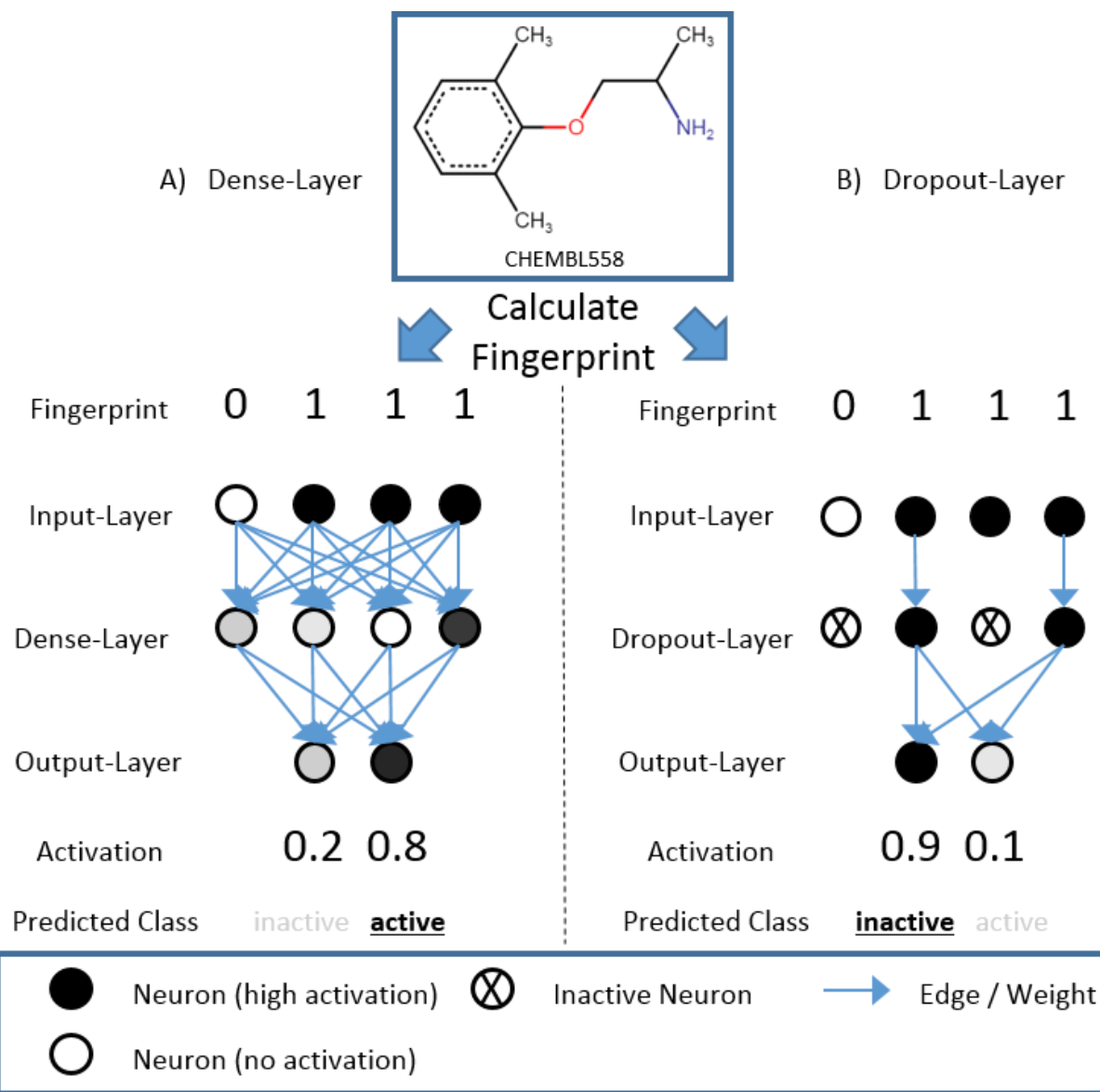


Automatic Generation of Neural Network Architectures using a Genetic Algorithm

Wolf-Guido Bolick, [Paul Czodrowski](#)
RDKit UserGroupMeeting, Berlin, September 20th 2017

MERCK

What is a DNN (deep neural network)?



- Morgan fingerprints as input
- Keras for learning deep neural networks
- scikit-learn for statistical measures & train/test set splits

DNN Architectures & Hyperparameters

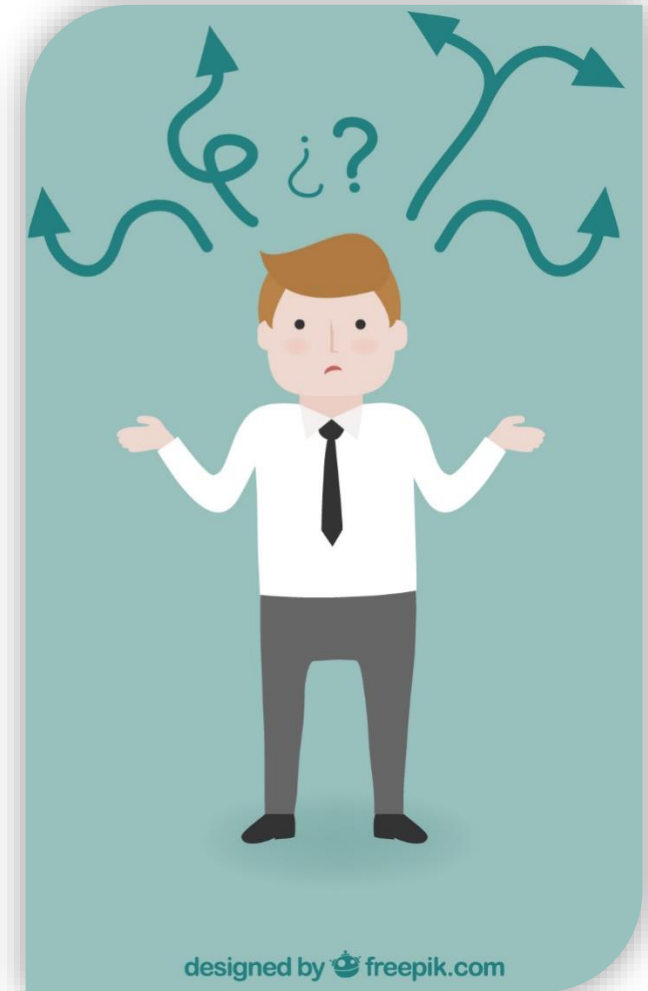
NN-Architecture

- Layer-Type
- Layers
- Neurons per Layer
- Activation-Functions

Training-Parameters

- Optimizer
- Learning-Rate
- Weight-Decay
- Batch-Size
- Loss-Function
- ...

Hyperparameters



Optimization of Hyperparameters

Expert



- Hyperparameters derived from literature & **experience**
- Hyperparameter search within promising parameter areas

Lucky People



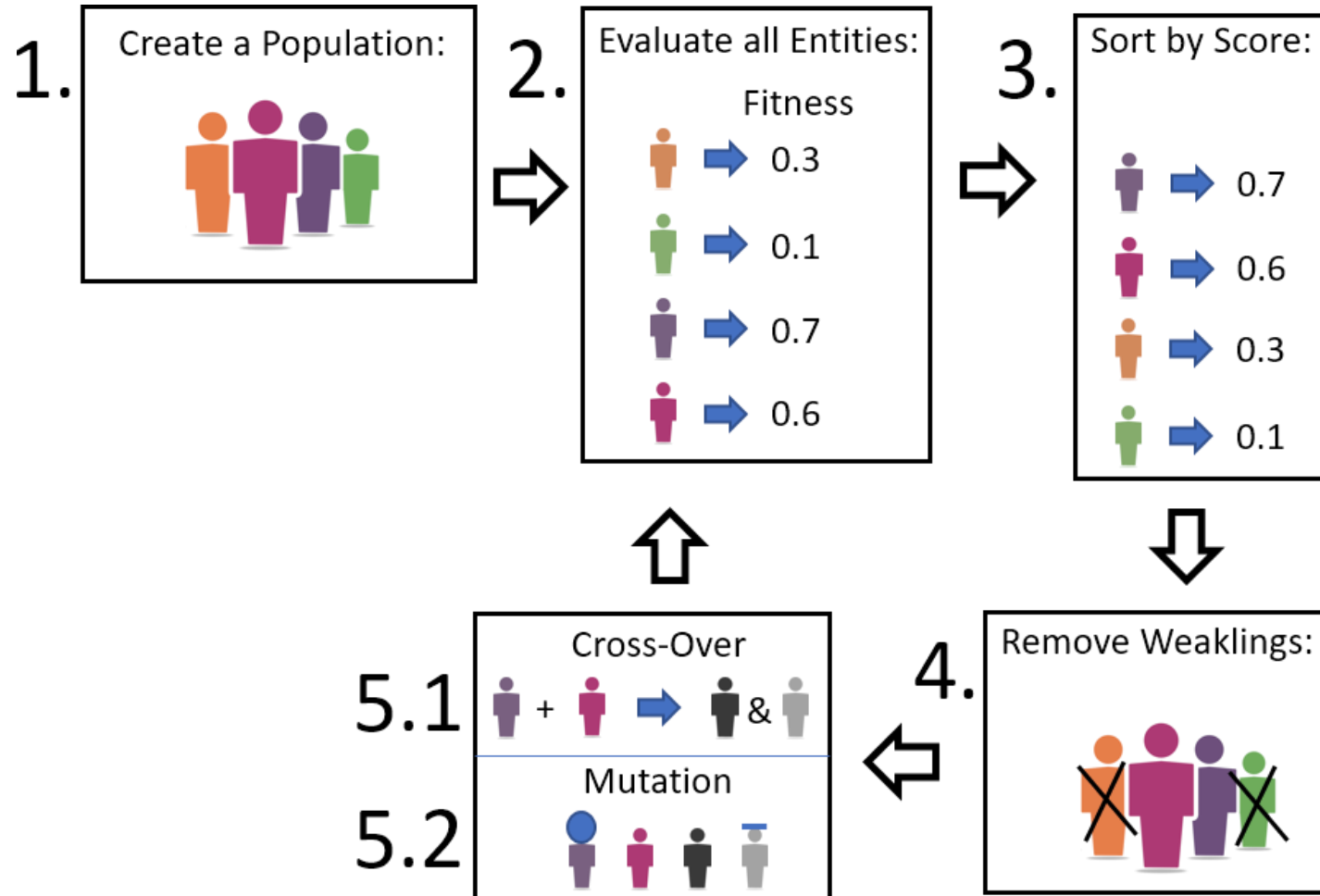
- Random-Search

Everyone

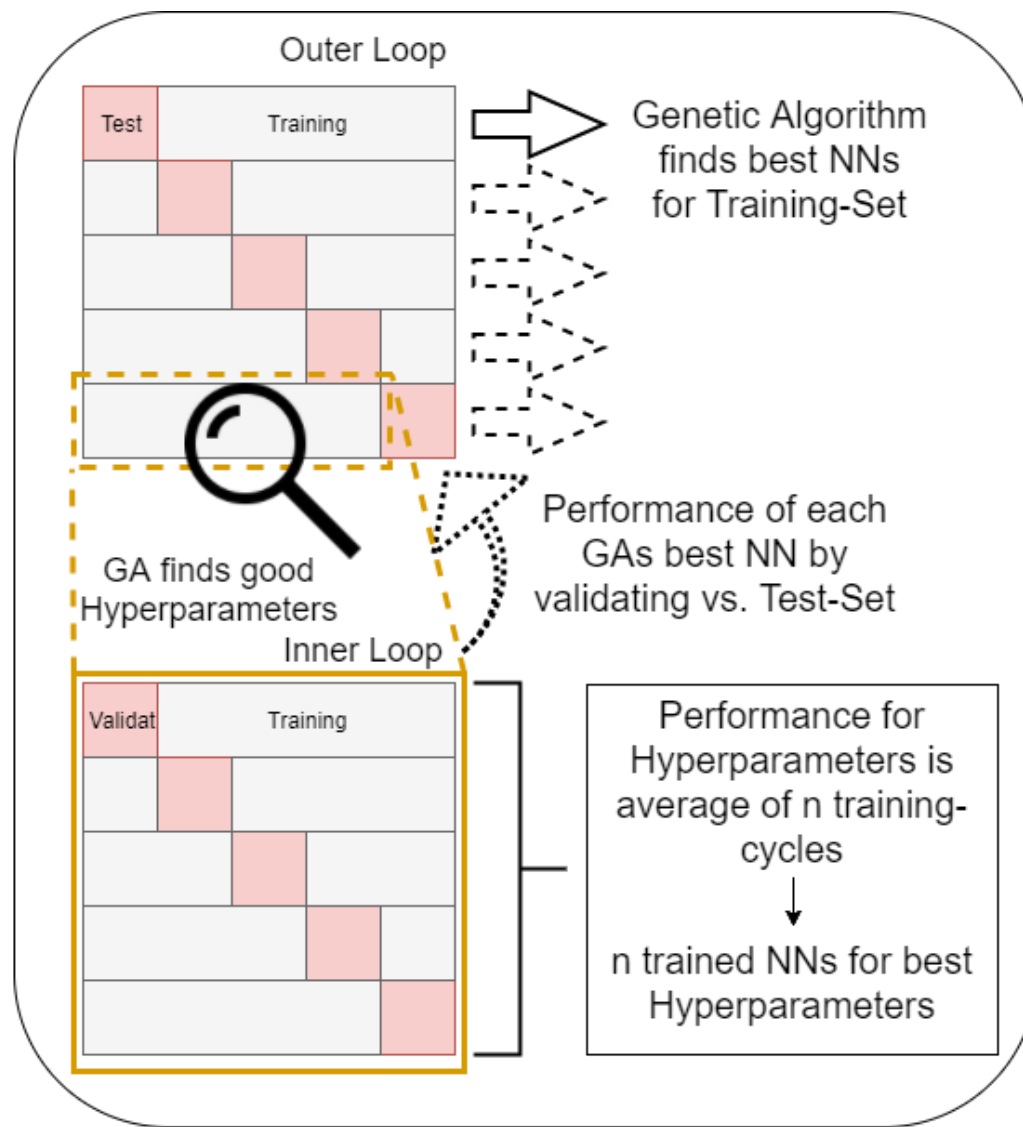


- Grid-Search
- Probability based algorithms
- Directed Random-Search (e.g. **genetic algorithms**)

What is a Genetic Algorithm (GA)?

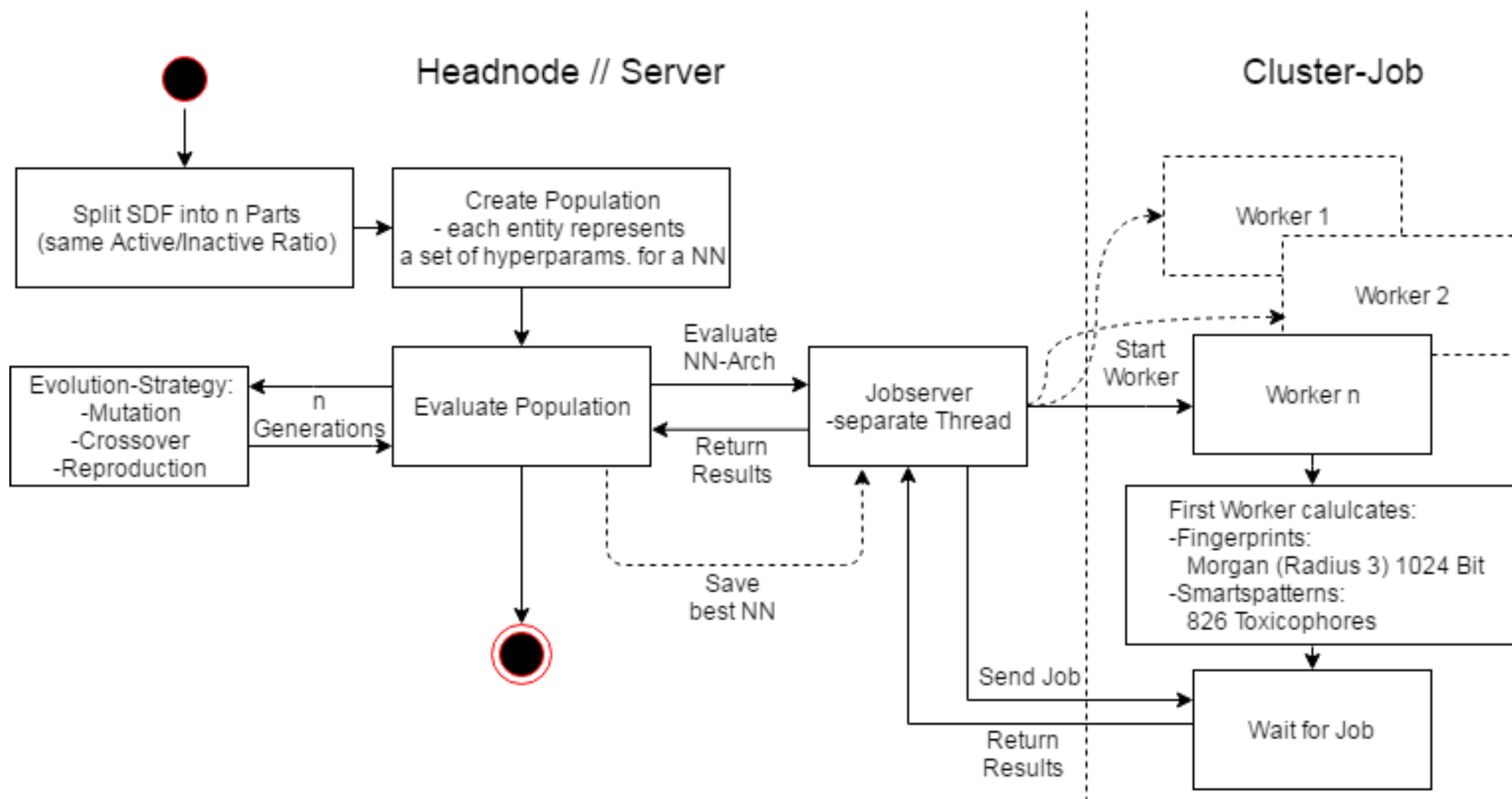


Validation strategy: nested cross validation

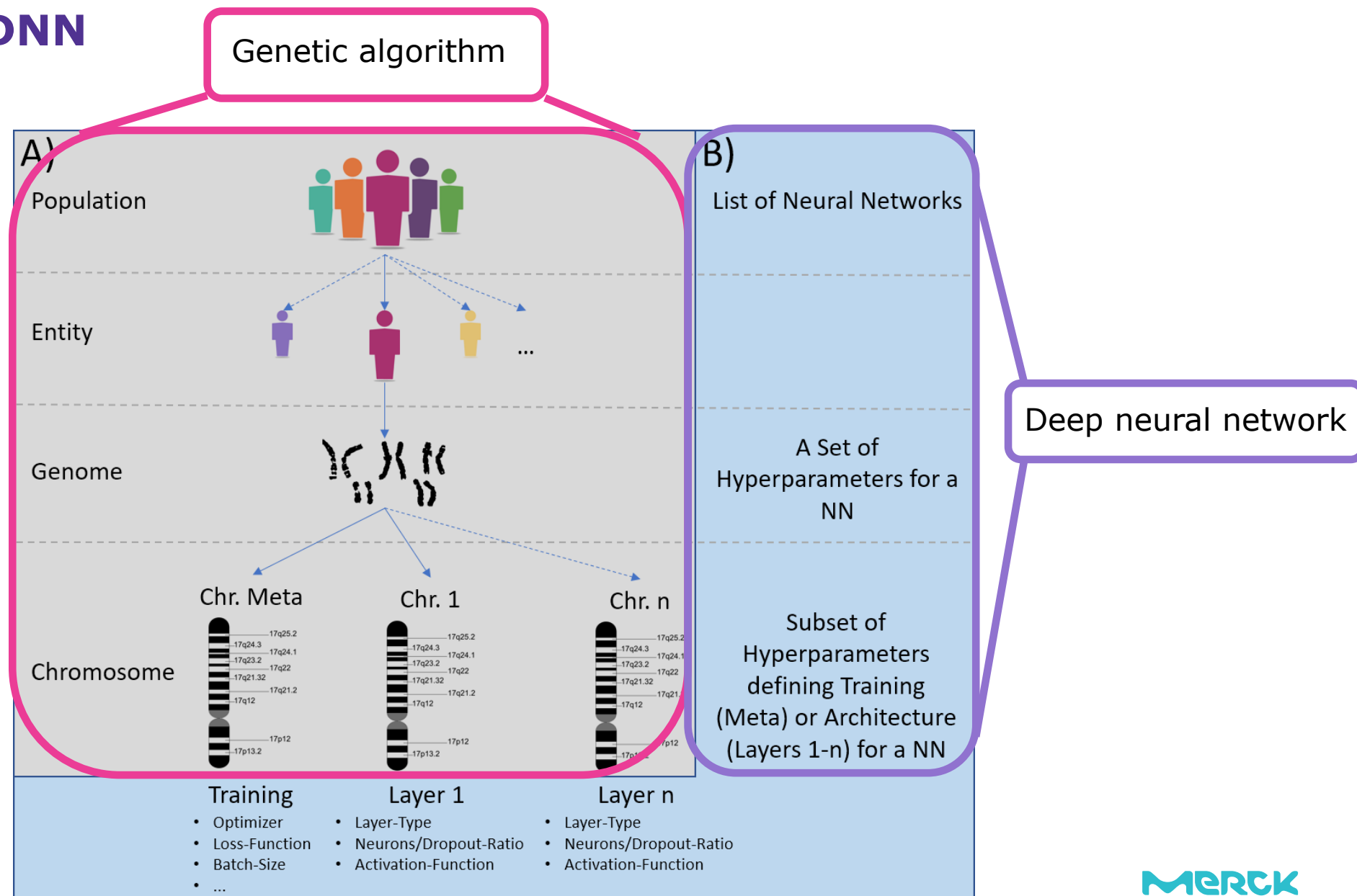


- **consensus Model of 25 individual Models**
- **Hyperparameter search inside inner loops & validated inside outer loops**
- **Every compound is represented in 16/25 individual Models**

Implemented workflow



Analogy GA & DNN



GA settings

Parameter	Default value
Population size	100
Worker	10
Fingerprint size	1024
Evolution strategy	drop worst 50 %
SMARTS-Patterns	826

Parameter	Default value
Optimizer	sgd, rmsprop, adagrad, adadelata, adam, adamax, nadam
Loss-Function	mse, mae, msle
Learning-Rate	5E-2, 1E-1, 5E-1, 1.0
Weight-Decay	5E-7, 1E-7, 0.0
Momentum	0.0, 0.1, ..., 0.9
Nesterov	0, 1
Batch-Size	5%, 6%, ..., 20%

Parameter	Default value
Chromosomes	2 – 5
Layer types	Dense, Dropout
Neurons	32 – 512
Neurons stepsize	32
Dropout ratio	5, 10, ..., 90%
Activation functions	linear, sigmoid, hardsigmoid, softmax, relu, tanh

Results for in house hERG data set

- outer kappa: almost 0.6 → Moderate Model
- nested cv: computation time ca. 8 – 14 hours on GPU cluster

NN-Architecture	Split 1	Split 2	Split 3	Split 4	Split 5
Optimizer	sgd	adadelta	sgd	sgd	adadelta
Loss-Function	mae	msle	msle	mae	msle
Learning-Rate	1.0	0.5	1.0	1.0	1.0
Batch-Size	0.2	0.09	0.14	0.15	0.12
Momentum	0.1	0.8	0.6	0.4	0.3
Nesterov-Momentum	1.0	-	1.0	1.0	-
Weight-Decay	5e-07	5e-07	-	-	1e-07

Layer 1

Layer-Type	dense	dense	dense	dense	dense
Activation-Function	relu	relu	tanh	relu	relu
Neurons	288	416	64	480	128
Dropout-Ratio	-	-	-	-	-

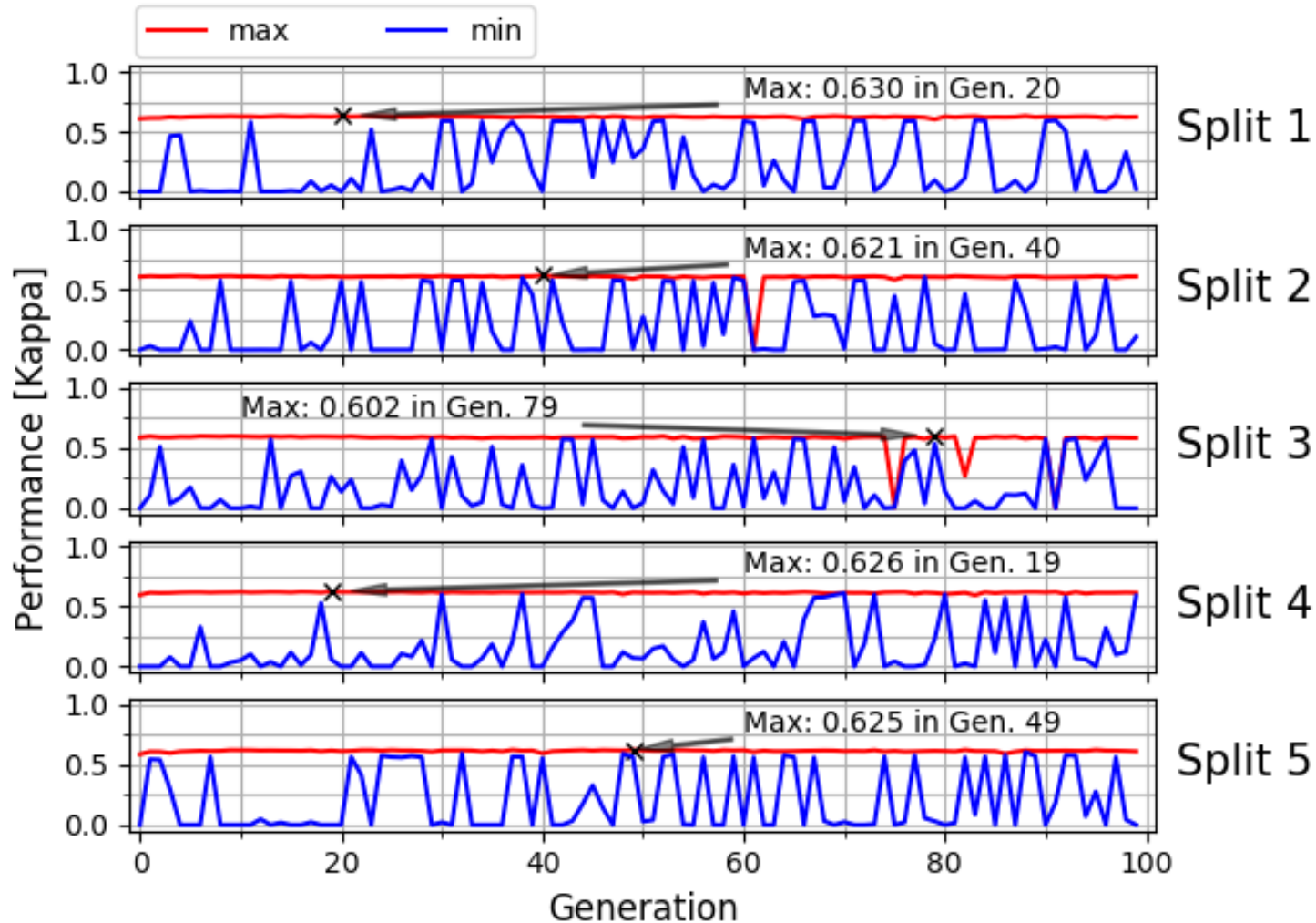
Layer 2

Layer-Type	-	dense	dense	dropout	-
Activation-Function	-	relu	softmax	-	-
Neurons	-	64	384	-	-
Dropout-Ratio	-	-	-	0.15	-

Performance

Inner Kappa	0.63 ± 0.02	0.62 ± 0.01	0.60 ± 0.03	0.63 ± 0.02	0.63 ± 0.02
Outer Kappa	0.57 ± 0.02	0.56 ± 0.04	0.56 ± 0.05	0.58 ± 0.02	0.59 ± 0.01
Generation	20	40	79	19	49
# Architectures	1318	1157	1146	1364	1403

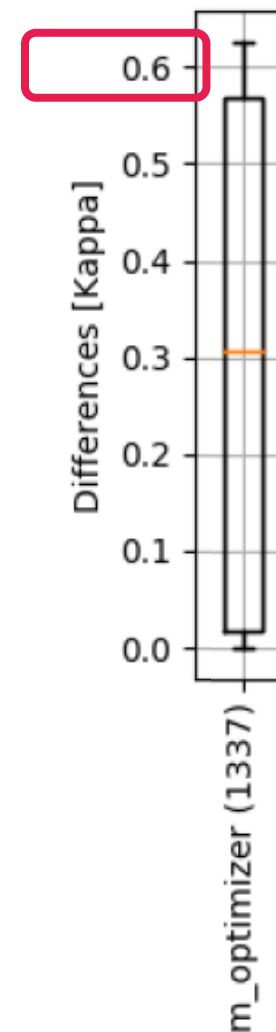
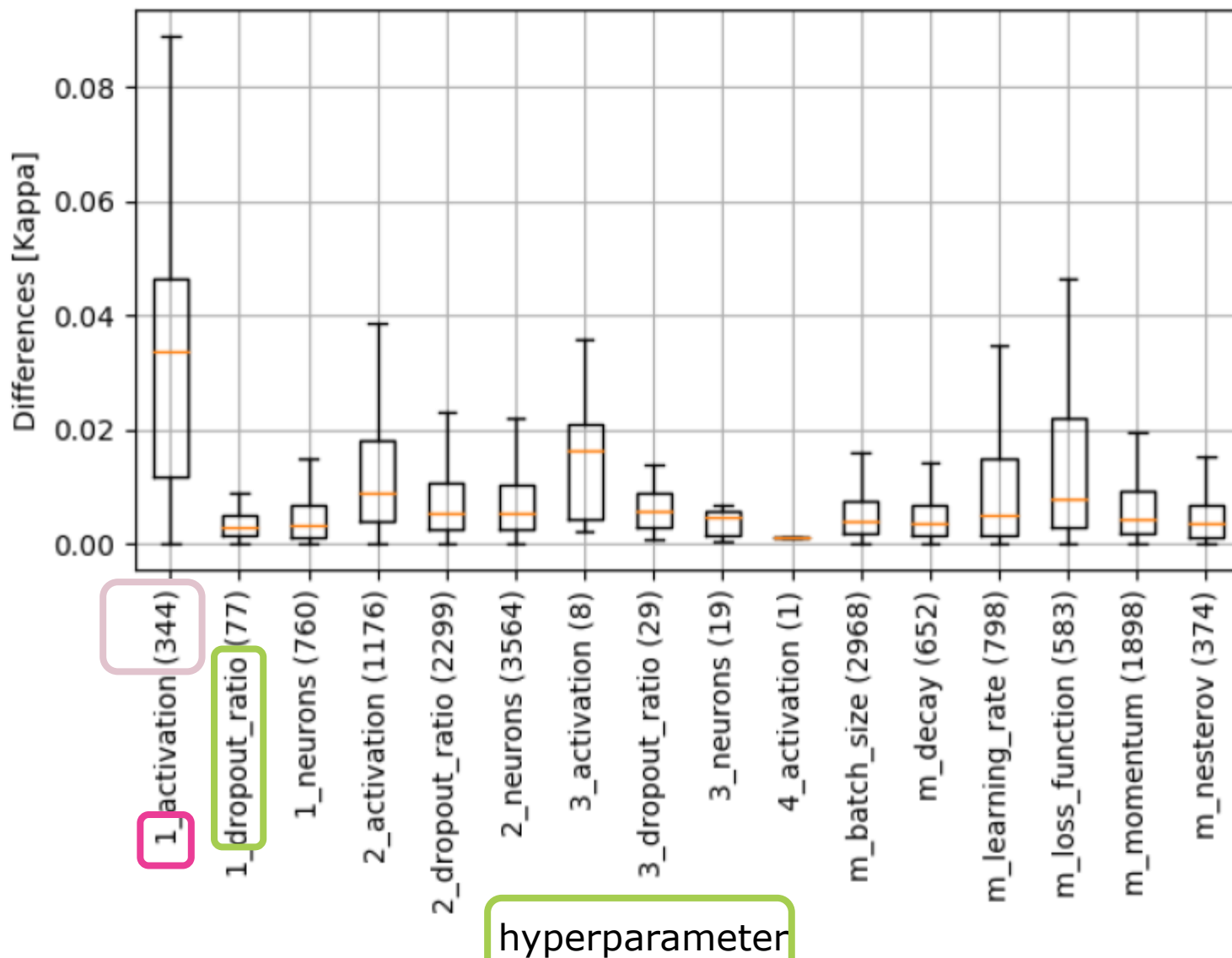
DNN performance over the GA generations



- strong fluctuation of minimum performance of DNN
- in contrast, stable maximum performance

Influence of individual parameters

Optimizer



Interface

```

##### 25 Aug 17 16:45:41 #####
MAL0D0XB15 55%-dropout 512-tanh kappa: 0.48 Penal: 0.011 State: 100.0%
MRL0D0XB2 35%-dropout 192-tanh kappa: 0.09 Penal: 0.027 State: 100.0%
MSL1D0M1nXB8 128-softmax 30%-dropout kappa: nan Penal: nan State: 80.2%
28-hard_si 60-tanh 5%-dropout kappa: nan Penal: 0.02 State: 80.0%
MAL1D2XB1 320-softmax 20%-dropout kappa: nan Penal: 0.01 State: 60.1%
MRL2D2XB15 256-linear kappa: 0.27 Penal: 0.003 State: 100.0%
MAL1D2XB2 0%-dropout 32-softmax kappa: nan Penal: 0.025 State: 40.3%
MRL3D2XB0 50%-dropout 128-hard_si kappa: nan Penal: 0.01 State: 60.1%
MAL1D0XB13 256-tanh kappa: nan Penal: 0.003 State: 80.2%

# <SPACE> - add a Worker
# <BACKSPACE> - kill a Worker
# <ESC> - Stop & Quit
# <ARROW-LEFT>- prev. Entity
# <ARROW-RIGHT>- next Entity
# 2 free workers
# last_key: 261
#
# Description:
# Entity of Gen 0
# Chromosome: 0
# Type: meta
# Optimizer: adamax
# Loss-Function: msle
# Learning-Rate: 0.1
# Decay: 1e-07
# Momentum: None
# Nesterov: None
# Last Activation: softmax
# Batch Size: 0.06
# epsilon: 1e-08
# Chromosome: 1
# Type: dense
# Neurons: 128
# Activation: hard_sigmoid
# Chromosome: 2
# Type: dense
# Neurons: 160
# Activation: tanh
# Chromosome: 3
# Type: dropout
# Dropout Ratio: 0.85
# Chromosome: 4
# Type: dense
# Neurons: 512
# Activation: relu
# kappa: nan
# Penal: 0.027
# State: 80.0%
#

MSL0D0M4NXB6 416-linear 45%-dro-out kappa: nan Penal: 0.027 State: 20.0%
MAL2D2XB4 384-reLU 32-sigmoid 85%-dropout kappa: nan Penal: 0.018 State: 20.1%
MAL0D0XB10 480-tanh 32-sigmoid 85%-dropout kappa: nan Penal: 0.018 State: 20.1%
MAL3D1XB9 20%-dropout 5%-dropout kappa: nan Penal: 0.003 State: 0.2%
MAL0D0XB15 320-tanh 352-sigmoid kappa: nan Penal: 0.027 State: 20.0%
MAL1D2XB2 0%-dropout 32-softmax kappa: nan Penal: 0.025 State: 40.3%
MRL3D2XB0 50%-dropout 128-hard_si kappa: nan Penal: 0.01 State: 60.1%
MAL1D0XB13 256-tanh kappa: nan Penal: 0.003 State: 80.2%

best NN

evaluated NN not evaluated yet

selected to see more details

current generation

evaluated entities in current gen.

inner kappa of best found NN

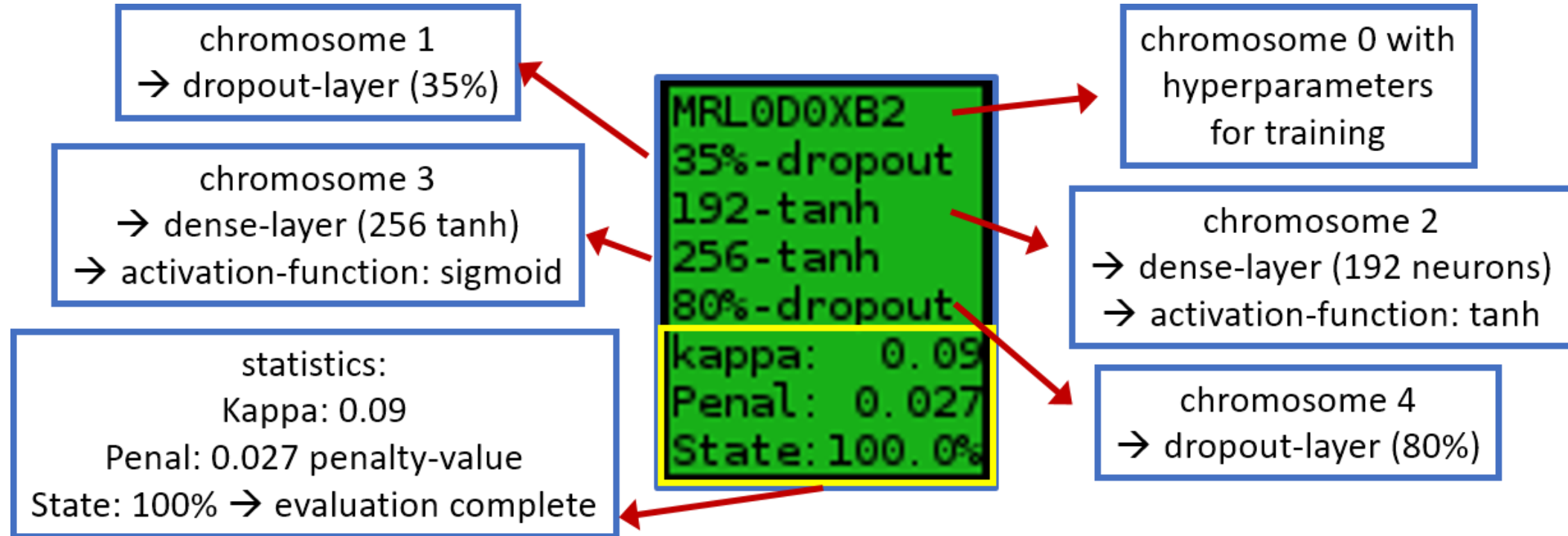
active workers

Generation: 0, Evaluated: 3/100, Best-kappa: 0.465, Workers: 13, Info: last_Key: 261 - press <SPACE>: to add a Worker <BACKSPACE>: to kill a worker - 0

```

• interactive tracking of GA in the console
 • implemented in curses

Interface, ctd



Detailed view of one particular architecture
Statistics given as well

Summary

- (Almost) stable implementation of a genetic algorithm the for hyperparameter search in Deep Neural Networks
 - ➡ Follow-up studies on-going at Merck
- Comparison with other machine learning techniques
 - ➡ Coming soon
- Roll-out of the code
 - ➡ First step: publish a paper (not a single line written so far)!

Acknowledgment

- Wolf-Guido Bolick
- Fabian Steinmetz
- Kristina Preuer, Günter Klambauer (Hochreiter group)
- Andreas Dominik (THM/JLU Gießen)