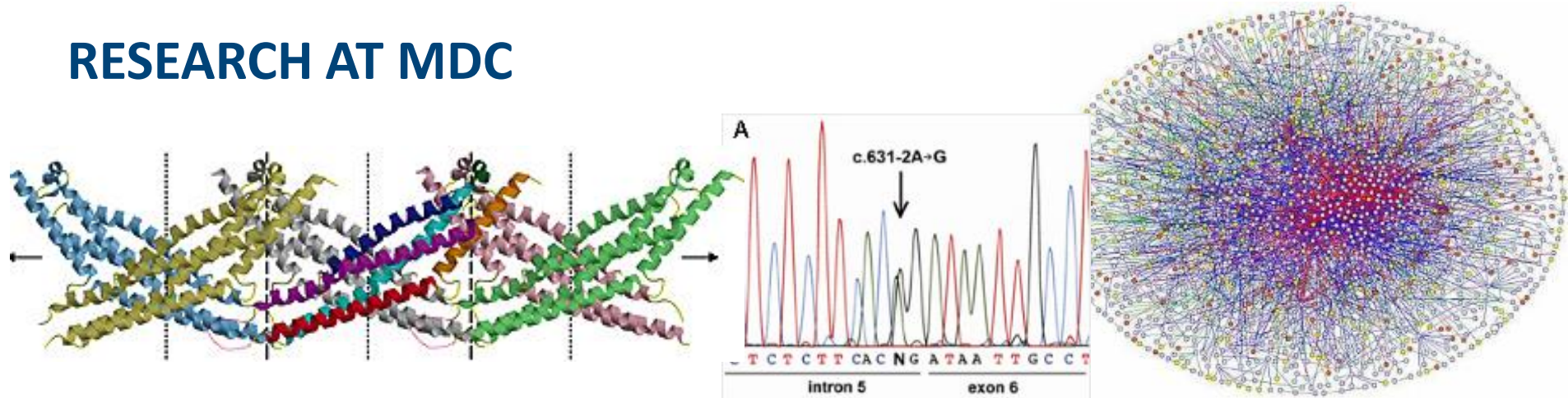
Max Delbrück Center for Molecular Medicine

Basic Facts 2016

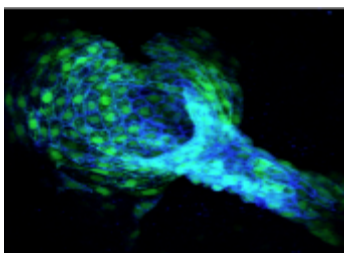
MDC Core Budget 2016	€ 87.5 M
Extramural Funding 2016	€ 30.7 M
Shared 2nd place for DFG-funding nationwide	
European Research Council (ERC) Grants	15
Research Groups	65
Staff (incl. guests), 25% international	1,660
Postdocs (incl. guests), 50% international	230
PhD Students (incl. guests), > 50% international	360
Patent families	95



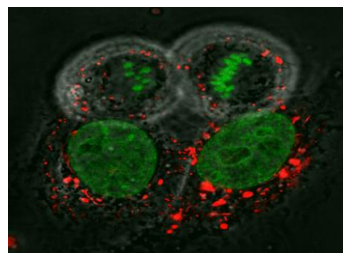
RESEARCH AT MDC



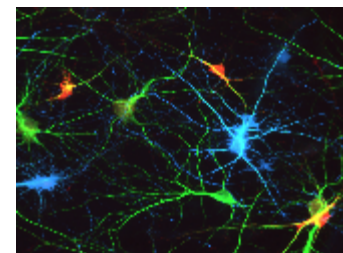
cell and molecular biology, signaling pathways, developmental biology,
physiology, structural biology, omics, systems biology, bioinformatics



Cardiovascular and
metabolic disease



Cancer



Diseases of the
nervous system



Medical Systems
Biology

INTERNAL DATAFLOWS



Gene Sequencers



CIFS to Spectrum
Scale CES



Mass Spectrometers



1 / 10 Gb/s
Network



Microscopes



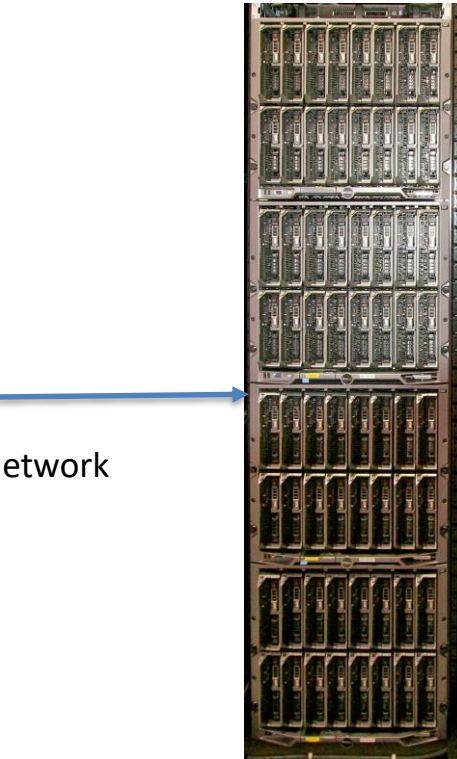
Slower Storage
for „cool data“
NFS, CIFS



Fast Storage for
„hot data“
GPFS

40 Gb/s Network

40 Gb/s Network



Compute Cluster
GPFS Clients

Tape Archive (IBM TS4500 with 9 TS1150 and
Versity Storage
Manager)



I/O CHARACTERISTICS

- One storage system and one compute cluster for all of our use cases
- Applications with very different I/O characteristics
- Data of very different sizes

- Tried Lustre in my previous job for similar workload: small-block I/O didn't work well
- Tried BeeGFS two years ago: seems to work well but not better than GPFS to switch

GPFS FEATURES IN USE AT MDC

- MDC is using only very few of the GPFS features
 - Quotas
 - Snapshots
- Plan to use rules engine to find „cool data“ and notify users
 - Later maybe migrate the data automatically
- Will investigate changing HSM to IBM Spectrum Protect (Tivoli Storage Manager) for better integration with GPFS next year

HPC

RFP is out for
~70 compute nodes with
a 40 Gb/s backbone and
5 Tesla P100 GPUs

GPFS is now ~830 TB usable

NFS space on
5 Oracle ZFS appliances (HA)
and 2 Dell servers.
Total of ~3.4 PB usable space

